The Experimenter Bias Effect and Its Relation to Submission to Authority

Anat Rutenberg

Wilfrid Laurier University
THE EXPERIMENTER BIAS EFFECT
AND ITS RELATION TO SUBMISSION TO AUTHORITY

By

Anat Rutenberg
B.A. Tel-Aviv University, 1967

THESIS

Submitted in partial fulfilment of the requirements
for the Master of Arts Degree
Wilfrid Laurier University
1975

Examinig Committee

Mr. Alan Auerbach, Chairman
Dr. Sidney Hellyer, Dept. of Psychology
Dr. Josephine Naidoo, Dept. of Psychology
Abstract

Scientific researchers are expected to control the "Experimenter" or "Rosenthal" effect, in which the experimenter's (E's) expectations somehow induce the experimental results. Some seventy studies dealing with experimenter bias effect (EBE), many of them recent, were critically reviewed. Contradictory results and failures to replicate were reported, and tentative explanations of inconsistencies offered. The most plausible explanation involved the personality characteristics of, and the dominance relation between, E and S. Accordingly, the present study examined one personality trait which seems to underlie the EBE, namely, submission to authority.

To elicit EBE, students serving as Es administered to Ss Rosenthal's Photo Rating Test (RPRT), a series of photographed faces which Ss rate on an ordinal "success-failure" scale. Some Es were led to expect that their Ss would perceive "success" in the faces; other Es expected their Ss to perceive "failure".

In a preliminary study RPRT was examined by having students rate the pictures on two occasions under neutral, i.e., "no-expectancy" conditions. Only 11 out of the 20 pictures received near-zero ratings, and a large rating dispersion was found. Consequently results were calculated separately for all the pictures and for the most "neutral" ones.
Focussing on S personality variables, Experiment One tested the hypothesis that Ss who were more submissive to authority were also more susceptible to EBE. Ss selected for either high or low dominance were given RPRT by medium dominant Es. Some Es were led to expect high ratings and others to expect low ratings from their Ss.

Simultaneously, focussing on E variables, Experiment Two tested the hypothesis that dominance and submission in E affected S's susceptibility to EBE. The task and the expectancies given to Es were the same as in Experiment One.

In both experiments Es and Ss were classified as dominant, medium dominant or submissive if they scored within the upper, middle, or lower range of their sample's distribution on Gold's Dominant-Submission Scale. The interaction between dominance level and experimental situation responses was examined by a hidden observer, recording smiles, glances and test duration, and by post-test questionnaires.

No EBE was found in either experiment, yet submissive Ss reported a significantly higher pressure to rate the pictures in a certain way than dominant Ss.

The results of both studies were interpreted in terms of problems surrounding this area of research. Finally, a proposed model of personality and situational variables which can be expected to elicit EBE was presented.
Acknowledgments

The author wishes to express her appreciation to several people who assisted in the preparation of this thesis. Mr. Alan Auerbach, thesis committee chairman, has been an advisor for the last two years. His contribution to the research project, steadfast support and encouragement are most gratefully acknowledged. Dr. Sydney Hellyer and Dr. Josephine Naidoo were members of the thesis committee. Their interest in the work and critical observations are greatly valued.

Gail and Dennis Jackson served as "the assistants" in the experiments. Their enthusiasm and dedication are sincerely appreciated. Drs Jim Dickenson and Mary Kay Lane, Miss Pauline Lamothe and Mr. Sheldon Rahn, devoted lecture time for some of the experiment. Dinah Emery spent many hours typing the manuscript. Her endeavour to complete the work quickly is deeply appreciated. Special thanks are due to my husband Avigdor for helping with the English.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Review of the Literature</td>
<td>3</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>100</td>
</tr>
<tr>
<td>Preliminary Study</td>
<td>101</td>
</tr>
<tr>
<td>Method</td>
<td>103</td>
</tr>
<tr>
<td>Results</td>
<td>107</td>
</tr>
<tr>
<td>Discussion</td>
<td>143</td>
</tr>
<tr>
<td>Appendix A</td>
<td>155</td>
</tr>
<tr>
<td>Appendix B</td>
<td>159</td>
</tr>
<tr>
<td>Appendix C</td>
<td>161</td>
</tr>
<tr>
<td>Appendix D</td>
<td>167</td>
</tr>
<tr>
<td>Appendix E</td>
<td>168</td>
</tr>
<tr>
<td>Appendix F</td>
<td>169</td>
</tr>
<tr>
<td>Appendix G</td>
<td>171</td>
</tr>
<tr>
<td>Appendix H</td>
<td>173</td>
</tr>
<tr>
<td>Appendix I</td>
<td>174</td>
</tr>
<tr>
<td>Appendix J</td>
<td>175</td>
</tr>
<tr>
<td>Appendix K</td>
<td>176</td>
</tr>
<tr>
<td>Appendix L</td>
<td>179</td>
</tr>
<tr>
<td>Appendix M</td>
<td>181</td>
</tr>
<tr>
<td>Appendix N</td>
<td>182</td>
</tr>
<tr>
<td>References</td>
<td>185</td>
</tr>
<tr>
<td>Table Number</td>
<td>Title</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>1.</td>
<td>EBE in Intelligence Tests, Summary of Studies Reviewed.</td>
</tr>
<tr>
<td>2.</td>
<td>EBE in Real Life Situations, Summary of Studies Reviewed.</td>
</tr>
<tr>
<td>3.</td>
<td>EBE and Es' Levels of Anxiety, Summary of Studies Reviewed.</td>
</tr>
<tr>
<td>4.</td>
<td>Assignment of Ss to Es, Experiment One.</td>
</tr>
<tr>
<td>5.</td>
<td>Assignment of Ss to Es, Experiment Two.</td>
</tr>
<tr>
<td>6.</td>
<td>Average Rating Changes and Significance Levels, Dominant vs. Submissive Ss: Set I Data.</td>
</tr>
<tr>
<td>7.</td>
<td>Average Rating Changes and Significance Levels, Dominant vs. Submissive Ss: Set II Data.</td>
</tr>
<tr>
<td>8.</td>
<td>Analysis of Variance: 2x2 Factorial Design. Es' Expectancy x Ss Dominance: Set I Data.</td>
</tr>
<tr>
<td>9.</td>
<td>Analysis of Variance: 2x2 Factorial Design. Es' Expectancy x Ss Dominance: Set II Data.</td>
</tr>
<tr>
<td>10.</td>
<td>Means, Standard Deviations and Significance Levels, Submissive vs. Dominant Ss: Mutual Glances, Es' Smiles and Time Spent.</td>
</tr>
<tr>
<td>11.</td>
<td>Means, Standard Deviations and Significance Levels, Submissive vs. Dominant Ss: Perception of Es.</td>
</tr>
<tr>
<td>12.</td>
<td>Average Rating Change and Significance Levels, Dominant vs. Submissive Es: Set I Data.</td>
</tr>
<tr>
<td>13.</td>
<td>Average Rating Change and Significance Levels, Dominant vs. Submissive Es: Set II Data.</td>
</tr>
<tr>
<td>14.</td>
<td>Analysis of Variance: 2x2 Factorial Design. Es' Expectancy x Es' Dominance: Set I Data.</td>
</tr>
</tbody>
</table>
15. Analysis of Variance: 2x2 Factorial Design. Es' Expectancy x Es' Dominance: Set II Data. 134


17. Results for Question 2 of Questionnaire for the E. 137

18. Results for Question 9 of Questionnaire for the E. 139

19. Means, Standard Deviations and Significance Levels, Submissive vs. Dominant Es: E's Questionnaire. 142

20. Preliminary Study: Rosenthal's Photo Rating Test: Means, Standard Deviations, Correlations Between Rating and Rerating and Significance Levels. 157
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>EBE Model: Diagrammatic Representation.</td>
<td>97</td>
</tr>
<tr>
<td>2.</td>
<td>Analysis of Variance: 2x2 Factorial Design Interaction. Ss' Dominance x Es' Expectancy.</td>
<td>123</td>
</tr>
</tbody>
</table>
Introduction

Whenever a behavioral scientist encounters a human research subject a social situation is created which must be taken into account in interpreting the results. Its general importance derives from the fact that the interaction of experimenter and subject, like other two-person interactions, may be investigated empirically with a view to teaching us more about dyadic interaction in general. Its specific importance derives from the fact that the interaction of experimenter and subject, unlike other dyadic interaction, is a major source of information in the behavioral sciences. An important role in this interaction is played by "experimenter effects", i.e., inadvertent influences of the experimenter on the results of his research. These include the experimenter's personal attributes, e.g., sex, and personality traits, and experimenter bias effects (EBE) which are produced by the experimenter's expectancies, desires, or biases. In real-life situations the "experimenter effects" have their counterpart in "self-fulfilling prophecies". These effects have been shown by Rosenthal and many others to exist under diverse laboratory tasks and in real life.

The present study deals with EBE and focuses on the personality traits which are associated with the experimenter's ability to affect his subjects' responses in the laboratory. The study examines experimentally the role of a particular personality trait which seems to underlie the
EBE, namely, submission to authority.

The personality, attitudinal and situational variables relevant to EBE are identified in the literature review. A model showing the interaction of the most important of these variables is proposed to be used as a framework for future research on EBE.
Review of the Literature

Rosenthal's original study in the laboratory situation used a person perception task as the criterion instrument. The experimenter (E) showed the subject (S) a series of photographed faces to be rated on the dimensions of "experiencing failure" or "experiencing success". The rating scale ran from -10 (extreme failure) to +10 (extreme success) with intermediate labeled points. Previously, the person-perception task had been administered under a "no-expectancy condition" to a large number of student Ss. The average rating given to the photographs had been very close to 0; i.e., Ss perceived the persons depicted as "neutral" with respect to having experienced failure or success. The EBE was created by telling the E's before running their Ss that "the subjects you are running should average about +5 rating" or that "the subjects you are running should average about -5 rating". In addition, Es were typically told that the expected results had been "well established" in previous studies which used Rosenthal's Photo Rating Test (RPRT), that they would conduct the experiments in order to obtain practice in "duplicating experimental results", and if their results came out "properly as expected" they would be paid $2.00 per hour; whereas if their results did not come out "properly as expected" they would be paid $1.00 per hour. Most of the studies on EBE were carried out by means of RPRT. Using this task, Rosenthal himself and many others (Friedman 1967; Duncan, Rosenberg and Finkelstein
1969; Minor 1970; McFall 1970; Smith and Flenning 1971) were able to demonstrate the EBE.

However, other investigators failed to find EBE, (e.g., Barber et al. 1969; Wessler and Strauss 1968; Bootzin 1971; Hertzog and Walker 1973). A detailed review of those studies is presented subsequently. For instance, Barber, Calverley, Forgione, McPeake, Chaves and Bowen (1969) conducted five investigations in an attempt to replicate Rosenthal's results, none of which elicited EBE.

Barber et al. randomly selected Es and Ss from introductory psychology, whereas in Rosenthal's experiments Es were usually graduate students who ran introductory psychology Ss. Since undergraduates regard graduate students as having higher professional status, it may be reasonable to assume that in the latter situation Ss were more affected by Es. At the outset of their experiment Barber et al. were sceptical about the ability of Es to communicate their expectations to Ss solely by means of standard instructions. They assumed that in Rosenthal's experiment Es reinforced their Ss to give "proper" ratings verbally, or misreported "improper" ratings. It may be argued that this negative attitude must have been to some extent apparent in Barber's instructions to Es, resulting in a low level of bias. It must be pointed out, however, that in this respect Barber is an exception; most investigators who failed to elicit EBE appear to have expected to find it.

The wide applicability of the EBE phenomenon as
well as some interpretive pitfalls, can be seen in a recent study by Uno, Frager, Takashima, Shibamoto and Rosenthal (1974). The graduate chemistry student E's were given, before running each S, either positive or negative expectations regarding S's ratings. The S's were 40 male and 40 female literature undergraduates. The results showed a "non-significant tendency" of S's to rate the pictures according to E's expectations. The S's tested later by the same E showed a greater bias effect than those tested earlier. Data obtained from S's tested later showed EBE occurrence at p<.075. This was accepted by the authors as significant, although one would expect a more conservative approach when dealing with a phenomenon whose existence is still a matter of controversy. Attempts to detect scoring errors were not reported; therefore, it is impossible to assess their contribution to the overall effect. Thus, the study cannot be regarded as a successful demonstration of E's ability to bias S's response.

Most of the early studies on EBE relied on the RPRT. Subsequent research by various workers was directed toward generalizing the situations by extending the range of tasks. Other researchers began to examine the manner in which E communicates his expectations to S, and the role of the participants. A third line of subsequent research examined the personality variables involved. These groups of studies are respectively reviewed later. In each category, studies reporting positive results (EBE demonstrated) are
followed by those reporting ambiguous findings, followed by those reporting negative results.

Task Variables

A number of substitutes for the photo rating test have been used. They include: animal conditioning, verbal conditioning, word association, simple motor task, judgement of visual stimuli, affective responses, psychological testing, real life situations.

Animal Conditioning

The EBE has been demonstrated using various animals. Rosenthal and Fode (1963a) asked undergraduate students to run rats through a simple maze. Six students were told that their rats were "bright" and should show rapid learning and the remaining six were told that their rats were "dull" and should show "very little evidence of learning". The rats were actually drawn at random from a homogeneous animal colony. An undergraduate research assistant who had worked for almost a year on a research program on EBE served as a control. She was consciously motivated to get as good a performance from her animals as possible. A t-test was performed on the mean number of correct responses per S for the six Es who believed they were running bright Ss, for the six Es who believed they were running dull Ss and for the research assistant E who was aware that the Ss were neither bright nor dull, but who was trying to get good results from her Ss. The data showed a
significant difference ($p<.01$, one tailed), indicating the presence of EBE. However, with the exception of the first day results, the control group run by the assistant (who knew that Ss were just ordinary rats and not bright ones) showed a better performance than the "bright" and "dull" groups. Assuming the presence of EBE, one would expect a better performance from the rats considered bright, but the authors failed to explain this discrepancy.

A further investigation of EBE using animals was carried out by Cordaro and Ison (1963). They asked students in an introductory psychology course to record the number of "contractions" and "head twins" manifested by flatworms when they were exposed to a conditioned stimulus (a light). Five students were told that their flatworms had been conditioned and "will probably show a high response rate" and five were told that their flatworms had not been conditioned and that they "shouldn't expect too much from them". Again the study showed an experimenter bias effect ($p<.001$).

A modification of the Rosenthal and Fode (1963a) experiment was carried out by Ingraham and Harrington (1966). Twenty-seven naive Es (freshmen in general psychology) conditioned fictitiously-typed "bright" and "dull" rats for bar pressing. Six Es ran only "bright" Ss, six Es ran only "dull" Ss and 15 Es ran both types. All three groups showed "a non-significant tendency towards EBE".

The research with animals described earlier has been justifiably criticized by Barber and Silver (1968) who
claimed that these experiments achieved the expected results only because the students who served as Es lacked previous experience with animals. With no clear criteria for the type of behavior they were to score, the Es merely reported what they were expected to see, not what their Ss were actually doing.

It may be concluded that the studies on EBE using various animals on the whole have failed to establish this phenomenon conclusively. The study by Cordaro and Ison indicates that when undergraduate students without prior experience in running laboratory animals are given ambiguous criteria for judging animal responses, their judgements are likely to be strongly influenced by their instructor's statements regarding the particular animal response. Ingraham and Harrington (1966) did not obtain significant results, while Rosenthal and Fode (1963a) apparently did. Yet the fact that in the latter research the performance of the control group was better than that of the high expectancy group, suggests that EBE research using animals needs further clarification.

Verbal conditioning

In a theoretical study, Lerner (1970) suggested that in the presence of E's expectancy and of S's set, interaction was plausible in the area of verbal conditioning and more specifically in the semantic generalization of classically conditioned responses. Lerner explained that E had an expectancy about his S's behavior. E's own behavior
(e.g., postural gestures, voice inflection, widening of the eyes) was a discriminative stimulus cueing S how to respond. The S responded accordingly and was reinforced.

Sheehan (1969) led different groups of Es (10 senior undergraduate psychology students) to expect that verbal conditioning either could or could not occur without awareness. In a Taffel-type task both groups tested independent sets of 19 Ss who constructed 100 sentences beginning with one of six pronouns and containing a past tense verb. For experimental Ss, sentences beginning with "I" or "We" were reinforced by the word "good". When the conditioning had been completed each E interviewed each S to determine whether during conditioning S was aware of the contingency between his behavior and the reinforcement used by E. Conditioning effects were matched by the expectancies of the Es. Ss who were unaware of the correct contingency were conditioned only in the group tested by Es who were led to expect that particular result. Unaware Ss in the group tested by Es who were led to expect that conditioning could not occur without awareness showed no significant difference from control Ss (p>.10).

The results of Sheehan's study seem to confirm the occurrence of EBE in verbal conditioning, but further studies are required to determine the factors mediating this effect. It is possible that Es who were led to expect that conditioning could occur without awareness unwittingly classified Ss as unaware even though the Ss were aware of
the conditioning process. Since the inquiry procedure was an open one, Es from this group could have been satisfied with answers of unawareness without deeper probing, whereas Es from the other group could have easily continued to look for awareness and might have sometimes found it.

Another study of EBE in verbal conditioning using a Taffel-type task was made by Page (1971). Page used 25 male undergraduates as Es and 193 female undergraduates as Ss. In that study, 49 Ss were run by Es under no formally-induced outcome expectancy and 144 under either a positive or negative outcome expectancy. Each E ran all three kinds of Ss. A t-test of the difference between the mean conditioning scores in the positive condition group and in the negative condition group showed a significant difference (p<.05). There was no significant difference between the positive condition group and the control group with no bias expectancy. Page concluded that the holding of a negative outcome expectancy might be especially crucial in determining results in verbal conditioning research.

Since there was no difference between the positive expectancy group and the control group, it may be argued that this conclusion is not sufficiently substantiated. However, it is likely that Es in the control group positively conditioned their Ss on their own initiative, being aware that they were participating in a conditioning experiment, even though they were not biased to expect positive conditioning. To put it simply, they tried to do a "good job".
Within the framework of verbal conditioning, Rosenthal, Kohn, Greenfield and Carota (1966), as part of a larger investigation, studied the separate and combined effects on research findings of Es' expectancy of certain results and the desirability to E of those results. The purpose of their study was to learn whether Es' expectations and desires might be partial determinants of the results of verbal conditioning. The experimental task was a Taffel procedure. They told 10 male graduate student Es that verbal conditioning would take place and told nine that it would not. The Es administered a Taffel-type task to a total of 60 female secretarial students. Half the Es [sic] in each of these groups were led to believe that it would be desirable if their Ss showed conditioning whereas the other half were led to believe that it would be undesirable. Apparently, neither Es' expectancy nor the desirability of the conditioning data by itself reliably affected the magnitude of conditioning scores, but the congruence between expectancy and data desirability did make a substantial difference. Those Es who (a) both wanted and expected, and (b) neither wanted nor expected their Ss to show increased use of "I" and "We" pronouns elicited significant conditioning (p=.001). Those Es who (a) wanted but did not expect, and (b) expected but did not want increased use of "I" and "We" pronouns obtained no significant conditioning. The authors could not adequately explain their findings and suggested further studies in the area. In fact, the table
presented in their paper showed that the group which neither wanted nor expected their Ss to show increased use of "I" and "We" pronouns got somewhat higher results than the group that wanted and expected these results, but the difference was not significant. Since the biasing of Es for this experiment was by no means a simple process, one would expect the bias of the Es to have been measured. This, however, was not done in the experiment, and it may well be that Es were not biased as intended; this in turn may have contributed to the experimental results.

From the results given in that study it is evident that EBE opposite to Es' bias and desire was found in one of the two groups with significant results. It may therefore be concluded that the experiment cannot serve as clear-cut evidence for the presence of EBE in verbal conditioning.

Authors of some studies dealing with verbal conditioning reported that they did not succeed in establishing the EBE. Kennedy (1969) conducted a Taffel-type verbal conditioning task. Es were six graduate males and Ss were 26 male and 34 female undergraduate students. His instructions to Es created three different expectancies, viz., positive, neutral, negative. That is, two male graduate student Es were led to believe that positive findings would result from the experiment, two Es expected negative results and two Es were not given an outcome expectancy indoctrination. The biasing of Es was done not by simply telling them what outcomes they should expect but by exposing them
to one of three indoctrination programs prior to the experiment. The two positively biased Es were told that they were replicating a study (Taffel) which consistently produced an overall conditioning. Es were also provided with a contrived proposal for the study in question, which reinforced the positive expectancy. In addition each E conducted two practice sessions prior to actual experimentation; however, Ss in this case were accomplices who were instructed to provide data which unmistakably demonstrated that conditioning had occurred. The negative-biased-outcome indoctrination was identical in format but opposite in intent. No overt references or special materials relating to the anticipated outcome of the experiment were provided to the two Es assigned to the neutral expectancy condition. In the trials proper the performance of Ss indicated no differences attributable to the expectations of the Es. Kennedy speculated that lack of explicit information of the expected outcomes might have caused these results, but Zegers' (1968) study, to be given later, did not support this explanation.

Zegers (1968) carried out another study which used a Taffel-type form of verbal conditioning as the experimental task, and two accomplices to bias Es' expectation. The participants were females randomly divided into two equal groups of Es and Ss. Both Es and Ss were taken from a highly homogeneous population with respect to age, education and marital status. The effect of the given expectancy on
Es' bias was verified by asking the Es to rate their expectancies before the experiment and after running each accomplice. The resulting ratings showed a significant difference between the groups (p<.001) which indicated that the Es were indeed biased. However, the study failed to demonstrate EBE. Equal amounts of conditioning were found in all treatment groups. As will be shown later in this study, the lack of sufficient difference in status between Es and Ss may have accounted for the failure to elicit EBE.

Summarizing the studies on EBE in the area of verbal conditioning it is observed that although all the studies used the Taffel-type procedure, they did not reach similar conclusions regarding the presence of EBE. One explanation perhaps lies in the different ways in which Es and Ss perceive the experimental situation, as was suggested by Orne (1962). This point will be elaborated on page 62. Page (1971) demonstrated the phenomenon quite clearly, especially when Es expected negative results. Sheehan (1969) also demonstrated EBE, but it is still an open question whether his Es really changed their Ss' responses or whether his results were due to the different ways of scoring used by Es. Rosenthal et al. (1966) found EBE in one group and reverse EBE in the other. Kennedy (1969) and Zegers (1968) failed to demonstrate the phenomenon. It seems that even though verbal conditioning experiments are sometimes influenced by EBE, the phenomenon is not easy to elicit and more studies are needed in order to define the situations in which it is
likely to occur. It seems likely that a study on the personality traits of the participants may help to clarify this problem.

**Word association**

Silverman (1968) studied EBE on performance in a word association test. Ten graduate and senior undergraduate Es administered a word association test to introductory psychology Ss with the expectancy that one group of Ss would show longer response latencies in association time between words than the other group of Ss. Based on a significance level of $p<.15$ Silverman concluded that "Ss tended to conform to the expectation". He also concluded that experimental Ss showed the difference to a larger extent than Ss from a control group comprising an equivalent group of Ss run by an equivalent group of Es who had not been given any expectancy ($p<.02$). In the experiment, response latencies in the two conditions of the two Es who had elicited the largest differential scores were later measured from recorded tapes by an assistant blind to the experimental situation. Systematic scoring errors by Es in the direction of their expectation were detected. When expecting long latencies Es timed or improperly recorded the latencies as significantly longer than they actually had been. Due to the poor quality of the remaining tapes Silverman was not able to determine the effect of scoring errors on all his results, i.e., the extent of Es' manipulation of Ss' latencies.
Considering the foregoing and the significance level (p<.15) it is difficult to see how Silverman could have concluded that EBE did in fact occur. Moreover, the reported significant difference between the two experimental groups and the control group (p<.02) cannot be taken as meaningful since the response latencies of the control group were somewhat shorter than those of either one of the two experimental groups, and were not, as might be expected, somewhere in between the two. Thus, it is doubtful that Ss did in fact change their responses in accordance with Es' expectations, and it is more likely that the results merely reflected the scoring errors.

Johnson and Adair (1970) replicated Silverman's study to determine whether these Es would obtain biased results in the absence of opportunities for committing errors in observing and recording Ss' responses. The Es (six male and six female undergraduate social psychology students) tested 96 male and 96 female introductory psychology Ss. The word association task was conducted in an attempt to assess: (a) the effects of high and low levels of inducement of Es' expectancy, and (b) the magnitude of observer (or recorder) error and bias affecting Ss' responses when data were obtained by biased Es. A significant expectancy effect was observed in the predicted direction (p<.05). The hypothesis that systematic observer or recorder error would account for some but not all the expectancy effect failed to reach significance (p<.10). Even
though the effect was in the predicted direction, the null hypothesis (i.e., no expectancy effect independent of observer or recorder error) cannot be rejected. Yet, Johnson and Adair found support for Rosenthal's position in their data. The difference between the group expecting a long latency response and the group expecting a short one "approached significance at the .08 level". The authors assert that, in view of the "repeated observations of EBE in the past" (for which, however, no reference was given), an effect at this probability level is acceptable. It seems, however, that due to the controversial nature of the EBE phenomenon a more conservative attitude should be taken.

In a modified experiment Johnson and Adair (1972) compared the effect of automatic instructions and verbal instructions on the EBE. Two groups of Es with different expectancy conditions performed the previous experiment using pre-recorded instructions, the timing and recording, however, being manual. Two equivalent groups of Es gave the instructions verbally. The overall EBE only "approached statistical significance (p<.08)", but Johnson and Adair nevertheless concluded that EBE was created, only by incautiously accepting the trends of the two previous studies (Silverman 1968, Johnson and Adair 1970). The bias effect was mainly accounted for by male Es testing Ss under conditions of non-automated stimuli and by female Es testing Ss under automated conditions (p<.05). No significant EBE was found from the data recorded by an independent
observer \((p<.08)\) but "the pattern of the means was partly consistent, or at least not inconsistent, with patterns obtained by the data [sic] obtained from the Es". From this Johnson and Adair concluded that some "true" EBE occurred, i.e., the Es did in fact change their S's behavior. Again it appears that the conclusion is not sufficiently supported by the data obtained in their study.

In summary it appears that the positive assertions made by the investigators regarding the presence of the EBE in word association situations are mainly based on a deeply felt notion that the phenomenon does indeed exist. Yet from their published data it is evident that this cannot be verified beyond reasonable doubt. It may well be that not every E can influence his S's, and it is of some interest to find out what personality traits are involved in the process.

**Simple motoric tasks**

Johnson (1970) studied EBE on a simple motor-performance task which did not involve a judgmental or decision-making process. Ss dropped marbles through holes in a table top. Es were told that the higher the S's intelligence the more marbles would be dropped. The counting of the results was made by the principal investigator in the next room, who was blind to the experimental situation. The expectancy effect was significant \((p<.01)\) and supported the hypothesis that experimenter expectancy would affect S's responses. In that study, the Es did not score the
data so that bias due to errors in recording and scoring was controlled. The task was a simple one, not involving judgment or decision-making by Ss or Es, and as such it was more difficult to bias. The study was well planned and can provide clear-cut evidence to the presence of EBE. Dusek (1971) performed a similar study, using as Ss 54 boys and 72 girls from grades one and two. In Dusek's experiment the Es, 18 males, were biased to expect that either boys or girls would drop the marbles faster. The scores indicated a significant effect due to Es' bias for girls ($p < .05$) but not for boys.

The study appears to confirm Rosenthal's (1966) findings that female Ss are more susceptible to bias than male Ss even in the case of children. This phenomenon may be a reflection of the dominant status of males in our culture, yet it also brings us back to the dependence of EBE on personality traits which is the subject of the present study.

**Judgement of visual stimulus**

Wessler (1969) had his Es (11 senior and six graduate students) ask their Ss (18 males and 22 females from introductory sociology course) to perform the RPRT and also to judge the length of lines. Some Es were led to expect their Ss to rate the pictures high (+5) and also to overestimate the length of the lines, and some Es expected their Ss to rate the pictures low (-5) and to underestimate the length of the lines. Each E held only one kind of expectation
regarding his Ss. The mean results of the photo-rating test for the two types of Es were opposite to their induced expectancies. The mean scores of the line-judging were in the expected direction but failed to reach significance. An additional analysis was performed using only those Ss whose scorings were in the same direction as Es' expectations in the RPRT, i.e., those who were positively biased in this test. This group of Ss comprised 10 Ss run by Es expecting long-errors and 10 Ss run by Es expecting short-errors. An examination of the type of errors made by each group revealed that for line-judging, the ratio of long-errors to all errors was 30:32 for the long-error group, and 17:32 for the short-error group. Disproportionately more of the errors made by the long-error group were long errors ($p < .05$). The study showed quite clearly that not every E was able to elicit biased responses from every S. However, the study did not control personality effects which could have been achieved by asking each E to run Ss under both kinds of expectations. Es' expectancies were not obtained by the original design but only from ex-post facto analysis on those Ss who revealed susceptibility in the RPRT. Thus, it appears that the study can be construed as suggesting new hypotheses to be validated in further research, i.e., under what circumstances are Ss biased by their Es. It should be noted that in a study of this nature Es' scoring errors should be controlled. This, however, was not done in Wessler's study.
The possibility of creating EBE by errors in judging experimental results was suggested by Bell (1971) who designed an experiment on taking measurements of pupillary changes. Seventeen graduate students were asked to help making measurements of change in pupillary responses. In Treatment I they were told that E had caused a pupillary dilation on the second of two photographs they were measuring and that an independent judgement of the size of the dilation was needed. The second photograph had previously been measured by 14 of 15 independent judges to be 1 mm smaller (on the average) than the first photograph. The discrepancy of the graduate students' response in the direction of the induced expectancy was significant (p<.01). Treatment II attempted to introduce a need for accuracy in addition to a dilation expectancy. An additional 15 students measured the same two photographs. Again the discrepancy between the independent measurement and the "expectancy treatment" measurements was significant (p<.01).

Bell's study gives a very clear example of scoring errors as a result of Es' bias. However, it does not shed light on the behavior of Ss under biased Es.

Summarizing the EBE in judgement of visual stimuli, it may be concluded that the susceptibility of Ss was not demonstrated beyond reasonable doubt. Wessler's study suggests that the phenomenon does exist, yet more clarification is needed.
**Affective responses**

The Es of Zoble and Seeman (1970) had each S immerse his hands in a waterbucket until his hands were temperature adapted. Then S placed each hand in a different bucket, one similar in temperature to the original, the other quite different. Ss rated which of the two they found more pleasant. Some Es were told that Ss would rate a small variation in water temperature as more pleasant; other Es were told Ss would enjoy a large variation in water temperature. Ss behaved in accordance with the Es' expectations (p<.01).

No attempt to detect Es' scoring errors was reported in the study. Since it is known that scoring and recording errors in the direction of Es' expectations cannot be ruled out (Silverman 1968, Adair and Johnson 1970, 1973), Zoble and Seeman's assertion could be questioned. It is impossible to assess the effect of these errors on the results.

**Psychological testing**

The possibility of the tester's preconception about his client influencing the test results is of serious applied concern. Substantial research effort has been devoted to the relation between test results and Es' expectancies. The results, however, are not conclusive one way or the other.
(i) Intelligence tests: Hersh (1971) explored the influence of a referral agent on the testing situation. The Es (28 male and female students in a graduate testing course) were presented with fictitious teacher referral reports before administering the Stanford-Binet Intelligence Scale to each of 28 culturally disadvantaged boys and girls. Each E tested both a "positively referred" and a "negatively referred" child. Under positive referral conditions, testers obtained higher IQ scores and made more favorable recommendations than under negative referral conditions ($p < .05$).

It appears that Hersh failed to differentiate between testers' scores and testees' IQ: i.e., the scoring of testees' answers was not checked by an unbiased tester blind to the experimental condition. Therefore, it is not possible to ascertain whether the IQs of the testees were really different, or whether the biased testers scored the testees who were positively referred more "generously". In other words, even though the bias did affect the results a further study should have been made to define the mediator in this effect: the quality of Ss' answers or merely the scoring by biased Es.

Schroeder and Kleinsasser (1972) dealt with the verbal portion of the Wechsler Intelligence Scale for Children (WISC) which was administered by biased Es to 18 fourth-grade children having a normal range of intelligence. Odd or even items were given by graduate student Es who
were led to expect Ss to be either bright or dull. The other part of the WISC was given by Es with the opposite expectation. Total verbal IQ scores were significantly (p<.05, two tailed) affected by the Es' expectations. Information, Similarities, and Vocabulary subtests were particularly susceptible to bias. The authors controlled the personality variables in the study by testing each S by two Es holding opposite expectations. They reported that few scoring errors by Es were detected. The protocols of the experiments were scored by an experienced clinician who was blind to the experimental conditions. An overall t-test of these independent IQ scores was conducted and the bias effect remained (p<.07, two tailed). Considering that the statistical procedures were two tailed even though directionality might have been predicted, the results can still be considered significant.

It was found by Witmer et al. (1971) that children presented with verbal approval scored higher on WISC than those presented with verbal disapproval. It may be argued that the same process took place in Schroeder and Kleinsasser's (1972) study. On the other hand, anyone familiar with the WISC test knows that sometimes it is quite difficult to write down exactly every word uttered by the testee and quite often the tester condenses Ss responses, especially when they are ambiguous. The tester often probes more in order to evaluate the response and he may not be quite accurate in recording the whole process. It
appears that in order to achieve scorings which are really objective the session should be mechanically recorded and the recordings should then be evaluated by an objective tester. As the study stands, there is no doubt that EBE was established, yet the question still remains whether it was demonstrated through Es' recording or through the answers of the Ss.

An investigation of EBE in the routine administration of intelligence tests to 80 retarded residents of a mental hospital was performed by Lasky, Felice, Moyer, Buddington and Elliot (1973). The study was designed to examine two questions: (a) the effect of Es' prior knowledge of Ss' IQ on their scores in an intelligence test, and (b) the effects of candy and standard reinforcement on Peabody Picture Vocabulary Test (PPVT) responses. The three Es were told only about the second purpose of the study. They were told that they would be given initial PPVT scores in order to know on which item to begin testing. These initial scores were inflated by 15 points for half the Ss in each condition (standard reinforcement or candy reinforcement). It was found that EBE was manifested only under standard reinforcement (p<.05) and not when candy was the reinforcing agent. In a study on EBE one would expect that full details about the Es be provided, but none was given. Of particular importance is the attitude of Es towards the two kinds of reinforcement, which might have provided an explanation of the data. Also, there was no mention of an
attempt to look for scoring errors. Yet the paper asserted that EBE was created by the examiners.

Several studies were carried out to determine the extent to which the scoring of WISC and WAIS (Wechsler Adult Intelligence Scale) might be influenced by biased Es who were given a written protocol without really testing the S. Very recently, Babad, Mann and Mar-Hayim (1975) tried to differentiate between bias in administration and bias in scoring of WISC, thus controlling the effect of actual administration and investigating only the bias in scoring. They asked 18 graduate students to score a WISC of a fifth-grade child. The same record was given to all students with one of two cover sheets portraying either an under-achieving child or a high-achiever. Significant differences were found for Comprehension ($p<.005$) and the total verbal part of WISC ($p<.05$). Students who received a disadvantaged description scored the record lower. The study used the same "S" for all Es, was well controlled and can be used as evidence of E's bias in scoring WISC.

Egeland (1969) investigated the influence of prior information about a child's past intellectual knowledge and academic performance on WISC scoring in Comprehension, Similarities and Vocabulary. Forty-six graduate students were given an identical WISC protocol with ambiguous answers in each of the subtests mentioned. Expectancy was created by supplementary information provided with the tests, in which the child's IQ (130 or 80) as "measured" in a prior IQ test
was mentioned. No bias was induced in about one third of the scorers. Results showed significant differences between groups in the scoring of Comprehension ($p < .01$) and Similarities ($p < .01$) but not in Vocabulary. The difference between the high expectancy group and the no expectancy group was significant ($p < .05$), but no significant difference was found between the low expectancy group and the control group. This might lead to the conclusion that the scorers gave more credit to the "intelligent" child but did not discredit the "unintelligent" one.

The extent to which the scoring of WISC Vocabulary items might be influenced by the scorer's expectations was investigated by Simon (1969). He gave 72 introductory psychology students the task of scoring 20 answers for this subtest. In the "bright condition" half the scorers were informed that the responses were those of a child reading far above his age level, while in the "dull condition" the scorers were informed that the responses were those of a child reading far below his age level. It was found that the mean score assigned to the protocol by the Ss in the "bright condition" was higher than that assigned in the "dull condition" ($p < .05$). Simon asserted that the results of his study could be generalized to real-life situations. However, in real life, the scorers are usually experienced psychologists rather than introductory psychology students to whom the test was completely new. Therefore, his assertion must be questioned.
Sattler and Winget (1970) reported an experiment designed to investigate the effect of referral reports and examinees' intellectual levels on scoring of intelligence test responses. Accomplices memorized WAIS responses which contained ambiguous answers in the verbal section. They played the role of examinees in "live" test sessions. Four scripts were prepared: two "superior" (IQ 130) and two "average" (IQ 96). Each script contained one of two parallel lists of 12 ambiguous responses. Es were graduate students who had administered at least 26 individual intelligence tests prior to the experiment. The testing sessions were taped with full knowledge of the participants. The results showed that the effect of the intelligence factor was significant ($p<.01$) while the referral reports were not ($p<.10$). Tape-recordings were used to count the number of probing questions, and it was found that the difference between groups was not significant. The study shows that Es tend to give higher scores to their examinees on ambiguous questions when they believe that the examinees are very intelligent. One may conclude that when an examinee is demonstrating an exceptional ability the referral report loses some of its influence. More generally, the study suggests that in an ambiguous situation the halo effect of the S's other responses might be stronger than the effect of an induced expectancy.

Contrary to previous research results, Saunders and Vitro (1971) failed to find EBE in the context of testing.
They asked graduate students to administer a Binet test to 60 normal second and third grade children. The children were randomly assigned to groups labeled retarded or gifted. The E was told that the purpose of the test was to refer the children to appropriate special programs.

Saunders and Vitro only told the testers that the teachers labeled the children as either retarded or gifted. No written referrals were given to strengthen this bias. This might have led to a lower expectancy level as compared with the other experiments in which written referrals were given.

Similarly, Auffrey and Robertson (1972) paired unscored record forms of the WAIS and WISC with differing case histories and gave them to examiners with three levels of experience: experts (at least 2 years of professional experience), interns (graduate students after training who had administered at least 10 Wechsler tests), and novices (students from a graduate course in Wechsler tests). It was hypothesized that the nature of the case history (optimistic, pessimistic, or neutral) would bias the scoring of the tests. Results revealed no significant scoring differences due to pretest case histories.

The foregoing review of literature on EBE as related to intelligence testing is summarized in Table 1. It leaves the impression that E bias does affect the results of intelligence tests. This does not necessarily mean that Es succeeded in biasing Ss' responses. Indeed, in some of
<table>
<thead>
<tr>
<th>Author</th>
<th>Test</th>
<th>Method</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egeland 1969</td>
<td>WISC</td>
<td>scoring</td>
<td>Comprehension, Similarity - biased, Vocabulary - no effect of bias</td>
</tr>
<tr>
<td>Simon 1969</td>
<td>WISC</td>
<td>scoring</td>
<td>Verbal IQ biased -</td>
</tr>
<tr>
<td>Sattler &amp; Winget</td>
<td>WAIS</td>
<td>scoring</td>
<td>Ambiguous responses biased</td>
</tr>
<tr>
<td>Saunders &amp; Vitro</td>
<td>Binet</td>
<td>testing</td>
<td>No effect of bias</td>
</tr>
<tr>
<td>Hersh 1971</td>
<td>Binet</td>
<td>testing</td>
<td>IQ affected by bias</td>
</tr>
<tr>
<td>Auffrey &amp; Robertson 1972</td>
<td>WISC</td>
<td>scoring</td>
<td>No effect of bias</td>
</tr>
<tr>
<td>Schroeder &amp; Kleinsasser 1970</td>
<td>WISC</td>
<td>testing</td>
<td>Verbal IQ, Information, Similarity, Vocabulary - biased</td>
</tr>
<tr>
<td>Lasky et al. 1973</td>
<td>PPVI</td>
<td>testing</td>
<td>IQ affected under regular reinforcement</td>
</tr>
<tr>
<td>Babad et al. 1975</td>
<td>WISC</td>
<td>scoring</td>
<td>Verbal IQ, Comprehension-biased</td>
</tr>
</tbody>
</table>
the experiments EBE was, or can be attributed to, scoring errors by biased Es. Only Schroeder and Kleinsasser (1972) made the distinction between Ss' responses and Es' scorings and yet they found EBE to be significant.

In most studies students were employed as testers and scorers and, except in Simon's (1969) study, they had at least some experience in administering the tests. The only study which employed professional psychologists as scorers failed to elicit EBE. Regrettably, this finding cannot be generalized to real life situations since, in that particular experiment, student scorers also failed to show EBE.

(ii) Projective tests All the investigators used the Rorschach test in their experiments. Marwit (1969) employed 20 clinical psychology graduate students as Es who administered Rorschachs to 40 undergraduates enrolled in introductory psychology. Ten Es expected Ss to give high total number of responses and high number of "animal" in relation to "human" responses, and ten Es expected Ss to give a low number of responses, most of them "human". In addition, each E, expecting "nothing remarkable", tested a control S. Generally, the results indicated tester bias. More specifically, in examining the productivity of the S, the results were in the expected direction but the difference was significant only for the low expectancy ($p < .05$). "Animal":"Human" ratios were significant for those who expected more "animal" responses ($p < .05$). For those who
expected more "human" responses the results were in the ex-
pected direction but failed to reach significance.

Marwit's study seems to have some weakness in de-
sign and in the interpretation of the data. The design did
not allow for testing every "kind" of S by every E. When
dealing with personal influence this variable should be
controlled. The scorings of Rorschach psychograms were
rated by two doctoral candidates. For validation purposes
some scorings were done by Marwit himself. Presumably, all
three of them were not blind to the purpose of the experi-
ment and probably even to the experimental conditions;
Marwit did not mention that they were. If the scorers were
not really blind, they might have increased the number of
the responses in one group and reduced them in another
since it is sometimes quite difficult to decide whether an
answer is a long combined one or a series of short
answers. Regarding the interpretation of the data obtained,
Marwit did not analyze the timing of the testing, even
though this might have been an intermediate channel through
which E's bias was communicated. When Es did not expect
many responses they simply removed the cards after a short
period, possibly without allowing the Ss sufficient time to
answer, and thus obtained significantly fewer responses.
Since there is a limit to the answers an ordinary testee
can give, the group expecting many responses did not differ
significantly from the control group. It is known that
"animal" responses are much more common than "human" responses, and the latter do not increase proportionally with increasing number of responses. Thus the larger the number of responses given by S, the larger is the number of "animal" responses likely to be given relative to "human" responses. This ratio was found to be significant in the group expecting "many responses and many animal", but not in the group expecting "few responses, most of them human". In short, it appears that the study demonstrated the EBE in those parts of the test which were easier to be manipulated by E, namely the total number of responses and the number of "animal" responses.

In a very recent experiment Marwit and Strauss (1975) tested the assumption that in a Rorschach test E influenced S during the instruction period. To test this assumption they used videotapes of four Es who had elicited the largest bias in an earlier experiment (Marwit 1969). Two Es expected "many answers, and a lot of animal responses" and the other two expected "few answers, mostly human ones". In the 1975 study the instructions were given to some Ss by means of the aforementioned videotapes, and to other Ss by means of tape recordings. After these instructions, Es tested Ss in the usual way, i.e., handling the cards and recording the answers. Es were senior psychology majors planning to attend graduate schools and Ss were 96 undergraduate females. No expectation was given to Es. Results showed no EBE. Marwit and Strauss concluded that
audio or audio-visual cues given to Ss during the instruction period were not sufficient to elicit EBE in a Rorschach test situation.

In explaining the results it may be argued that Es, not being biased, did not expect particular differences among Ss and therefore spent approximately the same time on each card; therefore no EBE occurred. However, no data on the time spent were presented. Apparently, stronger cues than those given in the experiment are required to elicit the EBE.

In contrast to Marwit's (1969) experiment, Jacob (1971), as part of a larger study, did not obtain any EBE in terms of the number of Rorschach responses. Jacob told half of his Es that they should expect 50 responses per S and the other half that they should expect 15 responses per S. After E tested all his Ss, he completed a short questionnaire which was designed to determine whether or not E received the biased expectancy. The answers to the questionnaire showed that 55 out of 72 Es reported some degree of suspicion regarding the real purpose of the experiment. Jacob did not rule out that this was the reason for the type of data obtained, although he was somewhat skeptical about the EBE phenomenon. A weak point in the design of the study seems to be that every E tested Ss only under one bias expectation.

Strauss (1968), in a well-controlled study, investigated the EBE on Rorschach responses regarding movement
and color percepts. E's were five female graduate students who tested 30 female undergraduate S's. Each E was led to expect two of her S's to give movement-dominated responses, two of her S's to give color-dominated responses. Two S's were tested without previous information about them and served as a control group. Sampling was restricted to females in order to avoid sex interaction. Before the test, E's predicted their S's psychogram in order to check whether they were really biased. All test sessions were recorded and checked. The results showed that E's were really biased ($p<.001$) but no significant EBE was found. It should be noted that the components of the Rorschach that were selected by Strauss, namely movement and color, are regarded as very essential in reflecting the personality traits of the testee. Apparently it is not easy to change S's responses when the responses are of this nature.

Summarizing EBE in projective tests, or rather more accurately in the Rorschach test, one may conclude that it is quite difficult to bias S's responses when using more essential components of the test like movement, color and "human". It seems easier to bias S's responses in the more common components of the test such as obtaining fewer responses or obtaining a large number of "animal" responses.

Real life situations

The leading study in a real life situation is that of Rosenthal and Jacobson (1966) who experimentally created
teachers' expectations, which changed students' performance. The experiment was conducted in a public elementary school in which most of the children came from lower-class homes, one sixth of them Mexican children. All the children in the school (18 classes of 6 grades) were included. The bias effect was achieved during the school year.

At the beginning of the school year, teachers were told to expect intellectual growth from about 20% of their students. Even though these potential "bloomers" had been randomly selected, they showed greater IQ gains as measured by Flanagan's Tests of General Ability (TOGA) than did the control children. The effects of teachers' expectancies were not uniform across the six grade levels. The two lower grade levels showed significant effect for grade One ($p < .002$) and grade Two ($p < .02$). In the remaining grades the "potential bloomers" did not show significant differences of IQ gained compared with the other children in the class.

Rosenthal and Jacobson's heuristic study evoked some criticism. Grieger (1971) asserted that the TOGA had relatively low correlations with other intelligence tests, and that its norms for young and low socio-economic children were inadequate. Apparently the majority of the teachers reported that they could not remember the names of the "bloomers". Rosenthal (1972) rejoined that the fact that the teachers did not recall the "bloomers" names did not mean that they never knew them, or that they did not
treat the children differently according to the bias given to them. However, this statement cannot be substantiated due to lack of data, and in fact one can plausibly argue from the opposite standpoint.

Meichenbaum, Bowers and Ross (1969) modified Rosenthal and Jacobson's (1966) study. The study was conducted on 14 female adolescents who were institutionalized in a training school. Six randomly selected girls were identified to their teachers as "potential intellectual bloomers". The expectancy induction was achieved through a discussion between teachers and the chief psychologists of the school about the "result" of the "test of late blooming" and other indications of probable intellectual growth. The effect was examined after one month. Since the elapsed time was short, no direct change in IQ was expected. Instead, the girls' academic performance and classroom behavior was measured. After a month, the "potential intellectual bloomers" showed significant improvement in course marks \( p < .025 \) which involved "objective" tests (mathematics, science and business practice). No significant improvement was found in marks given on the basis of essays and class participation (literature, history and english). All tests were given by the teachers as part of the school routine. Meichenbaum et al. tried to explain this unexpected result by stating that the lack of significant changes in the subjective courses might reflect the teachers' attempt to counteract any bias effect towards the
potential "bloomers". During the time of the experiment all 14 girls improved their behavior in class but the "late-bloomers" behaved significantly more appropriately \( (p < .05) \). The class behavior was measured by two observers who were blind to the hypothesis. In that study the bias was created in a most effective way. Not only were the "bloomers'" names given to the teachers but a group discussion was carried out about the meaning of "blooming". "Blind" observers and "objective" tests detected changes in the "bloomers'" performance. The study can be taken as a very good demonstration of the presence of EBE in a classroom situation.

Beez (1971) investigated the influence of teachers' bias on pupils' academic performance and teachers' behavior. He blindly assigned 60 six-year-old children to either a "low ability" group or a "high ability" group, each group consisting of 15 boys and 15 girls. Beez then randomly assigned 60 graduate education students with varied teaching experience to work individually with a child. The teachers were given a psychological evaluation that either interpreted the identical data positively or negatively depending on the child's group membership. Each teacher attempted to teach the child as many symbols as possible in a 10 minute period. Observers recorded the time spent on each symbol, the number of symbols the child attempted and the number of times the teacher re-read a word and explained it. Finally, the child moved to another room where an
independent examiner tested for recall of the symbols. Results showed that teachers who expected favorable performance attempted to teach significantly more symbols (p<.001), and to spend less time on each symbol (p<.01). Children who were said to possess "high ability" acquired significantly more symbols (p<.001). Sixty-seven per cent of the teachers who taught "low ability" children thought the tasks were too difficult for the pupils, whereas only one teacher of the "high ability" pupils thought so.

Beez demonstrated the teacher expectancy effect: children who were expected to perform better did in fact do so. Moreover, they were given better opportunities by the teachers. The teachers attempted to teach them more and actually thought the tasks to be within the pupils' abilities.

Several provocative studies have shown the effect of teachers' naturally-occurring bias on their students. Palardy (1969) sent questionnaires to 63 first grade teachers which included an item designed to elicit beliefs about the probable reading success of boys. He then administered the reading section of the Stanford Achievement Test to the children. A significant interaction effect was found for sex and teacher belief (p<.05). Male pupils of the teachers who believed boys would not read as well as the girls scored significantly lower than did the girls of the same teachers. These boys also scored significantly lower than boys and girls who were taught by teachers who did not
believe that there was a difference between the sexes regarding their reading ability.

Palardy did not manipulate the teachers participating in his study and so merely measured an existing phenomenon. One may argue that the teachers did not affect the children's performance but only formed their opinions about the reading ability according to the actual performance of the children in their class.

Support for Palardy's conclusion may be found in Seaver's (1973) study. Seaver obtained first-grade achievement scores for 79 younger siblings who had been preceded in school by bright or dull older siblings. It was hypothesized that pupils taught by the same teacher as their older siblings (expectancy condition) would perform better than those taught by a different teacher (controls) if their older siblings had been good students, and worse than the controls if their older siblings had performed poorly. The sample consisted of 79 pairs of siblings selected from the enrollment of two elementary schools and the associated junior high school. The schools were located in a high socio-economic status suburb and were clearly above the average in many respects (including IQ). Analyses of variance indicated a significant (p<.05) interaction between expectancy condition and older sibling performance on three achievement subtests: Word Meaning, Paragraph Meaning and Arithmetic.

Seaver's study was performed on data collected from
the permanent record files of the school; neither the teachers nor the pupils were manipulated. The study was done in a natural situation without any outside intervention, and therefore its conclusion may be generalized to first grade population.

EBE in different circumstances was created by Herrell (1971). A naive "guest lecturer" gave an identical brief talk to two introductory psychology classes. One class had been told he was "warm" and one had been told he was "cold". The two talks were recorded and the initial and terminal segments of his talk were played for 41 raters, who rated his talk along three dimensions: warm-cold, tense-relaxed and good teacher-poor teacher. Each one of them was presented in a seven-point Likert scale format. The terminal segment of his "cold" talk was rated as colder ($p < .001$), more tense ($p < .01$) and less competent ($p < .001$), than the initial and the terminal segments of his "warm" talk, although all the ratings were on the positive side of the scale. Note that these results were obtained by means of a t-test applied to ordinal data, i.e., Likert type scales.

An attempt to investigate the behavior and personality characteristics of the children who gained in IQ as a result of their teachers' expectations was made by Conn, Edwards, Rosenthal and Crowne (1968). In order to evaluate the long term effects of the teachers' expectations, the IQ of the children was measured again one year after
the teachers stopped instructing them. The study was conducted in six grades of an upper-middle-class suburban elementary school. The procedure was a replication of Rosenthal and Jacobson's (1966). In addition, the children were given a test in which they were required to identify vocal expression of emotions recorded by male and female voices. The IQ (TOGA) scores four months later showed that the experimental group gained in IQ scores more than did the control group. The higher differences were in verbal IQ ($p<.10$, two tailed), in total IQ ($p<.20$, two tailed) for boys, and in reasoning IQ ($p<.20$, two tailed) for girls. Conn et al. regarded those as significant. The second re-test made a year after the first one showed that in reasoning IQ and total IQ boys from the control group showed higher gains than did boys from the experimental group. The control group of girls showed higher scores in verbal IQ and total IQ, but no $p$ values were given. Regarding the children's ability to perceive vocal expression of emotion, girls who scored high on the scale assessing perception of emotion expressed by a female voice "gained more" in verbal IQ than low scorers ($p<.07$). Boys scoring high on the same scale gained more than those scoring low ($p<.05$). Boys scoring high on the scale assessing perception of emotion in the male speaker profited less than low scorers ($p<.05$). A comparison with the scores obtained in the IQ test conducted a year later showed opposite results for boys ($p<.01$). Conn et al. stated that male Ss, low in ability to
perceive vocally expressed emotion by males, showed personality characteristics associated with dependency. Since a significant positive relationship between dependency and susceptibility was established by several investigations, the IQ gain resulting from the expectancy is consistent with the literature. Regarding the opposite results obtained in the later IQ test, the authors suggested that male Ss showing a high ability to perceive emotions in males were usually more self-confident and more mature so that their gain from the expectancy was more "internalized" and "longer lasting". It is difficult to accept the latter argument since it suggests that the effects of bias after a long period of no reinforcement are more pronounced than during reinforcement. Also, the term "long lasting" is unfortunate since it implies that the effect was present at the time when the first test results showed otherwise. Since their conclusions are based on an analysis of variance performed on non-parametric data (Osgood scale), the significance levels obtained are of questionable value. Even if it is assumed that the p-values were correctly derived, the level of significance reported for the IQ gain in the experimental group is usually not considered to be sufficient. Therefore it seems that the study cannot be considered as a successful demonstration of the effect of teachers' expectations on pupils' gains in IQ.

The effect of teachers' expectations upon pupils' creativity performance and IQ in a black inner-city school
was recently examined by Rosenthal, Baraty and Clay (1974). Approximately one fifth of the school children were designated to their teachers as showing unusual potential for gain in creativity. Eight months later only the "high potential" children from the fifth grade (but not from the school as a whole) showed significant effects of having been "expected to bloom" ($p < .02$). Those particular children also showed significant gains in IQ scores as measured by TOGA ($p < .02$). The bias of the teachers was created by representing an IQ test (TOGA) as a measurement of creative potential. Before and after the bias was created, the children were given the TOGA, and a sheet of paper on which they were asked to draw a picture of a person on one side, and to draw as many different things as possible on the other side of the sheet. The pre-test was done at the beginning of the school year and the post-test at its end. Creativity scores were computed from the drawings by asking a group of four black and four white professional artists to rate the degree of creativity shown by the drawings. Rosenthal et al. could not explain why only one grade showed the expected results, and they were not able to specify just what were the factors that led to those specific results. They stated that about one-third of the reported studies showed a significant ($p < .05$) EBE. It should be noted that the scores obtained by judging children's drawings formed an ordinal scale and therefore the t-test used to differentiate
between the experimental and the control groups does not seem appropriate. Rosenthal et al. did not ascertain whether the teachers really believed in the expectancy given to them nor whether the teachers of the fifth grade felt differently about the bias than did the other teachers. This brings us again to the role played by the E's personality in the EBE: if the fifth grade teacher was indeed biased and the others were not, then certain personality traits should be sought to account for this difference. Similarly if all teachers were biased, but only the fifth grade teacher was effective in eliciting the bias, then perhaps some other personality traits were involved.

Clairborn (1969) replicated Rosenthal and Jacobson's (1966) study (page 35 in this thesis) in 12 first-grade classrooms. The study was done in three schools from two predominantly middle-class suburbs. Clairborn gave each teacher a list of approximately 20 per cent of her pupils who could be expected to show "intellectual blooming" when in fact these pupils were picked without regard to intellectual potential. Two months later tests showed no relative gain in IQ for pupils who were the object of the expectancy bias.

Clairborn concluded that his data did not even suggest the presence of an expectancy effect. Although the study was a very close replication of Rosenthal and Jacobson's, two differences that might have contributed to the different results should be mentioned. (a) The time
interval between the expectancy induction and the measurement of its effect was only two months. This appears to be too short a period to affect the IQ, even if such a change could have been possible. (b) The experiment was presented to the teachers as a "part of a requirement for a graduate education course" by a doctoral candidate, whereas in Rosenthal and Jacobson's study the experiment was presented as being conducted by university professors of psychology. It may well be that the difference in status of the two sources of bias led to different levels of belief.

Clairborn checked at the end of the experiment on whether the teachers still remembered the names of the "bloomers" and, indeed, they did remember. He did not ascertain whether the teachers really believed in the "intellectual potential" of those "bloomers".

Fielder, Cohen and Feeney (1971) replicated Rosenthal and Jacobson's (1966) study in 36 elementary school classes. In fact, it was almost an exact replication with the exception that measurements were taken after one semester rather than after one year. No effect of teacher expectation was found. Again no attempt was made to ascertain that the teachers were indeed biased. Indeed, one may argue that one semester is too short a period to bring about changes in IQ. However, it is very difficult to change IQ scores by modifying teachers' expectations even after a long period of contact, and it may be recalled
that the validity of the two studies reporting such changes has been questioned.

One study examined the effect of experimentally induced teacher's expectancy on subsequent behavior of institutionalized severely retarded children, when the bias effect was the result of optimistic psychological reports to cottage parents (Soule, 1972). In that institution most of the activities involving the teaching-learning process were carried out in the cottages between cottage parent and the child. They were primarily concerned with the learning of self-help skills. The children chosen to participate in the study (both experimental and control groups) were above the average for their cottage placement. The groups were homogeneous regarding their chronological age, ability to dress, feed and toilet training, and the ability to communicate with others. The teacher's bias effect was produced by biased psychological reports verbally presented to the cottage parents. They were told that:

the psychology department has developed what they hoped to be a more adequate way to predict the future functioning level of residents below the education range,

and that:

the psychology department will follow up on these students to help determine if our method of evaluation is a valid one.

During the following six months, weekly visits were made to each unit to follow up on each individual in the experimental group. This served as a method of reminding cottage
parents of the predictions made for the children. An evaluation of the children made at the end of the six months period did not show any significant differences between experimental and control groups. Soule concluded that teacher's bias effect has not the strength which is popularly attributed to it.

There is no doubt that the study failed to demonstrate the effect of teacher expectations on severely retarded children. It may well be that teachers' expectations cannot create great changes in the performance of such children, but it seems that the way Soule chose to bias the teachers might have contributed to his results. The test which was supposed to "predict the future functioning level" was introduced as a test which needed validation in a real life situation. It is reasonable to assume that the cottage parents doubted the efficiency of the test and that they were therefore not really biased. No attempt was made in the study to clarify this point.

Gozali and Meyer (1970) failed to replicate Rosenthal and Jacobson's findings in 16 special classes for educable mentally retarded pupils averaging 11 years of age. The subjects were examined at the beginning and the end of the academic year on the Stanford Achievement Test. No significant results were obtained. Again, the authors did not attempt to verify whether the teachers did in fact believe and remember the expectations given to them. It appears, however, that it is difficult to modify the
achievement of mentally retarded children by changing their teachers' expectations. A clarification of this point would require a better-controlled study.

The main findings of the "real life" research are summarized in Table 2. In teacher-student situations EBE was manifested in cases where the performance to be modified was the students' acquisition of academic material rather than their IQ scores. This is to be expected since IQ is considered to be a stable measurement, free from achievement effects. Another common feature of the studies eliciting EBE is the use of relatively strong methods of expectancy induction - either natural or artificial (by outsider). The studies which failed to demonstrate EBE used weaker methods to induce teachers' expectancies, and the failure may be attributed to the insufficiency of these methods to induce the expectancy. The relationship between teacher and pupil is often very personal and their interaction extends over a long period of time. It is therefore reasonable that teacher's expectations affect children's performance within certain limits provided that the expectations are regarding academic performance and classroom behavior.

As to mentally retarded children, the two studies examining their behavior failed to show EBE. However, in those studies, it was not ascertained that the teachers were in fact biased by the expectation given to them. More study is required to clarify this important point.
<table>
<thead>
<tr>
<th>Author</th>
<th>Sample</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rosenthal et al. 1966</td>
<td>Elementary school</td>
<td>EBE; IQ by TOGA, grades one &amp; two</td>
</tr>
<tr>
<td>Beez 1971</td>
<td>Culturally deprived children</td>
<td>EBE; achievement test</td>
</tr>
<tr>
<td>Conn et al. 1968</td>
<td>Elementary school</td>
<td>EBE; IQ by TOGA</td>
</tr>
<tr>
<td>Meichenbaum et al. 1969</td>
<td>Girl offenders</td>
<td>EBE; &quot;objective tests&quot; &amp; in classroom behavior</td>
</tr>
<tr>
<td>Palardy 1969</td>
<td>First grade</td>
<td>EBE; reading achievement</td>
</tr>
<tr>
<td>Clairborn 1969</td>
<td>First grade</td>
<td>No EBE; IQ by TOGA</td>
</tr>
<tr>
<td>Gozali et al. 1970</td>
<td>Educable retarded</td>
<td>No EBE; by achievement test</td>
</tr>
<tr>
<td>Herrell 1971</td>
<td>Guest lecturer</td>
<td>EBE; by objective judges</td>
</tr>
<tr>
<td>Fielder et al. 1971</td>
<td>Elementary school</td>
<td>No EBE; by TOGA</td>
</tr>
<tr>
<td>Soule 1972</td>
<td>Severely retarded</td>
<td>No EBE; by IQ tests &amp; behavior measurements</td>
</tr>
<tr>
<td>Seaver 1973</td>
<td>First grade</td>
<td>EBE; by achievement test</td>
</tr>
<tr>
<td>Rosenthal et al. 1974</td>
<td>Elementary school</td>
<td>EBE; in TOGA &amp; Creativity only fifth grade</td>
</tr>
</tbody>
</table>
Summary of EBE in various tasks

From the foregoing review of the research involving different tasks it can be seen that the studies using animals, word associations and teachers' expectations about children's IQ failed to demonstrate EBE. Regarding verbal conditioning, psychological testing, simple motor tasks, and teachers' expectations about academic performances, EBE was created in many cases. Judgment of visual stimulus and affective responses are still subject to controversy. Not all the studies demonstrated that the effect is indeed created by $S$ modifying his own behavior in response to $E$'s influence, but Page (1971), Johnson (1970), Dusek (1971), Marwit (1969), Meichenbaum et al. (1969), Beez (1971), Palardy (1969) and Seaver (1973) did. It is therefore reasonable to assume that $E$ does influence $S$ to produce the expected reaction, and some research work has been directed into probing the manner in which this influence takes place. Since it has been shown that not every $E$ is capable of biasing, and not every $S$ is susceptible to EBE, investigators started looking for personality variables in $E$ and in $S$ to link them with EBE.

Methods of Communicating Experimenter's Expectations

One possibility of communicating EBE is that $E$ may influence his $S$s to give expected-desired responses through unintentional paralinguistic and kinesic cues. Rosenthal and Fode (1963b) provided evidence that EBE can be mediated
by unintentional verbal and visual cues. The study involved four groups of student Es. Es in Group One (a bias replica) were told that they would earn more money if they obtained low (-5) rating from their subjects on the person perception task. Es in Groups Two, Three and Four were told that they would earn more money if they obtained high (+5) ratings. The latter three groups differed in the following: Es in Group Two sat behind a screen when they orally administered the task to their Ss; Es in Group Three were visible, presented all instructions in writing, and did not speak to their subjects; and Es in Group Four ran the Ss in the usual way presenting instructions orally and in full view of the Ss (bias replica).

The major findings were as follows: (a) Group Two (nonvisible and expecting +5) elicited significantly higher ratings ($p<0.005$) than Group One (bias replica, expecting -5) and significantly lower ratings ($p<0.02$) than Group Four (bias replica, expecting +5). These outcomes indicate that visual cues play a role but are not necessary in transmitting E's expectancies. (b) The ratings elicited by Group Three (nonverbal and expecting +5) did not differ significantly from those elicited by Group One (expecting -5), and indeed were nearly identical. This result indicates that in order to elicit EBE it might be necessary for Es to speak to their Ss.

One should take into consideration that Rosenthal and Fode obtained the statistical significance using
analysis of variance and t-tests, which are both parametric tests, whereas rating the pictures according to degree of success is obviously ordinal. The finding that the Es who were nonverbal and expecting "+5" obtained nearly identical results to those who were verbal, in full view and expecting "-5" is somewhat puzzling. The investigators did not deal with this result, but it appears that it is more difficult to obtain biased negative results (rating of failure) than biased positive results (rating of success). The tabulated results of the mean ratings obtained by the Es from all the groups support this argument: out of 28 mean scores only three were negative and all higher than -1.0. Again it should be noted that no attempt to examine the effect of the bias given to Es on the Es themselves, and to detect scoring error was reported.

McFall and Schenken's (1970) modified Rosenthal's photo-rating task so that Ss were given one of two tape-recorded E's instructions. The recordings were of one male E giving the same instructions, but in one case he expected generally positive rating, in the other, negative rating. Ss scoring high on need for achievement scale and those scoring as field dependent (on the concealed figures test) showed significant EBE, but an analysis of the ratings failed to show an overall expectancy effect when the personality variables were not taken into account. More details on McFall and Schenken's (1970) study are given subsequently.
The effect of differences in emphasis in reading the instructions to Ss was studied by several investigators. Duncan, Milton, Rosenberg and Finkelstein (1969) found differential emphasis scores, indicating the extent to which each experimenter had emphasized one side of the RPRT rating scale (+ or -) over the other side. All Ss who heard instructions that put greater emphasis on the rating alternative associated with success, subsequently rated the pictured persons as having had more success than the Ss who heard emphasis on failure (p=.004).

In that study no expectancy was given to the Es. The taped recordings of the instructions had been gathered by requesting some of the authors' male colleagues and students to give readings of Rosenthal's instructions first in an "objective and balanced" manner, and then in "slightly shaded" manner in either positive or negative directions. The authors chose nine tapes of three Es in which the differences in emphasis in reading the instructions were the greatest. The study is a clear demonstration of how emphasis on different phrases in the instructions serves as a cue for the S on the kind of responses he is expected to give.

Following Duncan et al. (1969), Marwit and Strauss (1975) tried to create EBE in Rorschach tests using recorded instruction by biased Es. A more detailed description of the experiment is given on page 33 of this thesis. The recorded instructions failed to elicit EBE. It should be
noted that in the Rorschach test the standard instruction to Ss is: "I am going to show you a series of 10 cards, what you are to do is tell me everything you see." Thus, in contrast to RPRT instructions, cues cannot be added and it is practically impossible to make an expectancy-loaded allusion by means of variation in intonation. It follows that no EBE would be created from merely listening to Rorschach instructions, in fact it would be surprising if it were otherwise.

In contrast to studies with RPRT reporting that EBE is mediated through paralinguistic cues, Hertzog and Walker (1973) could not obtain a significant degree of EBE while using tape-recorded instructions. They divided their Es into groups of two for the purpose of recording the instructions for RPRT, communicated to them the bias expectancy and engaged them in a competition for a first prize of $3.00 and a consolation prize of $1.50:

The money was awarded on the basis of a subjective appraisal of the probable effectiveness with which the tape would accomplish the purpose for which it was intended.

Hertzog and Walker questioned the efficacy with which auditory cues were able to elicit a significant EBE. However, they did not consider one major difference between their study and those which successfully demonstrated the EBE using only verbal cues. In Hertzog and Walker's study, Es never actually tested live Ss; after Es were biased they merely read the instruction to a tape recorder. It might be suggested that in order to elicit the EBE,
i.e. in order to bias $S$, $E$ should feel the presence of the $S$ he is trying to bias. When $S$ is not present the $E$ should be told the manner in which the instructions should be read, as was done by Duncan et al. (1969). Hertzog and Walker did not mention any attempt to check on the effectiveness of the bias they gave to their $Es$, so that the failure to elicit EBE might have been due to the fact that no bias was created.

In another example of a study investigating para-linguistic cues, Rosenthal (1967) analysed films taken during EBE experiments and found that male $Es$ had a more friendly interaction with their $Ss$ (male and female) than female $Es$ ($p < .05$). He also found that female $Ss$ evoked more smiling from their $Es$ - males and females ($p < .05$). Considering that male $Es$ succeed more in biasing their $Ss$ while female $Es$ find it difficult to bias male $Ss$, Rosenthal suggested that a condition for evoking EBE is a friendly relationship between the $E$ and his $Ss$. However, one cannot exclude another interpretation of these results. Considering the dominance of males in our culture, it may be argued that if EBE is to be obtained, $E$'s status cannot be inferior to that of $S$.

An investigation of $Ss$' perception of successfully biased $Es$ was performed by Rosenthal, Fode, Friedman and Vikan (1960). $Es$ were 12 male graduates and $Ss$ were 56 introductory psychology students, half of them females. The experimental task was RPRT in which, however, all 12 $Es$
were led to expect a +7 mean rating from their Ss. After the RPRT, the Ss completed a questionnaire designed to evaluate Es. To examine EBE, the data obtained by Es were compared with the results of an original RPRT standardization group, and with the results of a group that in an earlier experiment had been successfully biased to elicit low ratings from the Ss. Significant differences were found between the tested group and each of the two earlier groups (p < .001, for both). The results also showed that Es who biased their Ss more were rated by their Ss as more likeable (p < .10), more personal (p < .10) and more interested (p < .01). The authors considered the first two results as significant although only the third result reached the usually accepted level of p < .05.

Unfortunately, no reference is given to the earlier studies used so there is no way of evaluating them regarding the EBE obtained. As to the standardization group, no attempt was made to examine the validity of the results, and as shown on page 157 of the present work, even the results given by the same S on two different occasions are not necessarily the same for some of the pictures. Therefore, it appears that the only conclusion one can draw from the study of Rosenthal et al. is that those Es who elicited higher mean ratings were perceived by their Ss as more interested.

In a later study, Rosenthal, Persinger, Mulry, Vikan-Kline and Grothe (1964) asked 20 graduate male Es to
test 73 undergraduate Ss, 83 per cent (approximately) of them female, on RPRT. The results showed significant EBE ($p<.02$). The reported level was obtained, however, after excluding three Es (out of 20) who elicited ratings significantly opposite to the expectation given to them ($p<.05$). It is obvious that the reported significance levels cannot be accepted since they were not obtained from all the data. Rosenthal et al. themselves rated Es' behavior using films taken during the experimental sessions. Those results showed that Es who were rated as less professional elicited more positive ratings of the pictures ($p<.01$). One may question the objectivity of ratings made by those who are aware of the experimental conditions. It was not reported whether films of those Es who elicited "reverse" EBE were included in the analysis of Es' behavior.

A year later Friedman, Kurland and Rosenthal (1965) asked a group of observers, totally blind to the treatment conditions and experimental results, to rate the films taken during the experiment of Rosenthal et al. (1964). A random sample of 53 sessions with Ss was taken. A positive correlation was found between "successful" ratings of the pictures and (a) the number of glances E exchanged with S during the instruction period ($p<.01$) and (b) the duration of that period ($p<.05$). When the entire experiment was longer, Ss rated the pictures as more successful ($p<.01$). Another finding in this study was that S's responses were more biased when during the instruction period E exchanged
fewer glances with him ($p < .02$), read the instructions more rapidly ($p < .02$) and with fewer errors ($p < .02$). Es who were judged to be more dominant ($p < .001$) and more professional ($p < .01$) biased their Ss more.

Friedman (1967) obtained six significant correlations ($p < .05$) between the following variables and perceived success in the neutral photos: (a) the number of glances exchanged between the experimenter and the subject in the instruction period, (b) the number of times the experimenter glanced at the subject in the instruction period, (c) the longer the duration of the instruction period, (d) the longer the duration of the prerating period, (e) the longer the duration of the rating period, and (f) the longer the duration of the total experiment. Friedman suggested that the glances and the longer duration of E and S interaction made the Ss feel good because E was interested in them and they projected their good feelings on the pictures and judged the person as more successful.

Jones and Cooper (1971) investigated the effect of E's glances at S by instructing half of their Es to look at their Ss "at least 30 times" during the instruction period and by instructing the other half "not to look at the S while delivering the instructions". Eighty male high school students participated, 40 as Es and 40 as Ss. No expectancy about the results was given to the Es. After Jones instructed Es she left the room and observed the experiment through a one-way screen, counting the times that E glanced
Results showed that Es who were instructed to glance at their Ss did so 30.1 times on the average, while the others glanced only 0.15 times. Ratings of Ss in the frequent glancing condition were significantly higher ($p<.01$) than those in the other condition. The study clearly demonstrates the effect of E's glances on S's rating, although it is difficult to understand why Jones and Cooper, while designing a study to investigate EBE, did not take care that the investigator would be blind to the experimental condition. However, the difference between the number of times E glanced at S in the two glancing conditions was so large that counting errors by biased Es could not have affected the results significantly. Therefore, Jones and Cooper's conclusion can be accepted.

In summary, paralinguistic cues seem to be the medium through which E communicates his expectations to S. In the verbal instructions these cues probably take the form of different emphasis on particular words and phrases as the possible answers are being described by E. When no description of possible outcomes can be given, as in the Rorschach test, E cannot communicate his expectations through the instructions, so that other means have to be used. A positive rating is obtained by smiling at S, looking at him, and spending more time with him during the experiment. Es eliciting more positive ratings are also perceived as "more interested" by Ss. Es judged by blind observers as more dominant and more professional tend to
obtain higher EBE.

Role of Participants and Personality Variables

The role played by S in determining the results of psychological experiments has been examined by several investigators. Riecken (1962) stated that the Ss of psychological research, rather than being passive, compliant participants, are actively striving to discern the intent of the experiment so that they may maximize the rewards and positive evaluation they receive from the experimenter. Rosenberg (1965) similarly but more precisely specified the S's felt need to achieve positive evaluation from the E. He proposed that the typical human S approached psychological experiments with a preliminary expectation that the psychologist might undertake to evaluate his (S's) emotional adequacy or mental health. Even when S is convinced that his adjustment is not being directly studied he is likely to think that E is nevertheless bound to be sensitive to any behavior that bespeaks poor adjustment or immaturity. Rosenberg proposed that the experimental situations aroused within S an "anxiety-toned concern that he win a positive evaluation from the experimenter", which he labelled "evaluation apprehension". Rosenberg suggested that under the condition of high "evaluation apprehension" arousal, S's prime concern was with receiving positive evaluation from the experimenter, whom S perceived as assessing his maturity, intelligence, etc.
The implication of the "evaluation apprehension" phenomenon for experimenter-expectancy effects lies in considering the typical Rosenthal experimenter-expectancy experiment as a situation in which the proper or correct response is not easily discernable to S. The S is asked to rate pictures which are in fact neutral, i.e., so chosen as to evoke neither positive nor negative ratings from S. In this case he would attend to the subtle expectancy-indicating cues emitted by E. Minor (1970) suggested that the evaluative aspects of the psychological experiment might be the determinants of the S's reaction. He replicated Rosenthal's experimenter-expectancy effect findings, using Rosenthal's photo rating task. Some Ss were made to feel apprehensive or ego-involved in their performance, while the remaining Ss were assured that their performance would not be utilized to evaluate their functioning. The findings revealed that the expectation held by an E only led to confirmatory responses from Ss when the latter were personally concerned with their performances (p<.01).

While allowing that Ss are also interested in maintaining a positive self-image, Orne (1962) stated that much of S's interest in complying with the demand characteristics of experiments stems from his commitment to the advancement of science. According to this view, Ss will show more EBE when they consider the research to be scientifically important.

Adair (1972) seemed to find some support for Orne's
contention in a study purporting to measure dot estimation accuracy. Six graduate psychology student Es ran 87 undergraduate Ss. Es were led to believe that all Ss had previously scored either high or low on an inventory measuring the tendency to yield to social influence. In fact Ss were randomly selected. Results showed that female Ss who were aware of being deceived and nevertheless conformed to Es' expectations had a more positive attitude to psychological research (p<.05) than aware Ss who conformed less. The results for males, however, failed to reach significance (p<.07).

However, in a sharp contrast to Orne's assumption, Masling (1966) suggested that many Ss were negatively motivated in psychological experiments. He suggested that some Ss responded with a "screw you" attitude, i.e., with a desire to ruin the experiment, or at the least, to not cooperate. Similarly, Argyris (1968) equated S's behavior in resisting E with that of low-level employees who tried to "beat" the management.

It should be noted that neither the views of Riecken, Rosenberg and Orne, nor the views of Masling and Argyris necessarily embrace all the participants in psychological experiments. The attitude of many Ss towards psychological experiments may depend on the way they have been recruited. Masling and Argyris supported this point of view with examples of students who were "forced" to take part in an experiment to satisfy course requirements. They reported
the hostility expressed by those Ss who felt that they were being "used". Although many Ss serve in experiments as a course requirement, not all of them regard this as an injustice. Moreover, there are many Ss who volunteer to participate because they are genuinely interested, because they wish to do a personal favor to E, or because they feel that this is the right thing to do because of their positive attitude to psychological research. It seems that the hostility described by Masling and Argyris is not relevant to those Ss. The attitude of those Ss is more likely to be the "evaluation apprehension" of Rosenberg and Riecken and the "demand characteristics" of Orne. It appears that each of those views is valid for a certain type of Ss, depending on their attitude to the experiment and to the experimenter.

The recent focus of EBE research has been on the personality variables associated with this phenomenon. A substantial amount of research has examined the psychological aspects of the personalities of Es and Ss. The literature dealing with the effect of the following personality variables is reviewed: need for approval, anxiety, need for achievement, locus of control, field dependence, and dominance. Again, in each category studies reporting positive results are followed by those reporting ambiguous findings, followed by those reporting negative results.

**Need for social approval**

Smith and Flenning (1971) aroused S's approval motivation and thereby made her more susceptible to E's
influence by adding to RPRT instructions a single sentence designed to arouse S's motivation. This sentence occurred at the end of the instructions, when E (male) paused, looked at S (female) and said in a sincere and engaging tone:

"This is for my doctoral dissertation, and I'll appreciate it if you'll make the very best rating you can." Under these conditions, Ss high in need for approval, as measured by the Marlowe-Crowne Social Desirability Scale, demonstrated a significant susceptibility to EBE (p<.05) while low-approval-motivation Ss did not. These results were obtained by comparing the mean ratings of Ss who scored as having high need for approval in two groups: those tested by Es expecting "+5" ratings and those tested by Es expecting "-5" ratings. The comparison of Ss who scored low on the need for approval scale failed to reach significance. No overall EBE was found in the study. The two comparisons were done by means of one-tailed t-tests. The table given in that paper shows that the mean rating of Ss who scored as having high need for approval tested by Es expecting "-5" was the only one which was in the negative part of the rating scale. The mean rating of Ss scoring as low in need of approval tested by Es expecting "-5" was in the opposite direction to Es' expectations. No attempt to detect errors in Es' recordings was reported. One may assume that there was no difference in recording the scores when Es tested the two kinds of Ss, and errors, if any, probably affected both groups. Thus, the conclusion of Smith and Flenning
that Ss with high need for approval are more susceptible to EBE, may be accepted while noting that significance levels were obtained by means of a parametric test for ratings on an ordinal scale.

In a similar study Perlmutter (1972) used male Es and female Ss, half of each having a high and half a low need for social approval. The Es asked their Ss to undertake RPRT. (No scales for measuring social approval were indicated, and no significance levels were reported.) The results supported the hypothesis that Es and Ss need for social approval interact to affect expectancy outcomes. This was especially so for Es and Ss with high need for social approval.

Todd (1975) modified the procedure of RPRT for children. The children guessed how much each man won in a poker game on a scale of lost $10 to won $10. Ss were 24 boys and girls with above-average intelligence, ages ranging from 8-6 to 12-4, who had been given Nakamura's Hypothetical Situation Questionnaire (which consists of a number of subscales). Twelve boys and 12 girls who scored in the upper third of the distribution on "Interest in Social Recognition", in the lower third on "Interest in Task Performance", and in the lower third on "Self-Confidence", were labelled "Social Evaluation Oriented". Twelve boys and 12 girls who scored in the lower third on "Interest in Social Recognition", in the upper third on "Interest in Task Performance", and in the upper third on "Self-Confidence" were labelled "Task
Oriented". Three male undergraduates who served as Es were biased to expect high ratings from half of their Ss and low ratings from the others. After the data were collected Todd found that six of the pictures were given extreme ratings which were due to the stimulus and not to the expectancy. The ratings of these six pictures were eliminated. Results showed that those children who were labelled as "social evaluation oriented" demonstrated EBE whereas children who scored as "task oriented" did not ($p < .05$). Todd concluded that social evaluation oriented children produced responses consistent with E's expectation whereas task oriented Ss did not. The sub-scale measuring self-confidence which was used to assign the children to the two groups seems to have obscured Todd's conclusion. It is reasonable to assume that non-confident children would be looking for help from E when having to decide how to rate the pictures, and would respond to E's cues. The opposite seems to be the case with the self-confident children. Thus, the result obtained might be at least partly due to self-confidence. This problem could have been avoided by controlling Ss' self-confidence at the same level for both groups. Another weakness appears to be the need to eliminate six pictures from the raw data. One would expect the neutrality of RPRT to be verified prior to the experiment. Since such a test was not done, the reliability of all the pictures may be questioned.

Contradicting the foregoing results, Rosenthal,
Kohn and Greenfield (1966) found that Ss high in need for social approval (as measured on the Marlowe-Crowne Social Desirability Scale), arrived earlier at the site of the experiment ($p = .003$) and were less "aware" of the contingency ($p = .02$) but were no more likely to show conditioning in a Taffel-type task. Likewise, the need for social approval of Es was not related to their Ss' conditioning scores. The study is presented in detail elsewhere. No overall EBE was found in the study. A possible explanation for the failure to elicit EBE may lie in the rather complicated nature of the instructions, i.e., the Es may not have been affected by the bias. However, this is only a conjecture, since the effect on Es was not measured. This may also be the reason for the failure of Rosenthal et al. to relate EBE to the need for social approval.

Similarly, Bootzin (1971) could not relate S's or E's scores on the Marlowe-Crowne Social Desirability Scale to EBE as measured by RPRT. In that study Es were 20 male volunteer students from an undergraduate experimental psychology course. The rest of the class, 42 men and 77 women, served as Ss. No effects of the expectancy given to Es were found, but Es showed a significant effect of their own belief as to what Ss' rating should be. Es elicited ratings consistent with their guesses regarding Ss' average rating ($p < .05$). No significant correlation between the bias that was given to them and their own belief about Ss' ratings
was found. The latter correlation was taken as a measurement of Es' belief in the induced expectancy, i.e., how biased they really were. Bootzin concluded that they were not affected by the given expectancy and that therefore it did not affect the ratings of their Ss. However, their own belief about Ss' mean rating did cause a significant effect.

One possible explanation for Bootzin's failure to relate the need for social approval to EBE might be associated with the fact that Es and Ss were classmates, and as such have had some opportunity to gratify their need for social approval. In the usual situation E and S are strangers, and the need for social approval might become more dominant.

From the foregoing review of literature examining the relationship between EBE and the need for social approval it may be concluded that Ss scoring high on scales measuring need for social approval conform more to Es' expectations (Smith and Flenning 1971). Es' and Ss' need for social approval apparently interact to affect the expectancy outcomes, and thus high-need Es testing high-need Ss created more EBE (Perlmutter 1972). All the studies which successfully demonstrated the effect of need for social approval on EBE used RPRT as the experimental task.

The two studies which failed to establish a relationship between EBE and need for social approval also failed in establishing EBE. Rosenthal et al. (1969) used verbal
conditioning and a somewhat sophisticated biasing technique whose effect on Es was not measured. Bootzin (1971) used RPRT, but failed to establish EBE, apparently because his Es and Ss came from the same population. The relationship between EBE and the need for social approval in the case of children is not so clear-cut (Todd 1975).

Anxiety

Rosenthal (1966) concluded from the results of six studies, three of them unpublished, that E's level of anxiety was related to the occurrence of EBE. All studies used RPRT as the experimental task. A brief description of the three published studies follows. Rosenthal, Persinger and Fode (1962) used 10 male graduate students as Es and 56 undergraduates ("about equal numbers of each sex") as Ss. Es' level of anxiety was measured by the Pt scale* of Minnesota Multiphasic Personality Inventory (MMPI). The correlation between high-anxious Es and bias was found to be +.52 (p<.13). In Rosenthal, Greenfield, Persinger, Kohn and Corota (1965) Es were 26 senior undergraduates and Ss were 115 female undergraduates. Anxiety of Es was measured on Taylor's Manifest Anxiety Scale (MAS). Results showed that medium-anxious Es exerted the greatest expectancy effect at a significance level of p<.08. The study of Rosenthal, Persinger, Kline and Murly (1963) was different from the rest in that it tested vicarious EBE. Fourteen graduate students (three females) tested 76 introductory psychology students "about half of them females". Es'...

*Psychasthenia
anxiety levels were measured by the MAS. After administering the RPRT, every E was randomly assigned two advanced undergraduates as assistants (all but three were males). Es trained their assistants in administering the RPRT. Each assistant subsequently ran five to six Ss under his E's supervision. Ss were 154 introductory psychology students "about half of them females". The instructions to the assistants did not inform them about the kind of ratings they could expect from their Ss, and Es were warned not to tell them anything about the expected data. Also, Es were never explicitly told to expect their assistants to elicit from Ss the same ratings that the Es themselves had elicited from their own Ss. "It was subtly implied, however, that Es somehow knew the magnitude of the data that their assistants ought to obtain from their Ss". The results showed that assistants who were trained by those Es who had elicited EBE transmitted bias more than did assistants trained by Es who had not done so (p<.01). The assistants of the six Es who had obtained the most biased results elicited EBE (p<.05). More anxious Es showed a "non-significant tendency" to bias their Ss more than did less anxious Es (p<.08). The difference in EBE between more anxious assistants and other assistants was not significant (p<.30). When anxious Es and anxious assistants were combined the significance of the difference between them and others regarding EBE rose to p<.05.

Rosenthal (1966) did not give many details about
the following three unpublished studies relating anxiety level to EBE*. Fode (1965), using 16 Es and 167 Ss, found that Es who were medium anxious on the MAS exerted the most bias ($p<.001$). Rosenthal, Persinger, Murly, Vikan-Kline and Grothe (1962), with 29 Es and 86 Ss, concluded on the basis of $p<.08$, that medium anxious Es exerted the most bias. Persinger (1962), with 12 Es and 43 Ss found that low anxious Es exerted the most bias ($p<.05$). In the last two studies MAS or a "near relative" were used to rank anxiety levels.

In summary, EBE was shown to be significantly related to each of the three levels of anxiety: low level ($p<.05$), medium level ($p<.01$) and high level ($p<.05$). The results of the other three studies cannot be accepted as significant although $p<.08$ was so considered by the authors. These three results are to some extent contradictory, and it may be suggested that other personal attributes were also involved in the occurrence of EBE.

Since no detailed information on the three unpublished studies is available, it is difficult to make general observations on their design and analysis. Yet it is apparent that levels of anxiety were somewhat loosely defined in the whole series of the six studies. In the first place, more than one scale was used, namely, MAS and a "near relative", which was not identified in Rosenthal (1966).

*For references see Rosenthal (1966).
Secondly, in all experiments Es were ranked as high, medium, or low anxious if they fell into the upper, middle or lower third of their samples' distribution of anxiety score; therefore the definition of high, medium and low could have varied among studies.

The study of Rosenthal, Persinger, Kline and Murly (1963) deserves a more detailed review since it involved a vicarious situation, in which Es' expectancy was supposed to be transmitted to Ss via assistants. In that study elaborate measures were taken to prevent Es overtly biasing the assistants. In such cases one would expect an attempt to find out whether the bias was in fact transmitted. In the absence of any indication that the assistants were biased, the correlation between high-anxious Es and EBE \( p<.08 \) is the only meaningful result of that study regarding anxiety. This correlation, however, failed to reach the accepted significance level. It is evident that the relation between levels of anxiety and EBE is ambiguous.

The relation between anxiety levels and EBE in a Taffel-type situation was examined by Rosenthal, Kohn and Greenfield (1966) as part of a larger study. MAS was used to rank anxiety levels of Es. Based on a significance level of \( p<.08 \), the authors concluded that both high-anxious and low-anxious Es elicited higher conditioning than did medium-anxious Es. Apart from the fact that \( p<.08 \) is below the accepted significance level \( (p<.05) \), the EBE elicited in that study was not conclusive, as has been
suggested in more detail on page 11. It appears therefore that the study failed to establish a relationship between EBE and E's level of anxiety.

A short summary of the experimental results on the relationship between anxiety and EBE is presented in Table 3. It can be seen that no clear-cut conclusion can be drawn about this relationship, perhaps because, as has been observed, "anxiety" was not satisfactorily defined. Moreover, it is possible that another factor, not considered in the studies, could have been involved in the occurrence of EBE.

Need for achievement and motivation

The relationship between EBE and Ss' need for achievement was tested by McFall and Schenkein (1970). The study involved 48 female Ss who were tested by means of tape-recorded instructions, chosen from several recordings made during the administration of RPRT by eight male E's in a pilot study. In the pilot study each E tested from seven to 10 undergraduate Ss. The tape used in the experiment itself was of that E who had produced the most EBE, testing the two Ss who had been the most strongly biased. After taking RPRT, all Ss were tested on the need-for-achievement scale from the Edwards Personal Preference Schedule. Ss were assigned to a high or a low group if their scores on the scale fell above or below the median of the scores' distribution. The mean ratings of RPRT were compared by a t-test. Ss with high need for achievement showed EBE.
<table>
<thead>
<tr>
<th>Author</th>
<th>Task</th>
<th>Anxiety Scale</th>
<th>Results: Higher EBE by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persinger 1962</td>
<td>RPRT</td>
<td>MAS or &quot;near relative&quot;</td>
<td>Lower anxious E, ( p &lt; .05 )</td>
</tr>
<tr>
<td>Rosenthal, Persinger &amp; Fode 1962</td>
<td>RPRT</td>
<td>( p_t ) from MMPI</td>
<td>High anxious E, ( p &lt; .13^* )</td>
</tr>
<tr>
<td>Rosenthal, Persinger, Murly et al 1962</td>
<td>RPRT</td>
<td>MAS or &quot;near relative&quot;</td>
<td>Medium anxious E, ( p &lt; .08^* )</td>
</tr>
<tr>
<td>Rosenthal et al. 1963</td>
<td>RPRT</td>
<td>MAS</td>
<td>High anxious E, ( p &lt; .05 )</td>
</tr>
<tr>
<td>Fode 1965</td>
<td>RPRT</td>
<td>MAS</td>
<td>Medium anxious E, ( p &lt; .001 )</td>
</tr>
<tr>
<td>Rosenthal et al. 1965</td>
<td>RPRT</td>
<td>MAS</td>
<td>Medium anxious E, ( p &lt; .08^* )</td>
</tr>
<tr>
<td>Rosenthal et al. 1966</td>
<td>Verbal conditioning</td>
<td>MAS</td>
<td>High and low anxious E, ( p &lt; .08^* )</td>
</tr>
</tbody>
</table>

*Not significant
The study was well planned and was very well controlled, and thus it can serve as a clear demonstration of the effect of S's need for achievement on EBE. It should be noted, however, that Ss rated the pictures on an ordinal scale whereas the t-test is parametric.

A conclusion that can be drawn from the foregoing study, though it was not mentioned by the authors, is the need to investigate the kind of E who can bias his Ss. The study used the tapes of that E who had most successfully biased his Ss in the pilot study. It appears that in order to achieve EBE, an interaction between a special kind of E and a special kind of S is needed. This can be concluded from the fact that in the pilot study even the most successful E did not elicit significant differences between Ss' ratings under high and low expectation ($p < .10$), and that in the main study Ss low in need for achievement were not biased.

Hertzog and Walker (1973) expanded the work of McFall and Schenkein (1970). The details of their study were presented earlier (page 55). In the present context it is relevant that their sample consisted of four female and four male Es scoring high on the need to "avoid success" and four female and four male Es scoring low. Ss were 32 females and 32 males scoring high, and 32 females and 32 males scoring low on the need. The need was measured by Thematic Apperception Test (TAT)-type stories. Results failed to demonstrate EBE. It will be recalled that the
tape recordings used were obtained from Es who really were never in a testing situation (S was not present). This might explain the failure to obtain EBE. A more conventional way of testing is perhaps required to establish Es' need to "avoid success" on EBE.

In several studies attempts were made to manipulate the motivation of E and of S by influencing their ego involvement. Rosenthal and Fode (1963b) divided Es into moderate and high ego involvement groups. The high ego involvement Es were told they would be paid $5.00 an hour for a good job instead of $2.00. It was felt that sizeable difference in rate of pay might serve to ego-involve Es in the outcome of their experiments to a greater extent. In order to vary the motivation of the Ss, they were randomly assigned to a paid or an unpaid group. Results using RPRT showed no significant difference between the rating obtained by the higher and the lower paid Es, nor were there any differences between the ratings given by paid and unpaid Ss.

It may be argued that the manipulation employed by Rosenthal and Fode did not really achieve ego-involvement. Undoubtedly, the participants wanted to be paid as high as possible, yet it does not follow that they were personally involved, at least not in the sense of Rosenberg (1965). The participants did not feel that their personality or status would be endangered by any particular outcome of the experiment.

Minor (1970), in an attempt to replicate Rosenthal's
EBE findings, made some of the Ss feel apprehensive or ego-involved in their performance, while the remaining Ss were assured that their performance would not be utilized to evaluate their functioning. Members of the first group were told that the task they would be faced with would be judged by psychologists as indicative of a stable, mature and well-functioning personality. Minor wanted to make the Es feel personally responsible if they did not obtain the data they were led to expect. Therefore, the Es were told that if they followed the instructions and the proper procedure they would get the expected data. The finding revealed that the expectation held by Es only led to confirmatory responses from Ss when Ss were personally concerned with their performance (p<.01). Minor concluded that an overall EBE was demonstrated in his study (p<.05). His data showed that the mean photo rating of all Ss was "experiencing failure" but this was more the case where Es expected "failure" rating. The only sub-group yielding a "success" rating comprised Ss who were personally concerned about their performance when tested by Es who expected this rating.

Johnson (1973) investigated the effect on Ss of Es' intentional efforts at manipulating Ss' performance. Es were asked to attempt to communicate to their Ss the correct type of performance, i.e., increasing speed for Ss in one group and decreasing speed for Ss in the other group. The Es thus attempted to differentiate their treatment of high and low expectancy Ss without deviating from standard
instructions or without directly telling the Ss what was the correct type of performance. The Es obtained effects opposite to their expectancies (p<.02). An analysis of tape recordings of the experimental sessions revealed that Es who tried to manipulate their Ss spoke more than did other Es. Johnson tried to explain this "reverse EBE" in terms of the resistance of Ss who felt that they were being pressured to perform in a certain way.

Summarizing the relationship between need for achievement and between motivation and EBE, it appears that when Ss are motivated by high ego involvement, or when Ss are high in need for achievement, they are more susceptible to EBE. The studies which failed to establish this relationship either did not manipulate Ss' motivations (Rosenthal and Fode 1963) or they did not give E the opportunity to interact with S (Hertzog and Walker 1973). It appears, however, that E should not try to manipulate his Ss in an overt way because if he does, he is likely to achieve the opposite results. It should be noted that some motivation in E to achieve the expected results was taken for granted in all studies.

**Locus of control**

Felton (1971) examined the relationship between EBE and internal (self-controlled) vs. external (e.g., controlled by "fate") control for Es and Ss. Excluding the middle 20%, internal-external scores of 116 males were dichotomized. Three internal and three external Es (randomly selected from these Ss) each tested six internal and six
external Ss individually using RPRT. The results suggested that internal Es succeeded more in biasing their Ss than external Es (p<.025) and that external Ss were more easily biased than internal Ss (p<.005). No interaction effect for Es' locus of control and Ss' locus of control was found. The study was very well controlled, i.e., each E tested two kinds of Ss from his own sex only, and both E and S came from the same population (thus taking care of the effect of status). Again it is noted that the results were obtained by using a parametric test on ordinal data. Even though no attempt was made to detect errors in Es' recordings of Ss' responses, it seems that Felton's findings about the relation between EBE and the locus of control can be accepted.

Bootzin (1971), in a study described on page 68, failed to find a significant relationship between EBE and E's or S's locus of control. As mentioned before, his study failed to show an overall EBE. Ss and Es were not chosen according to their scores as externally or internally controlled but this scale was administered after the RPRT. The scores and the measure of difference between the internal and external Ss were not reported. It is therefore difficult to examine the reasons for Bootzin's failure to obtain significant results.

Field dependence

The dimension of field-dependence field-independence deals with an individual's responsivity to environmental
cues. A person with a field-dependent cognitive style tends to rely on external sources for definition of his situation, while a field-independent person tends to rely on himself.

As part of a larger study, McFall and Schenkein (1970) investigated the relation between EBE and field-dependence using RPRT in which recorded instructions were given. More details on the experimental design are given on page 74. The assignment of Ss to a field-dependent group or a field-independent group was based on their scores on the rod-and-frame and the concealed figure tests. Results analyzed by a t-test showed that only Ss scoring as field-dependent on the concealed figure test showed an expectancy effect ($p < .05$). Again it should not be overlooked that a parametric test was performed on ordinal data. The relation between EBE and field-dependence was not completely clarified by this work. Apparently the concealed figure test is not quite applicable to females who were the Ss in McFall and Schenkein's study.

**Dominance**

The relation between EBE and dominance was investigated by Bootzin (1971) as part of his larger study on the effect of personality variables using RPRT. He found that Es who were more dominant and confident in their ability to influence others were more likely to elicit ratings in agreement with the expectation given to them ($p < .05$). Bootzin tentatively suggested that Ss who were more
susceptible to EBE were more likely to be influenced by others (p<.10). In the study, the levels of confidence of the participants were measured by means of four nine-point self-rating scales. Es and Ss were asked to circle the percentage of people in the course who were: (a) better leaders than they, (b) easier influenced than they, (c) more submissive than they, and (d) whom they could influence more easily. The study failed to show an overall EBE, and Bootzin concluded that his findings should be considered as only exploratory. Considering the significance level of the finding that Ss who are more easily influenced are also easily biased, his conclusion seems acceptable.

The foregoing review of the relation of certain personality variables to EBE has shown that some variables are more likely than others to be conducive to the occurrence of EBE. When E has certain expectations as to the experimental outcome, and is motivated to fulfil them, or believes in his ability to influence others, or is internally controlled, he is more likely to elicit EBE. There may also be some relation between the EBE and: (a) E's level of anxiety, and (b) E's need for social approval, but these have not been satisfactorily established. Regarding Ss, it was found that S who has a high ego involvement in the experiment, a high need for achievement, a high need for social approval and who is externally controlled is likely to be more susceptible to EBE.
Summary of Literature Review on EBE

The foregoing review of the literature has shown that EBE was demonstrated in many studies using several experimental tasks. RPRT was the most popular task but successful results were also obtained in such diverse tasks as simple motor tasks (marble-dropping), psychological tests and academic achievement. Results of experiments involving judgement of visual stimuli and affective responses were less conclusive since none of the studies distinguished between S's responses and E's scoring. As to word association and animal conditioning, EBE, if demonstrated, was related to E's scoring errors and not to S's responses.

Barring extra-sensory perception, communication of E's expectation is presumably carried out by means of para-linguistic cues. Variation of intonation while delivering the instructions to achieve different emphasis might constitute the means of communication. However, not every E is capable of biasing every S. Ss who are more susceptible to EBE are those with high ego involvement, high need for achievement, high need for social approval, and those who are externally controlled. No clear-cut conclusions can be made regarding field dependent Ss. Regarding Es, it appears that all studies took their motivation to obtain the expected results for granted. Internally controlled Es and those who believe in their ability to influence others are apparently able to bias their Ss.

Certain limitations have been found in the studies
reviewed in this chapter: (a) Most authors did not investigate whether Es were in fact biased by the expectancy manipulation either in general or while running Ss. (b) In most studies Es were students serving under professors. It may be assumed that in such situations students would endeavor to obtain the results that are expected from them. However, most studies did not ascertain whether the student really cared about the experimental outcomes, i.e., whether in fact they tried to obtain the results. (c) In many studies Es recorded Ss' responses, and in most of them no attempt was made to ascertain whether EBE resulted merely from unintentional recording errors made by Es, or whether Es did in fact influence Ss' responses. (d) Most of the data, although ordinal, were analyzed with parametric tests. (e) In studies using RPRT no attempt was made to investigate the reliability of Ss' ratings. (f) In many RPRT studies Es tested under a single expectancy only, i.e. "high" or "low", and thus no control over personal variables was possible.

From the critical summary of the main experimental findings on EBE it is possible to identify certain relations between personality traits, situational variables and tasks which are associated with the creation of the EBE. A model describing this emerging pattern is introduced later in this study.

The personality traits reviewed earlier are not necessarily the only ones which underlie the EBE. It is
possible to identify other traits which, on theoretical grounds, are as likely to be involved. In the following section, one such trait, namely submission to authority, is discussed. This trait is the object of the experimental investigation undertaken in the present study.

Submission to Authority and its Relation to EBE

The foregoing review of research showed that EBE was demonstrated in situations where difference in status and sex existed between E and S. In most cases E is a graduate student where S is an undergraduate; E is usually male while S is often female. Although Rosenthal (1966) was not successful in showing the relation between status differences and EBE, his contention that E's status affects S's response seems logical. His failure may be due to the fact that the personality variables associated with status were not separately controlled.

Differences in status are likely to affect dominant and submissive persons in a different way. Submissive people, being in need of support and guidance, are likely to be more attentive to E's intonations and gestures, and thereby will be provided with cues as to his intentions and expectations. Regarding dominant Es, it is reasonable to assume that in order to dominate the situation they are likely to provide more discernible cues.
Dominance-submission

Block and Block (1952) focused on the interpersonal aspects of a formal experimental situation and suggested that the situation was almost invariably structured by the S in terms of a dominance-submission frame of reference. To S, E surely represents authority. E explains the situation, gives the instructions, remains aloof, and observes and records the S's behavior. In this setting it is not surprising that suggestions from E tend to be viewed as stemming from a person of authority and thus provide structure for S, who is in an uncertain situation. Differentiating authority and submitting to it may represent ways of avoiding an ambiguous environment by stabilizing it, since an ambiguous environment would otherwise involve endless decision-making and conflict.

Block and Block performed an experiment to examine this assumption. The sample consisted of 54 male students, relatively homogeneous with respect to age, intelligence and socio-economic status. E gave to S a spool-packing box and asked him to fill it. When the box was filled, it was emptied and S was asked to refill it and so on. Before S began to fill the box he had been told "When you don't want to do any more, you may stop". When S expressed his wish to stop the pointless task, E asked: "Don't you want to do some more?"

Ss who in previous tests showed submission to authority continued the task longest after they had wished
to stop ($p<.05$). It was reported that $E$ endeavored to ask the standard question with the same intonation for all $S$s. Since $E$ was not blind to the experimental situation, her success in concealing her own expectations might not have been complete.

The correlation between suggestibility and need for autonomy or, alternatively, between suggestibility and dependency, was investigated by Zuckerman and Grosz (1958), on a sample of student nurses. Sway's predictor test for hypnotizability was used as a measure of suggestibility, Edwards Personal Preference Schedule as a measure of autonomy and the TAT to reflect dependency. Those who scored low on Sway's test scored significantly high on the need-for-autonomy scale ($p<.01$). Those scoring high on Sway's test scored significantly higher on "dependency" as expressed by asking for help, sympathy or support in the TAT ($p<.05$). Zuckerman and Grosz stated that the high autonomy scores on Edward's scale expressed a desire "to be independent of others in making decisions, to feel free to do what one wants, and to avoid situations where one is expected to conform." They concluded that a person who was suggestible was likely to be a person of strong dependency needs, whereas a person who resisted suggestion was more likely to have stronger needs for independence or autonomy. We may conclude that dependent people are socially suggestible, and thus they are more likely to be biased by an $E$.

In a theoretical introduction to the presentation
experiments in which the task given was the Einstellung arithmetic problems. After testing hundreds of Ss in a friendly, easy-going atmosphere no significant relationship between authoritarianism and ability to solve the problem was found. This led him to compare Ss under two different atmospheres. In the first, E suggested to Ss that the tests to be taken were measures of intelligence and motivation, and that the results were of great importance. In the second, E described himself as a "psych. major" carrying out a class project in which he personally took very little interest. Ss in each atmosphere situation were drawn at random from freshmen enrolled in English. Results showed that the correlation between authoritarianism and Einstellung scores was significantly greater than zero for the ego-involved condition (\(p<.03\)), and this correlation was significantly greater than the comparable correlation for the non-involved group (\(p<.001\)). The performance of Ss who were submissive to authority was dependent on the atmosphere they worked in. Brown suggested that the Frenkel-Brunswik concept of "intolerance of ambiguity" as the genotype underlying the relationship between various perceptual and cognitive phenomena and submission to authority should be modified. It should read "intolerance of ambiguity as to the means of avoiding personal failure".

Authoritarianism, which subsumes within it attitudinal and behavioral predispositions toward persons in different status or power position, seems to be a relevant
variable of EBE. It is likely that Ss with different levels of conformity to authority will be affected to a different degree by E.

Most experiments on EBE were carried out by means of RPRT. When S has to rate neutral pictures, he is faced with an ambiguous situation which involves decision-making and conflict. It is likely that Ss who are more submissive to authority, i.e., place conformity above individuality, will turn to E, as a representative of authority, for cues as to what is required from them. Ss who are less submissive to authority, i.e., dominant Ss, will be more independent in their decision-making, will not fulfil expectations to conform, and will be less responsive to E's suggestions, i.e., less biased. This difference between submissive and dominant Ss, is likely to be manifested in a situation which involves a threatening ego-involving atmosphere, when performance is tied-in with personal success and failure. A dominant, non-conforming S will be more likely to resist E's suggestion to rate the pictures in a certain way.

Regarding E, it is reasonable to assume that a dominant E will be more successful in influencing S in the direction of his own bias. Support for this assumption may be found in the work of Friedman et al. (1965), who filmed Es conducting an experiment using RPRT. Analysis of the film made by five observers blind to the treatment conditions and to the experimental results showed that Es who were rated as more dominant biased their Ss more.
In this work it is hypothesized that a S who is submissive to authority will be more susceptible to EBE than a non-conforming S. It is also hypothesized that an E who is more dominant will obtain more bias from his Ss.

**Attitudes Toward Psychology**

When investigating interpersonal relations in a psychological experiment, the attitude of the participants to psychology and psychological experiments should also be considered. Adair and Fenton (1971) tested the assumption that the difference between the responses of Ss are due to differences in attitudes toward psychology and psychological research leading to a continuum of motivations to cooperate with E. The dependent variable was opinion-change. Although all Ss showed opinion-change to a significant degree, it was found that Ss with more positive attitudes toward psychology showed significantly greater opinion-change than Ss with less positive attitudes (p<.01).

In relation to EBE it may be assumed that Ss with a positive attitude to psychological research would be more susceptible to EBE than Ss with a negative attitude. The latter might even give results which are diametrically opposed to those predicted in order to "prove" the uselessness of psychological research. Also, an E with a positive attitude to psychological experiments might be more successful in inducing EBE. (An E with a negative attitude might bias S in the opposite direction, again to justify the negative attitude.) In order to control this variable it
is necessary to ensure that all Es and Ss have at least a non-negative attitude to psychological research.

**EBE Model**

In this section a model is presented relating Es' and Ss' attitudes and personality traits in the context of experimental tasks. At the outset two kinds of EBE should be distinguished. The first is the one created by observer errors, i.e., E errs in measuring, recording or scoring (when dealing with psychological tests) Ss' responses. The second kind of EBE, and perhaps a more fundamental one, is created when E, by some unintentional cues and personal attributes, sufficiently changes the experimental condition so as to produce changes in S's responses, i.e., creating changes in the dependent variable. It is difficult to assess the prevalence of EBE of the first kind due to lack of sufficient data in the literature. The model given later in this section deals with EBE of the second kind.

It was suggested by Johnson (1973) that EBE should be viewed as a two-person interaction where one person (E) communicates certain information to the other person (S). E knows the way in which S is "supposed" to perform. Only when E communicates this information, or part of it, to S and only when S acts upon this information, can the results be biased in the direction of E's expectations. If E is concerned with the outcome of the experiment he may be motivated, albeit unconsciously, to communicate hypothesis-
related information. Similarly, if $S$ is concerned about his own performance he may be set or motivated to "receive" and act upon such information.

It is clear from the literature review that EBE is associated with situational and personal variables. The situational variables are the experimental tasks and the kind of responses that are expected from $S$. The personal variables are $S$'s and $E$'s attitudes and personality traits.

Certain tasks are apparently less susceptible to EBE than others. It seems that when $S$ is asked to perform a task which is more fundamental to his personality, e.g. certain responses to projective tests and response latency of association, his responses are less likely to be biased.

Both $E$ and $S$ bring to the experimental situation certain attitudes and expectations. Some $S$s may have arrived naive but created their expectations at the very beginning of the interaction with $E$. When $S$ has a negative attitude to the experiment and/or experimenter, he most likely would not conform to the latter's expectations, and might even respond in the opposite direction. In order to respond to $E$'s expectations, $S$ presumably should have a cooperative attitude. This cooperation is taken for granted in most studies. A cooperative attitude, however, still does not ensure that $S$ will positively respond to the expectancy. Only when $S$ wishes to divine $E$'s covert expectation, succeeds in doing so, accepts it and wishes to conform, then EBE will be created.
Acceptance of bias is clearly related to S's personality traits. Following is a list of traits reportedly associated with EBE, together with some interpretive comments by the writer.

(a) High need of achievement and/or high ego involvement. Such an S, motivated to "do well", would look for cues from E to act so. But there would be less ceiling effect if S was biased to lower his performance level than to raise it.

(b) High need for social approval. Such Ss presumably seek E's approval by gratifying what they consider to be E's expectations. Other Ss would not be interested in E's expectations, and if those expectations are perceived, Ss may be indifferent to them.

(c) External control. Ss having this trait are not likely to resist external manipulation by E because they perceive the attainment of the expected result as minimally dependent on direct efforts on their part, i.e., beyond their personal control, and therefore would be more affected by E's suggestions. Internally controlled Ss are more likely to resist external manipulation and thus would be more difficult to bias.

(d) Field dependence. Ss having this trait may be more responsive to their social surroundings, including the E's running them, whereas field independent Ss are more likely to act according to their own notion and thus are less likely to be biased. It will be recalled that the experimental results on the relationship between EBE and field
dependence are not conclusive.

Several other personality traits of $S$ which appear to be positively associated with EBE are suggested for further study.

(a) Submission to authority. $S$s who are submissive to authority might turn to $E$ as a representative of authority for cues as to what are the appropriate responses. Submission to authority is the object of the present study.

(b) Self-confidence. The behavior pattern is likely to be similar to that suggested in (a).

(c) Tolerance of ambiguity - $S$s who are intolerant to ambiguity may try to avoid the ambiguous situation of the experimental task by looking for more clear cues from the $E$ running them.

The other component of EBE is $E$'s expectation and his ability covertly to communicate it. As demonstrated in the literature review, most authors did not consider whether $E$ had any expectations about the experimental outcome, or whether he sufficiently cared about the results to be motivated to achieve them.

Despite these limitations, several investigators were able to identify a number of variables which help $E$ to communicate his expectation to $S$, and to manipulate $S$ according to the expectations:

(a) Locus of Control. When $E$ is an internally controlled type he perceives the achievement of the desired results as a direct consequence of his actions and he might try more
actively to achieve his goal and to supply more cues for his Ss about the expected response. Conversely, an externally controlled E who feels no sense of personal control over the desired results, is likely to supply fewer cues for his Ss. (b) Dominance. When E is convinced of his ability to influence others he is likely to supply more cues. It is suggested in this thesis that dominant Es, who are usually used to controlling dyadic interaction, are more likely to supply more cues for their Ss.

Other personality variables of Es which might be positively associated with EBE are suggested for examination in future studies:
(a) Need for achievement and motivation. It is likely that E with these traits would supply more cues to his S than E lacking these traits.
(b) Self-confidence. It is likely that self-confident Es would be perceived by Ss as more reliable source of information.

Little is known on the interaction between the personality traits and the attitudes of E and of S. Apparently when both E and S have a high need for social approval, they are more likely to produce EBE than when only E or S have this trait. It is also known that overt attempts to sway Ss can result in reverse EBE.

A diagram illustrating the proposed model is given in Figure 1.
FIGURE 1

EBE MODEL

Diagrammatic Representation
Procedure Permitt

no/able to communicate expectation (?)

naive technician

create expectations re results ?

no

has apparent expectation?

yes

externally controlled submissive (?)
low in need for achievement (?)
not self-confident (?)
feels unable to influence (?)

NO EBE

REVERSE EBE

internally controlled dominant (?)
high in need for achievement (?)
self-confident (?)
feels able to influence (?)

overtly communicates expectations, how?

subtly
Permitting Interaction

1. Personally involved
   - High in need for:
     1. Achievement
     2. Social approval
   - Externally controlled
   - Field dependent (?)
   - Submissive (?)
   - Not self-confident (?)
   - Intolerant of ambiguity (?)

2. Not personally involved
   - Low in need for:
     1. Achievement
     2. Social approval
   - Internally controlled
   - Field independent (?)
   - Dominant (?)
   - Self-confident (?)
   - Tolerant of ambiguity (?)

3. Task requiring fundamental responses?
   - Yes
   - No

4. Co-operative (?)
   - Positive attitude to psych. (?)

5. EBE

6. No EBE

7. REVERSE EBE (?)
The Present Study

The aim of the present study is to examine the effect of dominance-submission on EBE. The first project was to test whether or not Ss who were submissive to authority were more susceptible to EBE. For this purpose two groups of Ss (dominant and submissive) were given RPRT, by two groups of Es whose scores were in the middle third of their sample's distribution on the Dominance-Submission scale. One group of Es was led to expect high ratings from their Ss (+5) and one group of Es was led to expect low ratings (-5).

In order to relate the measure of E's dominance to the measure of S's bias, namely to test whether or not dominant Es biased their Ss more, a second experiment was simultaneously carried out in which each of two groups of Es (one group consisting of dominant Es and the other consisting of submissive Es) tested a group of Ss whose scores were within the middle third of the distribution on the Dominance-Submission scale. Some Es were led to expect high ratings (+5) and the others were led to expect low ratings (-5).

A reliability study of the RPRT pictures was carried out prior to the above two experiments. The pictures were administered under a "no-expectancy" or a neutral condition to a group of students who happened to be present in the concourse of WLU. The same students were asked to rerate the pictures in the same situation a week later and about
half of them did.

In order to control the influence of the sex of S and E, E ran Ss of his or her own sex only. To control the attitude to psychological research, only Es and Ss without negative attitude to psychological research, as measured on Adair's Psychology Research Survey (PRS), were chosen.

The effect of Es' errors in recording the data was controlled by requiring Ss to record their own responses and to rate them out loud. The latter requirement was designed to create verbal interaction between E and S, and also to simulate the original RPRT situation in which S stated his responses out loud.

To create a situation in which submission to authority would be present (Brown 1953, reported on page 88 of the present work), a manipulation was carried out intended to establish a threatening ego-involving atmosphere. The manipulation chosen was proposed by Minor (1970) and is described on page 77 of this work.

During the trials a hidden observer who was blind to the experimental conditions recorded the number of glances exchanged between S and E, and the number of times E smiled at S. She also recorded the duration of the instruction period and the rating period. This procedure is similar to the one used by Friedman (1967) who found that the above four variables correlated positively with the S's tendency to perceive success in Rosenthal's photos.
Hypotheses

Restricting the sample to Es and Ss having a non-negative attitude to psychological research two hypotheses were suggested:

Hypothesis I: Ss who are submissive to authority are more susceptible to EBE than dominant Ss. It was assumed that submissive Ss faced with an ambiguous situation involving decision-making and conflict, would turn to E, as a representative of authority, for cues as to what was required from them.

Hypothesis II: Dominant Es are more likely to elicit EBE than submissive Es. It was assumed that dominant Es who are usually used to controlling dyadic interactions are more likely to supply cues to their Ss than submissive Es.
Preliminary Study

The writer carried out a reliability study on the pictures used in RPRT. She approached female students who were present in the concourse of Wilfrid Laurier University (WLU) and asked them to spend a few minutes on a psychological test. The first 50 volunteers were given written instructions on rating the pictures. As suggested by Rosenthal and Fode (1963), the use of written rather than oral instructions is likely to achieve a neutral expectation situation and to minimize the EBE. All participants were requested to repeat the test in the following week. Thus, the writer determined the test-retest correlation of the rating for each picture.

The mean and standard deviation were calculated to evaluate the proximity to "neutral" or zero and the dispersion of the ratings. It was found that only 11 pictures were rated around zero and there was a large dispersion in the ratings of the pictures. The ratings and re-ratings for most of the pictures were correlated at a significance level of at least $p \leq .05$.

In view of the foregoing results it seemed necessary to measure the ratings of every subject twice, (a) under no-expectancy condition and (b) under expectancy condition of a biased $E$. The mean of the differences between those measurements was used in the following experiments as an expression of S's susceptibility to EBE.

It was found that some of the pictures were more susceptible to random variations than the rest. Therefore,
for the purpose of the analysis, the pictures were divided into two sets. In order to replicate Rosenthal's experiments more closely, set I comprised the original 20 pictures. Set II comprised the nine pictures rated closer to zero and with smaller standard deviations. It was found that the pictures from Set II were more reliable since they were less susceptible to random variations, and therefore the EBE could be better revealed by them. Complete details on the Preliminary Study are given in Appendix A.
Method

Experimental Design

A 2x2 factorial design was used in Experiment One. One factor was type of S (dominant or submissive); the other factor was the picture-rating expectancy (high or low) given to medium dominant Es. The allocation of Ss to Es, which was determined by a random procedure on a same-sex basis is shown in Table 4.

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSIGNMENT OF Ss to Es</td>
</tr>
<tr>
<td>Experiment One</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E's Number</th>
<th>Es Expecting High Ratings</th>
<th>Es Expecting Low Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5</td>
<td>6 7 8 9</td>
</tr>
<tr>
<td>Number of submissive Ss tested</td>
<td>- - - 2 4</td>
<td>- - 2 2</td>
</tr>
<tr>
<td>Number of dominant Ss tested</td>
<td>3 2 1 - -</td>
<td>1 3 - -</td>
</tr>
</tbody>
</table>

Experiment Two can also be described as a 2x2 factorial design. The first factor was type of E (dominant or submissive) who tested the medium-dominant Ss; the second factor was the expectancy given to the Es (high or low). Ss were randomly assigned to Es on same sex basis as shown in Table 5.
TABLE 5
ASSIGNMENT OF Ss to Es
Experiment Two

<table>
<thead>
<tr>
<th>E's number</th>
<th>Dominant Es</th>
<th></th>
<th>Submissive Es</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expecting high ratings</td>
<td>Expecting low ratings</td>
<td>Expecting high ratings</td>
<td>Expecting low ratings</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

E_s and S_s were classified as dominant, medium-dominant or submissive if they scored in the upper, middle or lower third of their sample's distribution on Gold's Dominant-Submission Scale (GDSS).

Only E_s and S_s with a non-negative attitude to psychological research participated, the cut-off point being 156 points on Adair's Psychological Research Survey (PRS). The two scales are reproduced in Appendix B and Appendix C respectively.

All data collected either by E or by S were to be excluded if either one correctly guessed the purpose of the study. This was to be determined from the answers to post-experimental questionnaires for E_s and for S_s. Both experiments were carried out simultaneously.
Subjects - Experiment One

The Ss were chosen from WLU extension courses in Introductory and Social Psychology and from first year graduate students in the School of Social Work, on the basis of their scores on Gold's Dominant-Submission Scale (GDSS) reproduced in Appendix B. Ten students (two males and eight females) scoring in the lower third of their sample's distribution on the scale (seven to 24 points) were chosen as "submissive" Ss. Ten others (three males and seven females) who scored at upper third of their sample's distribution on the scale (37 to 41 points) were chosen as "dominant" Ss. An additional criterion for selection was a non-negative attitude towards psychological research, i.e., at least 156 points on Adair's Psychological Research Survey (PRS) reproduced in Appendix C. Those 20 students had in fact at least 171 points on the scale.

Nine Es (three males and six females) were selected from two Research Methods courses and a Social Psychology course. All nine Es scored within the middle third of their sample's distribution on GDSS (26 to 31 points) and had at least 156 points on Adair's PRS. In fact, 10 Es participated, but one of them suggested in a response to a questionnaire that the purpose of the experiment was "to see whether the E's knowledge of Ss' ability affects his actions with the Ss", so her data were excluded.
Subjects - Experiment Two

The Ss and Es were selected from the same classes as in Experiment One, and they were given the same tests. Thirty-two students (15 male and 17 females) scoring in the middle of their sample's distribution on GDSS (25 to 36 points) and scoring at least 156 points on PRS scale were selected to serve as Ss. In fact, all Ss scored at least 157 points. Six students (3 males and 3 females) scoring in the upper third of their sample's distributions on GDSS (33 to 40 points) were selected to serve as "dominant" Es. Six students (3 males and 3 females) scoring in the lower third of their sample's distribution on this scale (nine to 18 points) were selected to serve as "submissive" Es. In fact, all Es scored at least 176 on the PRS scale.

Two assistants participated in Experiment One and in Experiment Two, a male divinity student with a B.A. in psychology and his wife, a third-year student in psychology, each of whom dealt with same sex Es and Ss. A list of participants is given in Appendix D.

Apparatus

The apparatus comprised two group-administered tests for selection of the participants, an experimental task and two post-experimental questionnaires: (a) Gold's Dominance-Submission Scale was used to allocate Es and Ss to "dominant", "medium dominant" and "submissive" groups. (b) Adair's Psychological Research Survey was used to
eliminate Es and Ss with a negative attitude to psychological research. (c) Rosenthal's Photo Rating Test was the experimental task in both experiments. (d) A questionnaire for the Ss to examine their perception of the experiment and the Es. (e) A questionnaire for the Es to examine E's perception of the experiment and himself during the experiment. A detailed description follows.

(a) Gold's scale of Dominance-Submission (GDSS). The scale consists of 41 MMPI items. Gold, Leon and Swensen (1966) reported that the test-retest reliability was remarkably high (+0.939). To test the validity of the scale, Gold paired high-dominant and low-dominant male Ss who were given a decision-making task. Each pair of Ss was observed and their interaction rated by two judges.

The sample included 14 pairs of males. The agreement of scale score and judges' ratings in predicting which of the two Ss would dominate the situation were perfect on 10 out of the 14 pairs, but the \( \chi^2 \) was not significant. It was found, however, that of the 10 pairs of Ss with a dominance scale score difference of 12 or more points, nine were judged to be dominated by the member of the pair who scored higher (\( p = .065 \)).

In another validation study on female Ss, employing the same method, the scale and judges agreed on 13 of the 17 female pairs. A \( \chi^2 \) of 5.9 indicated significant agreement (\( p < .02 \)) between the scale scores and the judges' decisions as to which individual dominated. Combining the

*taken one week apart
results for the male and the female samples led to agreement between the scale and the judges on 23 out of 31 pairs studied ($x^2$, df=1 $p<0.001$). The GDSS is reproduced in Appendix B.

(b) Adair's Psychological Research Survey (PRS) (Adair, 1970). The scale of 52 items is presented in a five-choice Likert format and consists of statements which paraphrase Orne's detailed description of the attitudes toward psychology and psychological research (Orne 1962). The items are stated negatively and positively with equal frequency. The range of possible scores is 52 to 260, with high scores reflecting a positive attitude toward psychological research.

Adair and Fenton (1971) pointed out that:

While several checks of the reliability have yielded corrected split-half coefficients of .89 and .95, it is proposed that the construct validity of the scale is dependent upon demonstration of relationship between scale scores and external criteria.

A clear demonstration of the relationship was obtained in a series of studies on attitude change. The experiment consisted of three parts: a pre-test measure, a communication designed to change Ss' opinions and a post-test measure, all within the same session. Thus the demand characteristics of the experiment were rather obvious. Adair (1974) asserted that:

\textbf{Results overwhelmingly supported the hypothesis that Ss with more positive attitudes toward psychological research as measured by PRS would show greater opinion change than Ss with less positive attitude.}

However, no specific data or statistical analyses were reported.
In another attempt to examine the relationships between PRS and external criteria, Adair (1974) asked psychology students to volunteer for an experiment either for $1.50 per hour or gratis. Ss who volunteered to take part in the experiment had significantly higher scores on the PRS than Ss who did not (p<.05). When he solicited volunteers for a "Food and Nutrition Department taste experiment" he found no predictable relation between PRS scores and S's response. Other similar studies by Adair also indicated that the PRS has high construct validity for assessing attitudes towards psychological research (Adair 1974). The PRS is presented in Appendix C.

(c) Rosenthal's Photo-Rating Test (RPRT): E shows the S photographs of 20 different persons, one at a time, and asks S to rate how much success or failure is being experienced by the person photographed. A rating of +10 represents extreme success and -10 extreme failure. As ratings of zero may not be given S has in effect a 20 point rating scale with no neutral point. A modification of Rosenthal's scoring procedure was introduced in the present study: Ss rather than E recorded their responses, thereby eliminating recording errors by E. Rosenthal's standardization of these pictures was such that their mean rating was zero, or neutral with respect to success or failure. The pictures used in the test were chosen from 57 pictures which were given to standardization group of 70 male and 34 female students (Rosenthal, 1968). The photos used in the present
study were provided by Dr. Rosenthal. The answer scale of the RPRT is presented in Appendix E.

(d) Post-experimental questionnaire for Ss. This questionnaire consisted of one "open" question designed to reveal whether S was aware of the real purpose of the study, plus 16 "closed" questions to determine whether S was aware of any pressure exerted on him to rate the pictures in a particular way, and the way S perceived his E. The last 11 questions were given in a form of a 20 point rating scale with no neutral point. The questionnaire is similar to those used by Rosenthal and Fode (1963b) and by Levin (1961). It is presented in Appendix F.

(e) Post-experimental questionnaire for Es. The questionnaire consisted of one "open" question designed to reveal whether E was aware of the real purpose of the study and seven "closed" questions intended to evaluate his reaction to the experiment, to his Ss, and how he perceived himself as an E. The questionnaire is similar to that of Rosenthal and Fode (1963) and is presented in Appendix G.

Procedure - Experiment One

To select Ss, two group tests (GDSS and Adair's PRS) were administered to students enrolled in extension Introductory and Social Psychology courses as well as first year students in the Graduate School of Social Work. To select Es, the same two group-tests were administered to two classes of Research Methods and a class of Social Psychology.
It was found by Rosenthal et al. (1963) that $E$ is unlikely to produce the EBE when $E$ knows that he is the one who is actually being tested. The tests for the selection of $E$s were therefore administered by the professors who gave the courses. The test administrator told the students:

The Department of Psychology is carrying out a research project and you are requested to participate by completing two questionnaires. The information given by you will be used for research purposes only and will be treated as confidential. Please write your name at the top of the sheet.

To the class of $S$s the writer added that she had a special interest in the research since she would use it for her M.A. thesis, and that they would be contacted later to continue the research.

The $S$s and the $E$s were selected as non-negative toward psychological research and either high, medium or low in dominance as described in detail under "Subjects". To eliminate sex interaction $E$s tested same-sex $S$s.

The students selected to participate in the research as $E$s were asked by the writer to volunteer to serve as experimenters in her experiments and were told that the experiments were part of her M.A. research program. They were asked to write down the times when they would be available and their telephone number. The $S$s were asked to devote 15 minutes and $E$s to devote less than one hour for this purpose.

As a result of the preliminary study, the writer decided to measure the base-line, namely the rating of
Rosenthal's photographs under neutral conditions for each of the Ss individually, using the written instructions given in Appendix H. Written (rather than oral) instructions enabled her to obtain the base-line, both for male and female Ss. This follows from Rosenthal and Fode's (1963b) conclusion that no EBE occurs when E does not speak with S. This rating took place 21 to 71 days prior to Experiment One. The writer gave to the two assistants who were aware of the experimental conditions a set of written instructions to be read to the Es and the Ss. The instructions to the Es and the Ss are reproduced in Appendix I and J respectively.

Prior to the actual testing each E privately saw an assistant and was given a "Social Perception Test Experimental Procedure" sheet which is reproduced in Appendix K. The assistant saw to it that E read the sheet, asked him (her) if he (she) had any questions and then gave him (her) the instructions reproduced in Appendix I.

In order to create a difference between the Es assigned to the group which was led to expect failure (-5) and those led to expect success (+5), the Es assigned to the "-5" group were given the Social Perception Test Experimental Procedure sheet in which all the references to "+5" were changed to "-5" and the word "success" was replaced by "failure". In these instructions Es were told that if they followed the instructions and used the proper experimental procedure they should obtain high (or low) ratings, since
all their Ss had been found to be "success perceivers" (or "failure perceivers") according to personality tests and "other information" collected earlier this term. The aim of this part of the instruction was to make E feel personally responsible for the kind of data he collected, and motivate him to obtain certain kind of results.

At the end of the session the assistant handed E the set of Rosenthal's photos together with forms on which Ss were to rate the degree of success they perceived in the pictures. The assistant escorted E to a room with a one-way mirror in which the trial took place.

Most Ss came to the trial from their class. They found their names and the times they were needed written on the blackboard. Each S saw an assistant of his or her sex for five minutes before he (she) entered the test room. A manipulation to create a high ego-involvement atmosphere was then made by giving the Ss a "Background Information Sheet". The procedure was similar to that used by Minor (1970). The "Background Information Sheet" is given in Appendix L.

Through this manipulation the Ss were made to feel that the task they would be faced with was soluble in the sense that if they had the E's knowledge they could give responses which would be judged by a psychologist as indicative of a stable, mature and well-functioning personality. An attempt was made to make the Ss feel that they had a chance to do well or poorly in the experiment, and
that E was directly interested in how they did in fact do. 
This part of the "Background Information Sheet" was designed 
to create an ego-involvement atmosphere in which Ss would 
feel that their personality was evaluated and that their 
behavior was tied-in with personal success or failure as 
suggested by Rosenberg (1965). Such an atmosphere arouses 
defensive responsiveness and causes people who are sub­ 
missive to authority to behave in a different way than 
people who are not (Brown 1953). By creating a high ego- 
involvement atmosphere a new situation was created, and even 
though the Ss had seen the pictures previously, they prob­ 
ably made an effort to rate them "correctly" in view of the 
new information they had after the manipulation.

After the S had read the "Background Information 
Sheet", the assistant told him the following:

Today you are asked to rate the pictures 
under somewhat different conditions. The 
experimenter will read to you the instruc­
tions. Please listen to him carefully 
even if you still remember from the last 
meeting what you have to do. This will 
eliminate mistakes. After you have fin­
ished your ratings please come back to 
this room to answer a short questionnaire.

Then S was escorted to the test room and the trial 
began. During the trial an observer who was blind to the 
experimental conditions and who sat behind the one-way 
mirror recorded the number of glances exchanged between S 
and E, the number of times E smiled at S and timed the in­
struction period and the rating period on forms which are 
reproduced in Appendix M. The first eight subjects were
rated by two observers (the writer and the female assistant) in order to check on the reliability of the observations.

When S finished his rating he went back to the assistant, and was handed a questionnaire which is shown in Appendix F. The questionnaire was designed to evaluate S's perception of E and to find out whether S was aware of any pressure to rate the picture in a particular way. The completed questionnaire was taken from S by the assistant who thanked S for his participation.

When E had run all his (or her) Ss, the writer took him (her) to the observation room, gave E a short questionnaire which was intended to evaluate his reaction to the experiment and the way he perceived himself as E. The questionnaire is presented in Appendix G.

The aim of these two questionnaires was to find out whether people with different levels of dominance experienced the EBE in different ways. The questionnaires follow those of Rosenthal and Fode (1963b) and Levin (1961).

Procedure - Experiment Two

Experiment Two was designed to investigate the influence on the Ss of Es having different ratings on the Dominance-Submission Scale. The procedure followed that of Experiment One, but Es were six students who rated high and six who rated low on GDSS, and Ss were those who fell in the middle third of their sample's distribution on GDSS. Again, half of each group was led to expect "+5" rating and
In Experiment Two the time between the measurement of the base-line and the experimental measurements was between 21 and 75 days.
Results

Results for each of the two experiments are presented in the following sections. Experiment One examined the difference between dominant and submissive Ss regarding susceptibility to EBE. Experiment Two examined the difference between dominant and submissive Es regarding their ability to elicit EBE. Most of the statistical analyses were performed with nonparametric tests because the scales used were ordinal. The tests used in the present study are considered by Siegel (1956) to be the most powerful for testing the significance of the given data.

In each experiment, two separate analyses of the RPRT data were performed. The first analysis used all the 20 pictures as in Rosenthal's studies (Set I). The second used only those pictures which in the preliminary study were rated as closer to zero and had a smaller standard deviation (Set II). It was assumed that Set II pictures were more likely to reveal the appearance of EBE, but analyses using Set I pictures were performed in order to remain as close as possible to Rosenthal's own studies.

In both experiments most of the Ss were tested during class hours but they seemed co-operative in spite of the fact that they were missing lecture time. The Es were very co-operative, and most of them mentioned that they knew that they had been observed since they were familiar with the room in which the experiment had been conducted and its one-way mirror. Many said that they did not
feel comfortable in this situation, as shown by one who said, "I hate this room, there is always someone beyond the mirror watching you".

Results - Experiment One

The experiment was designed to test whether the experimenter's bias had different effects on the ratings of dominant Ss and submissive Ss. Four groups of Ss were compared: (a) dominant Ss tested by medium-dominant Es expecting "+5" rating; (b) dominant Ss tested by medium-dominant Es expecting "-5" rating; (c) submissive Ss tested by medium-dominant Es expecting "+5" rating; (d) submissive Ss tested by medium-dominant Es expecting "-5" rating. For each of the four groups of Ss the changes in ratings from the neutral condition measurement to the expectancy condition measurement were added and the average change per S was computed. This was done for the data associated with each of the two sets of pictures, (Set I - all 20 pictures, Set II - 9 pictures). The average rating changes for Set I and for Set II are given in Table 6 and Table 7 respectively.

Test of hypothesis. Hypothesis I: Ss who are submissive to authority are more susceptible to EBE than Ss who are dominant. Three tests were performed: (a) Kruskal-Wallis one way analysis of variance by ranks was chosen to test whether ratings given by the four groups of Ss were drawn from the same population. (b) Mann-Whitney U test
<table>
<thead>
<tr>
<th>Es' Expectation</th>
<th>Dominant Ss</th>
<th>Submissive Ss</th>
<th>Mann-Whitney</th>
<th>Kruskal-Wallis</th>
</tr>
</thead>
<tbody>
<tr>
<td>+5</td>
<td>-3.67</td>
<td>-2.83</td>
<td>&lt;.305</td>
<td>&lt;.90</td>
</tr>
<tr>
<td>-5</td>
<td>+6.75</td>
<td>-10.25</td>
<td>&lt;.457</td>
<td></td>
</tr>
</tbody>
</table>

### TABLE 7

AVERAGE RATING CHANGES AND SIGNIFICANCE LEVELS
Dominant vs. Submissive Ss: Set II Data

<table>
<thead>
<tr>
<th>Es' Expectation</th>
<th>Dominant Ss</th>
<th>Submissive Ss</th>
<th>Mann-Whitney</th>
<th>Kruskal-Wallis</th>
</tr>
</thead>
<tbody>
<tr>
<td>+5</td>
<td>+5.67</td>
<td>+1.33</td>
<td>&lt;.238</td>
<td>&lt;.50</td>
</tr>
<tr>
<td>-5</td>
<td>+10.5</td>
<td>-9.00</td>
<td>&lt;.238</td>
<td></td>
</tr>
</tbody>
</table>
examined separately the effect of $E_s$ expectations for "+5" or "-5" rating on dominant $S_s'$ ratings and on submissive $S_s'$ ratings. (c) Analysis of variance (2x2 factorial design) tested the effect of the factors ($E_s'$ expectations for "+5" or "-5" rating or the level of $S_s'$ dominance) and their interaction. The results of the last analysis should be considered with some reservation since it was performed on ordinal data.

Each of the three tests was performed on Set I data as well as on Set II data. The significance levels obtained by the non-parametric tests, namely, the Kruskal-Wallis test and the Mann-Whitney U test, are given in Table 6 and in Table 7 for each set of data respectively (page 119).

The results for the analysis of variance (2x2 factorial design) are summarized in Table 8 and in Table 9 for the two sets of data respectively. It can be seen from Tables 8 and 9 that on the basis of the non-parametric tests no EBE was found for either set of data. The analysis of variance failed to show a significant effect for either $E_s'$ expectation or for $S_s'$ dominance level in either set of data. It did show, however, a significant interaction effect for Set II data ($p<.025$). This effect is shown in Figure 2. It is noted that in Set II data (based on ratings of nine pictures) all dominant $S_s$ raised their ratings. Those tested by $E_s$ expected "-5" actually raised their ratings more than did those tested by $E_s$ expecting "+5". Thus the direction of the change was opposite to $E_s$' expectations.
### Table 8

**Analysis of Variance: 2x2 Factorial Design**

_Es' Expectancy x Ss' Dominance: Set I Data_

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>m.s.</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>7844.55</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Es' Expectation</td>
<td>10.7</td>
<td>1</td>
<td>10.7</td>
<td>0.02</td>
<td>n.s.</td>
</tr>
<tr>
<td>Dominance 'Level of Ss</td>
<td>198.36</td>
<td>1</td>
<td>198.36</td>
<td>0.45</td>
<td>n.s.</td>
</tr>
<tr>
<td>Es' Expectation x Dominance Level of Ss</td>
<td>590.89</td>
<td>1</td>
<td>590.88</td>
<td>1.34</td>
<td>n.s.</td>
</tr>
<tr>
<td>Error</td>
<td>7704.61</td>
<td>16</td>
<td>440.29</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### TABLE 9

**ANALYSIS OF VARIANCE: 2x2 FACTORIAL DESIGN**

_Es' Expectancy x Ss' Dominance: Set II Data_

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>m.s.</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>3515.80</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>Es' Expectation</em></td>
<td>36.30</td>
<td>1</td>
<td>36.30</td>
<td>0.28</td>
<td>n.s.</td>
</tr>
<tr>
<td>Dominance Level of <em>Ss</em></td>
<td>540.80</td>
<td>1</td>
<td>540.80</td>
<td>4.15</td>
<td>&lt;0.062</td>
</tr>
<tr>
<td>_Es' Expectation x Dominance Level of <em>Ss</em></td>
<td>853.14</td>
<td>1</td>
<td>853.14</td>
<td>6.55</td>
<td>&lt;0.025</td>
</tr>
<tr>
<td>Error</td>
<td>2085.56</td>
<td>16</td>
<td>130.29</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
FIGURE 2

ANALYSIS OF VARIANCE: 2x2 FACTORIAL DESIGN INTERACTION
S_s' Dominance x E_s' Expectancy
The rating changes of submissive Ss, were congruent with Es' expectations.

The reliability of the unseen observer (writer of the present study) was tested by comparing her observations with those made by the female assistant on the eight Ss who were tested first. This was done by performing Pearson's correlation coefficient test between their observations and between their time measurements. The results were \( r = +0.827 \) (\( p < .05 \)) for the number of times Es smiled at Ss, \( r = +0.86 \) (\( p < .01 \)) for the number of mutual glances exchanged between Es and Ss and \( r = +0.97 \) (\( p < .001 \)) for the times.

The analysis of the behavior of Es towards Ss was carried out using the ratings of the unseen observer only. The significance of the difference between medium-dominant Es' behavior towards submissive Ss and their behavior towards dominant Ss was tested by means of a t-test (two tailed) performed on the scoring obtained by the unseen observer. No significant difference was found. The results of this comparison are given in Table 10.

Answers to the first question in the Questionnaire for the Ss showed that none of them was aware of the real purpose of the experiment. One of the Es in an answer to a question suggested that the purpose of the experiment was "to see whether the E's knowledge of the S's ability affects his actions with the S"; accordingly, her data were excluded from the analysis.

To test whether there was a significant difference
TABLE 10
MEANS, STANDARD DEVIATIONS AND SIGNIFICANCE LEVELS
Submitive vs. Dominant Ss:
Mutual Glances, Es' Smiles and Time Spent

<table>
<thead>
<tr>
<th></th>
<th>Instruction Period</th>
<th>Rating Period</th>
<th>The Whole Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mutual Glances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submissive Ss</td>
<td>$\bar{X}$</td>
<td>2.67</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>1.67</td>
<td>1.99</td>
</tr>
<tr>
<td>Dominant Ss</td>
<td>$\bar{X}$</td>
<td>3.70</td>
<td>3.10</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>1.85</td>
<td>1.22</td>
</tr>
<tr>
<td><strong>p</strong></td>
<td></td>
<td>&lt;.22</td>
<td>&lt;.36</td>
</tr>
<tr>
<td><strong>Es' Smiles</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submissive Ss</td>
<td>$\bar{X}$</td>
<td>1.22</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.79</td>
<td>1.47</td>
</tr>
<tr>
<td>Dominant Ss</td>
<td>$\bar{X}$</td>
<td>1.20</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>1.33</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>p</strong></td>
<td></td>
<td>&lt;.50</td>
<td>&lt;.06</td>
</tr>
<tr>
<td><strong>Time Spent (seconds)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submissive Ss</td>
<td>$\bar{X}$</td>
<td>130.25</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>28.25</td>
<td>-</td>
</tr>
<tr>
<td>Dominant Ss</td>
<td>$\bar{X}$</td>
<td>126.30</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>31.80</td>
<td>-</td>
</tr>
<tr>
<td><strong>p</strong></td>
<td></td>
<td>&lt;.50</td>
<td>-</td>
</tr>
</tbody>
</table>
between the two kinds of Ss regarding their awareness of being pressured to rate the pictures in any particular way, Fisher's exact probability test was performed on the answers to the following questions of the Questionnaire for the Ss.

2. While going through the pictures, did you think that you were supposed to rate them in any particular way?

5. Did the experimenter want you to rate in a certain way?

6. Select the rating from +10 to -10 you believed the experimenter wanted to obtain.

7. On the scale below circle the number which you feel best expresses your average rating of the photographs.

<table>
<thead>
<tr>
<th>Extreme Failure</th>
<th>Moderate Failure</th>
<th>Mild Failure</th>
<th>Mild Success</th>
<th>Moderate Success</th>
<th>Extreme Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>-9</td>
<td>-8</td>
<td>-7</td>
<td>-6</td>
<td>-5</td>
</tr>
<tr>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td>+4</td>
<td>+5</td>
<td>+6</td>
<td>+7</td>
<td>+8</td>
</tr>
<tr>
<td>+9</td>
<td>+10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For Question 7 the comparison was made between those who rated in the same direction as the bias given to their E's and those Ss who rated in the opposite direction. The main difference between dominant and submissive Ss was in their notion that E's expected them to rate the pictures in a certain way (p=0.043) but they did not significantly differ in their ability to guess the direction of E's expectation, and in their estimate of their own average rating.

In order to examine whether there were significant differences between dominant and submissive Ss regarding their effort to rate the pictures "correctly", the Median test was performed on answers to the following questions of
the Questionnaire for the Ss.

3. How hard would you say that you tried to figure out what was a "good rating"?
   very hard    fairly hard    not hard at all

4. Would you say that you wanted to make a good rating?
   very much    some    didn't care one way or other

No significant differences were found between dominant Ss and submissive Ss regarding their attempt to determine what was a "good rating" (p=.24) and regarding their interest in making a good rating (p=.32). The selection of the Median test was based on the fact that the answers were rated on a three-point ordinal scale. A summary of results for answers to Questions 2 to 7 is given in Appendix N.

To determine the significance of the differences between dominant Ss and submissive Ss regarding their perception of the medium-dominant Es, a Mann-Whitney U test was performed on answers to the following questions:

8. On each of the rating scales listed below circle the number which best characterizes your experimenter during the experiment.

<table>
<thead>
<tr>
<th>Extremely Honest</th>
<th>Moderately Honest</th>
<th>Mildly Honest</th>
<th>Moderately Dishonest</th>
<th>Extremely Dishonest</th>
</tr>
</thead>
<tbody>
<tr>
<td>+10 +9 +8 +7 +6 +5 +4 +3 +2 +1</td>
<td>-1 -2 -3 -4 -5 -6 -7 -8 -9 -10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

9. Extremely Unfriendly | Moderately Unfriendly | Mildly Unfriendly | Moderately Friendly | Extremely Friendly |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-10 -9 -8 -7 -6 -5 -4 -3 -2</td>
<td>+1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The main points of the questions, means, standard deviations and significance levels for the comparison of the two kinds of Ss are given in Table 11. It can be seen that significant differences between dominant Ss and submissive Ss were found for: enthusiasm ($p < .002$), interest ($p < .049$),
**TABLE 11**

MEANS, STANDARD DEVIATIONS AND SIGNIFICANCE LEVELS

Submissive vs. Dominant Ss: Perception of Es

<table>
<thead>
<tr>
<th>No.</th>
<th>Main Point</th>
<th>Submissive Ss</th>
<th>Dominant Ss</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \bar{X} )</td>
<td>S.D.</td>
<td>( \bar{X} )</td>
</tr>
<tr>
<td>8</td>
<td>Honesty</td>
<td>6.43</td>
<td>3.06</td>
<td>7.29</td>
</tr>
<tr>
<td>9</td>
<td>Friendliness</td>
<td>4.43</td>
<td>2.19</td>
<td>7.43</td>
</tr>
<tr>
<td>10</td>
<td>Personal</td>
<td>0.57</td>
<td>4.87</td>
<td>1.86</td>
</tr>
<tr>
<td>11</td>
<td>Enthusiasm</td>
<td>1.43</td>
<td>2.19</td>
<td>5.43</td>
</tr>
<tr>
<td>12</td>
<td>Interest</td>
<td>2.71</td>
<td>1.75</td>
<td>5.57</td>
</tr>
<tr>
<td>13</td>
<td>Courtesy</td>
<td>4.50</td>
<td>2.81</td>
<td>7.00</td>
</tr>
<tr>
<td>14</td>
<td>Business-like</td>
<td>4.57</td>
<td>3.50</td>
<td>7.14</td>
</tr>
<tr>
<td>15</td>
<td>Professionality</td>
<td>3.86</td>
<td>3.64</td>
<td>7.00</td>
</tr>
<tr>
<td>16</td>
<td>Encouraging</td>
<td>0.86</td>
<td>1.96</td>
<td>3.43</td>
</tr>
<tr>
<td>17</td>
<td>Pleasant</td>
<td>3.57</td>
<td>1.99</td>
<td>2.29</td>
</tr>
</tbody>
</table>

*\( p < .01 \)
†\( p < .05 \)
courtesy $(p<.049)$, being business-like $(p<.003)$, professional $(p<.049)$, for being encouraging $(p<.006)$ and pleasant $(p<.009)$. For all questions dominant Ss rated their Es higher than submissive Ss.

Results - Experiment Two

The experiment was designed to test the effects of dominance and submission of the biased Es on the ratings of medium-dominant Ss. Four groups were compared: (a) ratings obtained by dominant Es expecting "+5"; (b) ratings obtained by dominant Es expecting "-5"; (c) ratings obtained by submissive Es expecting "+5"; (d) ratings obtained by submissive Es expecting "-5". For each of the four groups changes in ratings from the neutral condition to the expectancy condition were added and the average change per S was computed. This was done for the data associated with each of the two sets of pictures (Set I - 20 pictures, Set II - 9 pictures). Those average rating changes are given in Table 12 and Table 13 for Set I and Set II respectively.

Test of hypothesis. Hypothesis II: Es who are dominant are more likely to elicit EBE than Es who are submissive to authority. Three tests were performed:

(a) Kruskal-Wallis one way analysis of variance by ranks, chosen to test whether the ratings obtained by the four groups were drawn from the same population. (b) Mann-Whitney U test to examine separately the ability of dominant Es and submissive Es to elicit EBE from their Ss. (c) Analysis of variance (2x2 factorial design) to test the effect
### TABLE 12

**AVERAGE RATING CHANGES AND SIGNIFICANCE LEVELS**

Dominant vs. Submissive Es: Set I Data

<table>
<thead>
<tr>
<th>Es' Expectation</th>
<th>Dominant Es</th>
<th>Submissive Es</th>
</tr>
</thead>
<tbody>
<tr>
<td>+5</td>
<td>+10.13</td>
<td>-2.00</td>
</tr>
<tr>
<td>-5</td>
<td>+10.17</td>
<td>+8.75</td>
</tr>
</tbody>
</table>

Mann-Whitney: <.29, Kruskal-Wallis: <.30

### TABLE 13

**AVERAGE RATING CHANGES AND SIGNIFICANCE LEVELS**

Dominant vs. Submissive Es: Set II Data

<table>
<thead>
<tr>
<th>Es' Expectation</th>
<th>Dominant Es</th>
<th>Submissive Es</th>
</tr>
</thead>
<tbody>
<tr>
<td>+5</td>
<td>+10.63</td>
<td>-0.25</td>
</tr>
<tr>
<td>-5</td>
<td>+6.50</td>
<td>+6.63</td>
</tr>
</tbody>
</table>

Mann-Whitney: <.14, Kruskal-Wallis: <.16
of the factors, namely, Es' expectation (for "+5" or "-5" rating) or the level of Es' dominance and their interaction. Conclusions based on the last test should be considered with the reservation that the scale used was ordinal.

Each of the three tests was applied to each of the two sets of data. The significance levels obtained by the Kruskal-Wallis and the Mann-Whitney U test are given in Table 12 and in Table 13 respectively for each set of data. The Tables are given on page 131. A summary of results for the analysis of Variance is given in Table 14 and in Table 15 for Set I and Set II respectively. It can be seen from the tables that on the basis of the non-parametric tests no EBE was found for either set of data. Similarly, no main effects or interaction between main effects were found by means of the analysis of variance, as can be seen from Table 14 and Table 15 for Set I and Set II respectively.

To determine the significance of the differences between dominant Es and submissive Es in their behavior towards their Ss, a t-test (two tailed) was performed on the scoring obtained by the unseen observer. During the instruction period, dominant Es shared with their Ss an average of 2.27 mutual glances while submissive Es shared an average of 1.25 glances. This, and only this difference was significant (p<.02). The results of the comparison between the behaviors of the two kinds of Es as obtained by the t-test (two tailed) and the summary of the observations
<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>ms</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>16152.47</td>
<td>33</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Es' Expectation</td>
<td>348.70</td>
<td>1</td>
<td>348.71</td>
<td>0.73</td>
<td>n.s.</td>
</tr>
<tr>
<td>Dominance Level of Ss</td>
<td>506.56</td>
<td>1</td>
<td>506.56</td>
<td>1.07</td>
<td>n.s.</td>
</tr>
<tr>
<td>Es' Expectation x Dominance</td>
<td>1061.26</td>
<td>1</td>
<td>1061.26</td>
<td>2.24</td>
<td>n.s.</td>
</tr>
<tr>
<td>Level of Ss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>14235.94</td>
<td>30</td>
<td>474.53</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 15

**ANALYSIS OF VARIANCE: 2x2 FACTORIAL DESIGN**

*Es' Expectancy x Es' Dominance: Set II Data*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>ms</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>5873.53</td>
<td>33</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Es' Expectation</td>
<td>50.30</td>
<td>1</td>
<td>50.30</td>
<td>0.31</td>
<td>n.s.</td>
</tr>
<tr>
<td>Dominance Level of Ss</td>
<td>332.82</td>
<td>1</td>
<td>332.82</td>
<td>2.04</td>
<td>n.s.</td>
</tr>
<tr>
<td>Es' Expectation x Dominance Level of Ss</td>
<td>618.03</td>
<td>1</td>
<td>618.03</td>
<td>3.50</td>
<td>n.s.</td>
</tr>
<tr>
<td>Error</td>
<td>4847.33</td>
<td>30</td>
<td>161.58</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
are given in Table 16.

The answers to the first question in the Questionnaire for the E and the Questionnaire for the S showed that none of them guessed the real purpose of the experiment.

To test whether there was a significant difference between dominant E's and submissive E's regarding the way they perceived the average ratings given by their S's, a Fisher's Exact Probability Test was performed on answers to Question 2 from the Questionnaire for the E:

2. On the scale below circle the number which you feel best expresses your average obtained rating. Do this without actually scoring or referring back to your data.

<table>
<thead>
<tr>
<th>Extreme Failure</th>
<th>Moderate Failure</th>
<th>Mild Failure</th>
<th>Mild Success</th>
<th>Moderate Success</th>
<th>Extreme Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>-9</td>
<td>-8</td>
<td>-7</td>
<td>-6</td>
<td>-5</td>
</tr>
<tr>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td>+4</td>
<td>+5</td>
<td>+6</td>
<td>+7</td>
<td>+8</td>
</tr>
<tr>
<td>+9</td>
<td>+10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In Question 2, the comparison was between those who chose ratings in the same direction as the bias given to them and those who chose ratings in the opposite direction. No significant difference was found between dominant E's and submissive E's (p=.43). The summary of data and results are given in Table 17.

To examine whether dominant E's and submissive E's felt differently regarding their S's conforming to the expectations that were given to E's, Fisher's exact probability test was performed on the answers to Question 9 of the Questionnaire for the E.
TABLE 16
MEANS, STANDARD DEVIATIONS AND SIGNIFICANCE LEVELS

Submissive vs. Dominant Es:
Mutual Glances, Smiles and Time Spent

<table>
<thead>
<tr>
<th></th>
<th>Instruction Period</th>
<th>Rating Period</th>
<th>The Whole Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mutual Glances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submissive Es</td>
<td>$\bar{X}$</td>
<td>1.25</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.77</td>
<td>0.84,</td>
</tr>
<tr>
<td>Dominant Es</td>
<td>$\bar{X}$</td>
<td>2.77</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.96</td>
<td>1.74</td>
</tr>
<tr>
<td>$p$</td>
<td></td>
<td>&lt;.02</td>
<td>&lt;.50</td>
</tr>
<tr>
<td><strong>Es' Smiles</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submissive Es</td>
<td>$\bar{X}$</td>
<td>0.50</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.59</td>
<td>0.67</td>
</tr>
<tr>
<td>Dominant Es</td>
<td>$\bar{X}$</td>
<td>0.73</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.62</td>
<td>1.26</td>
</tr>
<tr>
<td>$p$</td>
<td></td>
<td>&lt;.29</td>
<td>&lt;.09</td>
</tr>
<tr>
<td><strong>Time Spent (seconds)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Submissive Es</td>
<td>$\bar{X}$</td>
<td>123.73</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>67.66</td>
<td>-</td>
</tr>
<tr>
<td>Dominant Es</td>
<td>$\bar{X}$</td>
<td>107.73</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>32.70</td>
<td>-</td>
</tr>
<tr>
<td>$p$</td>
<td></td>
<td>&gt;.50</td>
<td>-</td>
</tr>
</tbody>
</table>
TABLE 17
RESULTS FOR QUESTION 2 OF QUESTIONNAIRE FOR THE E

<table>
<thead>
<tr>
<th></th>
<th>Dominant Es</th>
<th>Submissive Es</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gave rating in the same direction to his expectations</td>
<td>4</td>
<td>3</td>
<td>0.43</td>
</tr>
<tr>
<td>Gave rating in opposite direction to his expectations</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
9. On the average did your Ss rate: (DO NOT LOOK BACK OR SCORE YOUR DATA!)

1. Too high? (By how many Points? 

2. Too low? (By how many Points? 

3. Just right?

The Fisher test was performed twice: (a) to test for significance in the difference between the number of dominant E's and submissive E's who thought their S's rated the pictures "too low" or "too high", (b) to test for significance in the difference between the number of dominant E's and submissive E's who thought that their S's rated the pictures "just right". None of these differences was found to be significant. The summary of the answers to Question 9 is given in Table 18.

The Mann-Whitney U test was applied to test for significance in the difference between dominant and submissive E's regarding their satisfaction with their participation in the experiment as given in answers to Question 3 of the Questionnaire for the E.

3. On the scale below circle the number which you feel best expresses your satisfaction with your participation in the experiment.

<table>
<thead>
<tr>
<th>Extremely Satisfied</th>
<th>Moderately Satisfied</th>
<th>Mildly Satisfied</th>
<th>Mildly Unsatisfied</th>
<th>Moderately Unsatisfied</th>
<th>Extremely Unsatisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>+10</td>
<td>+9</td>
<td>+8</td>
<td>+7</td>
<td>+6</td>
<td>+5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+4</td>
<td>+3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+2</td>
<td>+1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-5</td>
<td>-6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-7</td>
<td>-8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-9</td>
<td>-10</td>
</tr>
</tbody>
</table>

No significant difference between the two kinds of E's was found ($p < .057$).

The difference between dominant E's and submissive E's regarding liking their S's was tested for significance by the
TABLE 18

RESULTS FOR QUESTION 9 OF QUESTIONNAIRE FOR THE E

<table>
<thead>
<tr>
<th></th>
<th>Dominant Es</th>
<th>Submissive Es</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Too high</td>
<td>1</td>
<td>2</td>
<td>.26</td>
</tr>
<tr>
<td>Too low</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Just right</td>
<td>2</td>
<td>3</td>
<td>.38</td>
</tr>
<tr>
<td>Others*</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

*"Too high" ratings + "too low" ratings.
Mann-Whitney U test on the answers to the following question:

4. On the scale below circle the number which you feel best expresses your liking for the subjects.

\begin{tabular}{cccccccc}
\text{Extreme} & \text{Moderate} & \text{Mild} & \text{Mild} & \text{Moderate} & \text{Extreme} \\
\text{Dislike} & \text{Liking} & & & & \\
-10 & -9 & -8 & -7 & -6 & -5 & -4 & -3 & -2 & -1 & +1 & +2 & +3 & +4 & +5 & +6 & +7 & +8 & +9 & +10 \\
\end{tabular}

No significant difference between the two kinds of Es was found ($p<.545$).

The differences between the way dominant Es and submissive Es perceived themselves during the experiment was tested for significance by the Mann-Whitney U test performed on answers to the following questions from the Questionnaire for the E.

5. On each of the rating scales listed below circle the number which best characterizes yourself during the experiment.

\begin{tabular}{cccccccc}
\text{Extremely} & \text{Moderately} & \text{Mildly} & \text{Mildly} & \text{Moderately} & \text{Extremely} \\
\text{Honest} & \text{Dishonest} & & & & \\
+10 & +9 & +8 & +7 & +6 & +5 & +4 & +3 & +2 & +1 & -1 & -2 & -3 & -4 & -5 & -6 & -7 & -8 & -9 & -10 \\
\end{tabular}

6. Extremely Moderately Mildly Mildly Moderately Extremely

\begin{tabular}{cccccccc}
\text{Unfriendly} & \text{Friendly} & & & & \\
-10 & -9 & -8 & -7 & -6 & -5 & -4 & -3 & -2 & -1 & +1 & +2 & +3 & +4 & +5 & +6 & +7 & +8 & +9 & +10 \\
\end{tabular}

7. Extremely Moderately Mildly Mildly Moderately Extremely

\begin{tabular}{cccccccc}
\text{Personal} & \text{Impersonal} & & & & \\
+10 & +9 & +8 & +7 & +6 & +5 & +4 & +3 & +2 & +1 & -1 & -2 & -3 & -4 & -5 & -6 & -7 & -8 & -9 & -10 \\
\end{tabular}

8. Extremely Moderately Mildly Mildly Moderately Extremely

\begin{tabular}{cccccccc}
\text{Uncooperative} & \text{Cooperative} & & & & \\
-10 & -9 & -8 & -7 & -6 & -5 & -4 & -3 & -2 & -1 & +1 & +2 & +3 & +4 & +5 & +6 & +7 & +8 & +9 & +10 \\
\end{tabular}
The only significant difference between dominant Es and submissive Es was found in the answers to Question 8: dominant Es perceived themselves as more co-operative ($p < .032$). A summary of the results for the answers to Questions 3 to 8 of the Questionnaire for the E is given in Table 19.

Two major hypotheses and several related questions which were investigated yielded the following results:
(a) No EBE was found for either submissive or dominant Ss.
(b) No EBE was elicited either by dominant or submissive Es.
(c) Possible interaction between Ss' level of dominance and Es' expectations was suggested by the results obtained from data based on the nine photos which were found to be more reliable.
(d) Post-experimental Questionnaires suggested that submissive Ss felt they were expected to rate photos in a certain way.
TABLE 19
MEANS, STANDARD DEVIATIONS AND SIGNIFICANCE LEVELS
Submissive vs. Dominant Es: E's Questionnaire

<table>
<thead>
<tr>
<th>No.</th>
<th>Main Point</th>
<th>Submissive Es</th>
<th>Dominant Es</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X</td>
<td>S.D.</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>Satisfactory</td>
<td>-3.25</td>
<td>5.85</td>
<td>4.50</td>
</tr>
<tr>
<td>4</td>
<td>Liking of Ss</td>
<td>4.75</td>
<td>3.56</td>
<td>3.17</td>
</tr>
<tr>
<td>5</td>
<td>Honesty</td>
<td>6.75</td>
<td>2.49</td>
<td>6.17</td>
</tr>
<tr>
<td>6</td>
<td>Friendliness</td>
<td>3.75</td>
<td>4.09</td>
<td>5.83</td>
</tr>
<tr>
<td>7</td>
<td>Personal</td>
<td>1.50</td>
<td>5.55</td>
<td>1.17</td>
</tr>
<tr>
<td>8</td>
<td>Cooperative</td>
<td>6.25</td>
<td>2.58</td>
<td>4.84</td>
</tr>
</tbody>
</table>

*p < .05
Discussion

In this section, conclusions based on statistical tests of the hypotheses are discussed, results are compared with previous findings, and exploratory questions are posed. With the aid of the model presented in the study, suggestions are made for future research on the relation between personality variables and the EBE. The section closes with a brief discussion on the importance of EBE research in the behavioral sciences.

Experiment One was designed to test the hypothesis: Ss scoring as submissive on GDSS are likely to be more susceptible to EBE than dominant Ss (with all Ss scoring as having a non-negative attitude to psychological research on PRS). This hypothesis was not supported for either set of data (20 or 9 pictures) by the two nonparametric tests used.

However, a parametric test, analysis of variance, performed on Set II data (the 9 pictures found to be more reliable), indicated an interaction effect between Ss' level of dominance and Es' expectation (p<.025). Submissive Ss tended to change their ratings in the same direction as the expectation that was given to their Es, while dominant Ss tended to change their rating in a

*Fig. 1, page 97*
direction opposite to their Es' expectations. The same tendency was observed in Set I data but the interaction failed to reach significance. The interaction between Ss' level of dominance and Es' expectation should be regarded as tentative: the interaction was detected by a parametric test performed on ordinal data while the nonparametric tests failed to yield significant results. Moreover, the parametric test failed to show a significant main effect. Somewhat similar results were reported by Bootzin (1971). His data failed to show an overall EBE but did suggest an exploratory hypothesis, viz. Ss who consider themselves as more influenced by others tend to rate RPRT in the same direction as their Es' expectations.

In summary, the hypothesis that submissive Ss are more susceptible to EBE was not sufficiently supported by the data, but there are indications that the direction of Es' expectations might have different effects on dominant and submissive Ss.

A tendency of dominant Ss to rate the pictures opposite to their Es' expectation was found despite the fact that a non-negative attitude toward psychological research was a prerequisite. Adair (1972) concluded that those Ss who had such an attitude tended to conform to their Es' expectations. The results of Experiment One suggest as an explorative hypothesis that Adair's findings may not be correct for dominant Ss. Some support for the present contention may be found in Adair's report that his results were
significant for females ($p<.05$) but not for males ($p<.07$) who are usually considered to be more dominant. No examination of the variables associated with "reverse" EBE has been reported in the literature. One such variable is likely to be $S$'s dominance; it is suggested that this possibility merits further research.

It is felt that in the context of the present study, an a priori negative attitude as suggested by Masling (1966) and Argyris (1968) can be excluded. However, dominant $S$s may have developed a negative attitude to an $E$ who was trying to manipulate them and reacted with an attempt to spoil $E$'s efforts (Argyris 1966).

Several explanations for the failure to establish EBE can be offered. The first one is associated with the "demand characteristics" as suggested by Orne. Whereas in most studies which successfully demonstrated EBE $E$s were students of the investigator and had at least some motivation to obtain the expected results; in the present case the writer was an unfamiliar graduate student who had no relation whatever with the $E$s. The second explanation, also associated with the "demand characteristics" of the situation is the credibility of the source giving $E$s their expectations. Similarly, in the other studies the source of bias was the student's professor, in the present case it was a graduate student. This might have led to a lower level of credibility or to less bias. Since the effect of the expectation given to $E$s was not measured, this
explanation cannot be substantiated. The third explanation
is related to the fact that most Es felt that they were
being observed since they were familiar with the room in
which the experiment was conducted and its one-way mirror.
As suggested by Rosenthal et al. (1963), no EBE was likely
to occur when E felt that he was tested. Yet another ex-
planation for the results may be associated with the small
samples used in the two experiments. The failure to obtain
significant EBE may be due to Type II error (accepting the
null hypothesis when in fact it is false) whose probability
of occurrence increases with decreasing size of sample.

Support for the notion that submissive Ss felt that
Es expected them to rate the RPRT in a certain way was
found in the answers to the Questionnaire for the Ss. A
larger number of submissive Ss than of dominant Ss answered
"yes" to the question: "While going through the pictures,
did you think that you were supposed to rate them in any
particular way?" (p<.043). However, only 60% of the sub-
missive Ss who felt that Es expected certain ratings guessed
the direction of Es' expectation. An additional difference
found between submissive and dominant Ss was the way they
perceived their Es. Submissive Ss rated their Es as signi-
ficantly less enthusiastic (p<.002), less interested (p<.049),
less courteous (p<.049), less business-like (p<.003), and
less professional (p<.049). An interpretation of these
findings may be associated with the pressure to rate the
picture according to Es's expectations that submissive Ss
presumably felt. However, the design of the study did not permit every E to examine both kinds of Ss. Thus, even though Ss were randomly assigned to Es it may be argued that the difference in the perception of E by each group of Ss may have been due to real differences in behavior among Es, rather than an expression of the pressure felt by submissive Ss. Another explanation for the different way in which dominant and submissive Ss perceived their Es may be tied-in with the way they usually perceived other people. One may argue that submissive people usually tend to rate others lower than do dominant people. However, this last argument cannot be supported without further study.

The fact that Ss, reporting a feeling that E expected certain ratings, perceived their Es as less professional (p<.049) appears to contradict the findings of Friedman et al. (1965) who concluded that Es able to bias their Ss were perceived as more professional. However, one should remember that Friedman et al. obtained their correlation by comparing EBE with the way blind observers perceived the Es. It might well be that while E's behavior appeared professional to a blind observer, he was not so perceived by a biased S since a professional E is expected to be objective.

Based on the data of the unseen observer, no significant difference was found in the way Es behaved towards submissive and dominant Ss. This fact does not necessarily contradict the finding that submissive Ss perceived more pressure than dominant Ss, since this pressure may have
been expressed in a more subtle way. Friedman et al. (1965) suggested that short duration of instruction reading and fewer mutual glances between E and S are positively correlated with EBE. The present study failed to find both components of this correlation.

Experiment Two was designed to test the hypothesis that Es scoring as dominant on GDSS are more likely to produce EBE than Es scoring as submissive, (all Es scoring as having a non-negative attitude to psychological research). The two non-parametric tests used failed to support the hypothesis for both sets of data (20 pictures, 9 pictures). Dominant Es seemed to obtain ratings in the direction of the given expectation for Set II data. Submissive Es seemed to obtain ratings opposite to the given expectation in both sets of data. However, the interaction failed to reach significance. It is of interest that the only significant difference found between dominant and submissive Es regarding the way they perceived themselves as Es was that submissive Es rated themselves as less co-operative than dominant Es ($p<.032$). It appears that submissive Es perceived themselves as less cooperative because they felt that they failed to obtain the expected results. Conversely, one might argue that they failed to obtain the results because they were not cooperative. However, this point should be further investigated. Note that neither contention can be substantiated, since the difference between the ratings failed
to reach significance.

Several explanations can be offered for the failure to demonstrate EBE in Experiment Two. Some are similar to those given for Experiment One, namely, the differences between the way Es perceived the demand characteristics in the present study and in those studies which successfully demonstrated EBE: Es' motivation to obtain the expected results, the effect of the bias given from a source of lesser reputation, and Es' awareness of being observed. One can also speculate that dominant Es are likely to be less biased by the given expectation while submissive Es are likely to be more biased, but are less capable of biasing their Ss. Regrettably the effect of the bias given to the Es was not measured in the experiment so this explanation may only serve as an exploratory hypothesis to be confirmed in future research.

The only significant difference found between the behavior of dominant Es and submissive Es was that dominant Es shared with their Ss more glances than submissive Es during the instruction period (p<.02). Friedman et al. (1965) suggested that the number of mutual glances was positively correlated with Ss' perception of the pictures as "successful". The present study failed to confirm this finding: there was no significant difference between ratings obtained by dominant and by submissive Es.

Before summarizing the experimental findings, several limitations should be pointed out: (a) In Experiment
One, dominant and submissive Ss were not tested by the same E. Thus, although Ss were randomly assigned to their Es, the design did not sufficiently control the influence of Es' different personalities on Ss' ratings. The writer points to this as the most serious methodological flaw in the present work. (b) In both experiments the design did not permit each E to test Ss under the two kinds of expectations, ("+5" or "-5"). Thus the interaction between a given expectation and a certain E was not controlled. (c) The influence on E of the bias and ego involvement treatment given to him was not investigated. Thus no information is available regarding Es' belief in the expectations given to them, or their interest and effort in eliciting the experimental results. (d) The effect of the manipulation carried out to ego-involve the Ss was not measured, making it impossible to ascertain whether Ss did in fact feel that their personality was evaluated, and whether they tried to avoid failure, i.e., whether they satisfied the definition of "submissive". (e) The sample used was rather small. Since the power of statistical tests increases with the sample size, the probability of type II error (accepting the null hypothesis when it is false) increases. (f) Most Es felt that they were observed, thereby reducing the probability of eliciting EBE (Rosenthal et al. 1963).

It should be noted, however, that most studies which did successfully establish EBE did not measure the effect of the manipulation given to Es or Ss, including Minor, whose manipulation of the Ss was followed in the
present study. Moreover, during the performance of the test under "no-expectancy" conditions some Ss explicitly stated that they "knew" that rating the pictures was taken as expressing their personalities.

In summary, it can be concluded that although no EBE was found, some relations between submission to authority and EBE can be suggested for further study. It seems that submissive Ss who have a non-negative attitude to psychological research tended to conform to Es' expectation, while dominant Ss tended to respond in the opposite direction. A study of the relation between E's ability to elicit EBE and his level of dominance still seems promising even though the present work failed to tie down this relationship conclusively. A future study could test the hypothesis that dominant Es, who have a non-negative attitude toward psychological research will bias their Ss in the direction of their expectations, while submissive Es, even with a non-negative attitude to psychological research, will obtain ratings opposite to their expectations. It is suggested that these two hypotheses be tested by a better controlled study, using a larger sample and a more reliable experimental task.

Several additional suggestions for future EBE research may be considered. It seems important to explore the generality of the phenomenon through a systematic variation of tasks, experimental procedures, personality traits of the participants and especially the interactions among them. An
identification of the personality traits which reportedly and presumably underlie the EBE was attempted in this study under "EBE Model". A systematic investigation of these traits and their interactions might lead to a better understanding of the EBE and thereby allow prediction and control. However, the task is by no means easy since interaction within E or S and between E and S should be investigated.

The EBE associated with scoring errors by Es (EBE of the first kind) is of some practical importance; however, this aspect has been somewhat neglected by behavioral scientists. An investigation of the situational and personality variables of the type of E who tends to commit those errors might be useful.

One may question whether any future research in EBE should be undertaken. The interest in EBE derives mainly from the fact that experiments are a major source of information in the behavioral sciences. The presence of EBE in experimental situations limits the extent to which generalizations can be drawn about the effect of the experimental variables in non-experimental situations. In this context, the two kinds of EBE are important. It is obvious that when E biases his Ss, this bias is exclusive to their specific interaction in the laboratory. It is also obvious that E's recording or interpretation errors distort our knowledge even when S's responses are not affected.
Secondly, EBE research provides some information on dyadic interactions in general. It can teach us about the variables associated with the ability to influence and to be influenced. Similarly, research on the way in which E communicates his expectations to S may shed some light on covert communications in general.

The practical impact of EBE on professional psychologists appears to be negligible. Page and Yates (1973) reported that 90 percent of their sample of 250 American and Canadian psychologists from various fields felt that EBE did indeed have serious implications for psychology. Yet, an analysis of the 1971 literature (sample of 303 papers) revealed that "Hardly any study featuring more than one E reported E characteristics or controlled E's variables in any way". Similar findings were reported by Silverman (1974). It is not difficult to understand this state of affairs: present-day knowledge is insufficient to control the phenomenon satisfactorily. Yet, some suggestions might be offered. One possibility is to hire a technician, naive to the experimental hypotheses, to run Ss. Obviously, a technician might form his own hypotheses, but he is less likely to be motivated to do so. Also, a post-experimental inquiry could be designed to reveal his hypotheses. An alternate but less practical solution might be to replace the human E by a mechanical instruction and recording device.
Psychological researchers have historically paid close attention to the selection of Ss, but have invariably ignored the systematic selection of those who "run" the Ss. At the very least, it is suggested that until a better understanding of EBE is attained, investigators should report in some detail on their Es.
APPENDIX A

DETAILS OF PRELIMINARY STUDY
Method

Subjects

The first fifty female students present in the concourse of WLU who responded to the Researcher's request for "a few minutes on a psychological test" served as subjects.

Procedure

The writer carried out a reliability study on the pictures of Rosenthal's Photo Rating Test. She approached female students who were present in the concourse of WLU, requesting them to spend a few minutes on a psychological test. Fifty students who volunteered to do the test were given written instructions rather than oral ones in order to achieve a neutral expectation situation for the picture rating and to minimize the EBE (Rosenthal and Fode 1963). These instructions are reproduced in Appendix H. When each S finished rating the pictures the writer arranged a further meeting with her to re-rate the pictures the following week in the concourse at the same time. This enabled the writer to determine the correlation between two ratings of each picture on two occasions.

Results

Fifty students out of about 55 who were asked to participate in the preliminary study agreed to do so.
They seemed cooperative, trying to rate the pictures as well as they could. Most of them found this task difficult since it was not easy to judge whether a person experienced success or failure. Some of them mentioned that the pictures seemed to be strange or weird.

The results for these ratings and for the second set of ratings which took place a week after the first set, in which 22 out of 50 Ss participated, are summarized in Table 20. The mean and the standard deviation were used to evaluate how close to zero the pictures were on the average. It was found that only 11 pictures were rated around zero. The standard deviations of the 20 pictures were quite high and varied between 3.495 and 5.042. Spearman's rank correlation, a non-parametric test, was chosen since the scale was ordinal. The results using Spearman's rank correlation were between +0.2616 and +0.7158 with significance levels between \( p < .10 \) and \( p < .001 \).

Discussion

As can be seen from Table 20, there was a large dispersion in the ratings of every picture, and the average rating of some of the pictures was quite far from zero. The ratings and re-ratings for most of the pictures were correlated at a significance level of at most \( p < .05 \).

In view of the foregoing results it seemed necessary
**TABLE 20**

**PRELIMINARY STUDY: ROSENTHAL'S PHOTO RATING TEST: MEANS, STANDARD DEVIATIONS, CORRELATIONS BETWEEN RATING AND RERATING AND SIGNIFICANCE LEVELS**

<table>
<thead>
<tr>
<th>Picture Number</th>
<th>Mean Rating</th>
<th>Standard Deviation</th>
<th>Spearman's Correlation</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>-0.94</td>
<td>3.495</td>
<td>0.7104</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>5.042</td>
<td>0.5843</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>3</td>
<td>-4.10</td>
<td>3.661</td>
<td>0.6478</td>
<td>&lt;&lt;0.01</td>
</tr>
<tr>
<td>4*</td>
<td>-0.52</td>
<td>4.491</td>
<td>0.6233</td>
<td>&lt;&lt;0.01</td>
</tr>
<tr>
<td>5*</td>
<td>0.60</td>
<td>3.805</td>
<td>0.6207</td>
<td>&lt;&lt;0.01</td>
</tr>
<tr>
<td>6*</td>
<td>-0.70</td>
<td>4.428</td>
<td>0.5296</td>
<td>&lt;&lt;0.05</td>
</tr>
<tr>
<td>7*</td>
<td>0.52</td>
<td>3.837</td>
<td>0.4427</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>8*</td>
<td>0.18</td>
<td>3.734</td>
<td>0.6511</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>9</td>
<td>-2.26</td>
<td>3.769</td>
<td>0.4120</td>
<td>&lt;&lt;0.10</td>
</tr>
<tr>
<td>10*</td>
<td>1.62</td>
<td>4.376</td>
<td>0.7158</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>11</td>
<td>1.40</td>
<td>4.534</td>
<td>0.5454</td>
<td>&lt;&lt;0.005</td>
</tr>
<tr>
<td>12*</td>
<td>-0.20</td>
<td>4.372</td>
<td>0.5228</td>
<td>&lt;&lt;0.05</td>
</tr>
<tr>
<td>13</td>
<td>2.60</td>
<td>4.626</td>
<td>0.5249</td>
<td>&lt;&lt;0.05</td>
</tr>
<tr>
<td>14*</td>
<td>0.40</td>
<td>4.204</td>
<td>0.5248</td>
<td>&lt;&lt;0.05</td>
</tr>
<tr>
<td>15</td>
<td>-2.24</td>
<td>4.756</td>
<td>0.5869</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>16</td>
<td>-1.36</td>
<td>4.502</td>
<td>0.3954</td>
<td>&lt;0.10</td>
</tr>
<tr>
<td>17</td>
<td>4.12</td>
<td>3.603</td>
<td>0.6041</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>18</td>
<td>-0.26</td>
<td>4.803</td>
<td>0.4712</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>19</td>
<td>-0.34</td>
<td>3.937</td>
<td>0.2616</td>
<td>&gt;0.10</td>
</tr>
<tr>
<td>20</td>
<td>3.66</td>
<td>4.082</td>
<td>0.3055</td>
<td>&gt;0.10</td>
</tr>
</tbody>
</table>

*Set II pictures*
to measure the ratings of every subject twice: (1) under no expectancy conditions and (2) under expectancy conditions of a biased $E$. The mean of the differences between those measurements was used in the following experiments as an expression of $S$'s susceptibility to EBE.

As can be clearly seen from Table 20 some of the pictures were more susceptible to random variations than the rest. In view of the above, the pictures were divided into two sets for the purpose of the analysis. Set I comprised all the 20 pictures and Set II comprised the nine pictures which were rated closer to zero and had a smaller standard deviation. Set II consisted of pictures 1, 4, 5, 6, 7, 8, 10, 12, 14. It was found that the pictures from Set II were more reliable since they were less susceptible to random variations, and therefore the EBE could be better detected by them.
APPENDIX B

GOLD'S SCALE OF DOMINANCE - SUBMISSION
G.S.D.S.

Name ___________________________ Date ____________

Read question 1. If the statement is true circle the letter T on the lefthand side; if the statement is false circle the letter F. In the same way answer all the 41 questions.

T F 1. Once in awhile I think of things too bad to talk about.
T F 2. I find it hard to keep my mind on a task or job.
T F 3. My daily life is full of things that keep me interested.
T F 4. I prefer to pass by school friends, or people I know but have not seen for a long time, unless they speak to me first.
T F 5. I am certainly lacking in self-confidence.
T F 6. I certainly feel useless at times.
T F 7. I have often lost out on things because I couldn't make up my mind soon enough.
T F 8. I am a good mixer.
T F 9. My feelings are not easily hurt.
T F 10. I frequently have a fight against showing that I am bashful.
T F 11. I find it hard to make talk when I meet new people.
T F 12. I am happy most of the time.
T F 13. I brood a great deal.
T F 14. When in a group of people I have trouble thinking of the right things to talk about.
T F 15. I am not likely to speak to people until they speak to me.
T F 16. Life is a strain for me much of the time.
T F 17. I seem to make friends about as quickly as others do.
T F 18. In school I found it very hard to talk before the class.
T F 19. Even when I am with people I feel lonely much of the time.
T F 20. I am not unusually self-conscious.
T F 21. I blush no more often than others.
T F 22. I usually feel that life is worthwhile.
T F 23. I am easily embarrassed.
T F 24. I seem to be about as capable and smart as most others around me.
T F 25. The sight of blood neither frightens me nor makes me sick.
T F 26. I cannot keep my mind on one thing.
T F 27. I forget right away what people say to me.
T F 28. I usually have to stop and think before I act even in trifling matters.
T F 29. I have no dread of going into a room by myself where other people have already gathered and are talking.
T F 30. I very seldom have spells of the blues.
T F 31. At parties I am more likely to sit by myself or with just one person than to join in with the crowd.
T F 32. It is great to be living in these times when so much is going on.
T F 33. I often think, "I wish I were a child again".
T F 34. It makes me feel like a failure when I hear of the success of someone I know well.
T F 35. I am apt to take disappointments so keenly that I can't put them out of my mind.
T F 36. It bothers me to have someone watch me work even though I know I can do it well.
T F 37. I do not mind meeting strangers.
T F 38. I like to let people know where I stand on things.
T F 39. I sometimes find it hard to stick up for my rights because I am so reserved.
T F 40. The future seems hopeless to me.
T F 41. People can pretty easily change me even though I thought that my mind was already made up on a subject.
APPENDIX C

PSYCHOLOGY RESEARCH SURVEY
As you may know, some of the practices commonly used in psychological experiments employing human subjects are coming under review. The Director of the U.S. Public Health Service has made known his concerns in this area. As a result, prominent psychologists at Harvard, Columbia and Northwestern Universities are now investigating the psychological experiment from the subject's point of view.

As most subjects are drawn from University students, their opinions are being sought. The attached questionnaire is being sent to certain North American universities to sample student feelings about psychology and psychologists, as they function within the framework of the psychological experiment. From this and other work, it is hoped to establish a set of guide lines which will govern future investigations.

This is the first large scale and systematic enquiry into students' feelings about acting as subjects. We would ask you, then, to complete the questionnaire frankly and honestly.

A standard IBM answer sheet is provided for your responses. Do not make any marks on the questionnaire itself.

1. Enter your name, sex, age and today's date on the top row of the answer sheet.

2. In the space labelled "school", indicate the Faculty in which you are enrolled (for example: Arts, Science, etc.)

3. In the space provided for "grade or class", indicate your University year.

Now turn to the questionnaire and read question one. Select the response which best describes your feelings on this statement in accordance with the following scale.

STRONGLY DISAGREE DISAGREE UNDECIDED AGREE AGREE

If, for example, you strongly agree with the statement, blacken in the number 5 space for question 1 on the answer sheet like this:

1. 1 2 3 4 5
PSYCHOLOGY RESEARCH SURVEY

As you may know, some of the practices commonly used in psychological experiments employing human subjects are coming under review. The Director of the U.S. Public Health Service has made known his concerns in this area. As a result, prominent psychologists at Harvard, Columbia and Northwestern Universities are now investigating the psychological experiment from the subject's point of view.

As most subjects are drawn from University students, their opinions are being sought. The attached questionnaire is being sent to certain North American universities to sample student feelings about psychology and psychologists, as they function within the framework of the psychological experiment. From this and other work, it is hoped to establish a set of guidelines which will govern future investigations.

This is the first large-scale and systematic enquiry into students' feelings about acting as subjects. We would ask you, then, to complete the questionnaire frankly and honestly.

A standard IBM answer sheet is provided for your responses. Do not make any marks on the questionnaire itself.

1. Enter your name, sex, age and today's date on the top row of the answer sheet.

2. In the space labelled "school", indicate the Faculty in which you are enrolled (for example: Arts, Science, etc.)

3. In the space provided for "grade or class", indicate your University year.

Now turn to the questionnaire and read question one. Select the response which best describes your feelings on this statement in accordance with the following scale.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRONGLY DISAGREE</td>
<td>DISAGREE</td>
<td>UNDECIDED</td>
<td>AGREE</td>
<td>STRONGLY AGREE</td>
</tr>
</tbody>
</table>

If, for example, you strongly agree with the statement, blacken in the number 5 space for question 1 on the answer sheet like this:

1. 1 2 3 4 5

If you strongly disagree with it, blacken in the number 1 space on the answer sheet for question 1.

1. 1 2 3 4 5
As you can see, you have a choice of: (1) strongly disagree, (2) disagree, (3) undecided, (4) agree, and (5) strongly agree for each statement. Make your judgments in accordance with your degree of acceptance or rejection of the statement. However, you should try to avoid the "undecided" response as much as possible, as it is your feelings (either positive or negative) towards each of the statements that is being sought.

4. Proceed to answer each of the items, recording your answers on the answer sheet.
1. Most psychology experiments are worthless since even the most carefully controlled experiments lead to inconclusive results.

2. Through experimentation psychologists have made a real contribution to the understanding of man.

3. Psychologists would be better advised to forget the laboratory, and go into the field where the "real people and problems" are.

4. Many of the questions asked in testing are personal and are none of the experimenter's business.

5. Given a free choice, most students would be willing to volunteer for experiments.

6. Many experimenters are smug and take a pretty high-handed attitude with subjects.

7. Most experiments in psychology are concerned with trivial observations of artificial behavior.

8. Tests and other experimental manipulations are generally not reliable measures of personality and behavior.

9. Most experiments deal with such a small segment of behavior that they are meaningless in the broad picture.

10. People generally express their real feelings on psychological tests.

11. Psychology experiments are fun but do not prove anything.

12. Human behavior is too complex to cut up and study piece by piece in the laboratory.

13. Most people would say that their experience as a subject in psychological experiments was favourable.

14. When an individual signs up for an experiment, it involves a commitment to do what is asked to the best of his ability.

15. Most students participate willingly in experiments.

16. People rarely express their "real" selves in psychology experiments.

17. Experiments in psychology have no value because of the inherent diversity of man and his environment.

18. Many experimenters ask too much from their subjects.

19. Experiments are nothing but "busy work" for psychologists.

20. Psychology experiments are too time consuming.

21. Some experimenters just seem to be waiting for the subjects to make fools of themselves.
22. As a matter of personal pride, most individuals would try to do their best when acting as a subject.

23. Experimentation is of no practical value in the understanding of the fundamental causes of behavior.

24. The psychological journals are mostly filled with unimportant trivia.

25. It doesn't matter too much what subjects do; the experimenter usually manipulates the data to prove his hypothesis anyway.

26. Psychological tests are generally reliable measures of personality.

27. Laboratory studies in psychology are too artificial to produce valid data.

28. Most students are "good" subjects, that is, they perform well in their role as experimental subjects.

29. Many subjects in psychological experiments go through the motions without really trying.

30. The experimental method can be used effectively in the study of human behavior.

31. Subjects in most psychology experiments are treated with respect.

32. The experimental approach to psychology has been both fruitful and helpful in understanding human nature.

33. Most experimenters are considerate and polite in their treatment of subjects.

34. Participation in psychology experiments is not a great imposition on students.

35. Psychologists sometimes forget that subjects are still human beings.

36. Through psychological tests and experiments psychologists have acquired the knowledge to predict behavior in many real life situations.

37. Most students follow the experimenter's instructions carefully so that they will be able to perform as a good subject.

38. Laboratory studies in psychology have contributed significantly to the knowledge of mankind.

39. The complexity of individuals make it necessary to study human behavior under controlled conditions.

40. From experiments, psychologists can validly generalize to the population-at-large.

41. Subjects in most psychology experiments are treated as guinea pigs.

42. Many students do not cooperate and therefore make poor subjects.
43. Psychology has proven its worth as an experimental science.

44. Any minor discomfort that subjects may go through such as electric shock, embarrassment, etc., is worth it in the long run.

45. Psychological data is useless because its interpretation is based on the manipulation of statistics.

46. Many students feel a responsibility to cooperate in any way possible in the pursuit of knowledge.

47. Subjects frequently feel manipulated by the experimenter.

48. Participation in psychological experiments is a waste of the students' time.

49. Students should not be asked to give up their time to serve as subjects.

50. College students tend to share with experimenters the hope that the study in which they are participating will in some material way contribute to science.

51. Subjects in psychology experiments are "contributors to science."

52. Experiments in psychology almost always involve deception or "tricking" the subject in some way.
<table>
<thead>
<tr>
<th>Name</th>
<th>Sex</th>
<th>Age</th>
<th>Date</th>
<th>School</th>
<th>Grade or Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>27</td>
<td>***</td>
</tr>
<tr>
<td>2</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>28</td>
<td>***</td>
</tr>
<tr>
<td>3</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>29</td>
<td>***</td>
</tr>
<tr>
<td>4</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>30</td>
<td>***</td>
</tr>
<tr>
<td>5</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>31</td>
<td>***</td>
</tr>
<tr>
<td>6</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>32</td>
<td>***</td>
</tr>
<tr>
<td>7</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>33</td>
<td>***</td>
</tr>
<tr>
<td>8</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>34</td>
<td>***</td>
</tr>
<tr>
<td>9</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>35</td>
<td>***</td>
</tr>
<tr>
<td>10</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>36</td>
<td>***</td>
</tr>
<tr>
<td>11</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>37</td>
<td>***</td>
</tr>
<tr>
<td>12</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>38</td>
<td>***</td>
</tr>
<tr>
<td>13</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>39</td>
<td>***</td>
</tr>
<tr>
<td>14</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>40</td>
<td>***</td>
</tr>
<tr>
<td>15</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>41</td>
<td>***</td>
</tr>
<tr>
<td>16</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>42</td>
<td>***</td>
</tr>
<tr>
<td>17</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>43</td>
<td>***</td>
</tr>
<tr>
<td>18</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>44</td>
<td>***</td>
</tr>
<tr>
<td>19</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>45</td>
<td>***</td>
</tr>
<tr>
<td>20</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>46</td>
<td>***</td>
</tr>
<tr>
<td>21</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>47</td>
<td>***</td>
</tr>
<tr>
<td>22</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>48</td>
<td>***</td>
</tr>
<tr>
<td>23</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>49</td>
<td>***</td>
</tr>
<tr>
<td>24</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>50</td>
<td>***</td>
</tr>
<tr>
<td>25</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>51</td>
<td>***</td>
</tr>
<tr>
<td>26</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>52</td>
<td>***</td>
</tr>
</tbody>
</table>
APPENDIX D

LIST OF PARTICIPANTS
The following persons participated in the study:

(a) The writer: writer of present study.

(b) Experimenters (Es): students from WLU courses in Research Methods and Social Psychology, who administered the Picture Rating Test.

(c) Subjects (Ss): students from Extension courses in Introductory and Social Psychology, and first year students from WLU School of Social Work who took the Picture Rating Test and rated the pictures.

(d) Assistants: a psychology graduate divinity student and his wife, a third year student in psychology who briefed the Es and the Ss before the trials.

(e) Test Administrators: persons who administered Gold's Scale and Adair's Scale - WLU professors to the Es' classes and writer of present thesis to the Ss' classes.
THE SOCIAL PERCEPTION TEST RATING SCALE

<table>
<thead>
<tr>
<th>Extreme Failure</th>
<th>Moderate Failure</th>
<th>Mild Failure</th>
<th>Mild Success</th>
<th>Moderate Success</th>
<th>Extreme Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX F

QUESTIONNAIRE FOR SUBJECTS
Questionnaire for Subjects

1. The purpose of this experiment was ....

2. While going through the pictures, did you think that you were supposed to rate them in any particular way?

3. How hard would you say that you tried to figure out what was a "good rating"? very hard fairly hard not hard at all

4. Would you say that you wanted to make a good rating? very much some didn't care one way or other.

5. Did the experimenter want you to rate in a certain way?

6. Select the rating from +10 to -10 you believed the experimenter wanted to obtain.

7. On the scale below circle the number which you feel best expresses your average rating of the photographs.

<table>
<thead>
<tr>
<th>Extreme Failure</th>
<th>Moderate Failure</th>
<th>Mild Success</th>
<th>Mild Success</th>
<th>Moderate Success</th>
<th>Extreme Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>-9</td>
<td>-8</td>
<td>-7</td>
<td>-6</td>
<td>-5</td>
</tr>
<tr>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td>+4</td>
<td>+5</td>
<td>+6</td>
<td>+7</td>
<td>+8</td>
</tr>
<tr>
<td>+9</td>
<td>+10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8. On each of the rating scales listed below circle the number which best characterizes your experimenter during the experiment.

<table>
<thead>
<tr>
<th>Extremely Honest</th>
<th>Moderately Honest</th>
<th>Mildly Honest</th>
<th>Mildly Dishonest</th>
<th>Moderately Dishonest</th>
<th>Extremely Dishonest</th>
</tr>
</thead>
<tbody>
<tr>
<td>+10</td>
<td>+9</td>
<td>+8</td>
<td>+7</td>
<td>+6</td>
<td>+5</td>
</tr>
<tr>
<td>+4</td>
<td>+3</td>
<td>+2</td>
<td>+1</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>-3</td>
<td>-4</td>
<td>-5</td>
<td>-6</td>
<td>-7</td>
<td>-8</td>
</tr>
<tr>
<td>-9</td>
<td>-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extremely Unfriendly</th>
<th>Moderately Unfriendly</th>
<th>Mildly Unfriendly</th>
<th>Mildly Friendly</th>
<th>Moderately Friendly</th>
<th>Extremely Friendly</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>-9</td>
<td>-8</td>
<td>-7</td>
<td>-6</td>
<td>-5</td>
</tr>
<tr>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td>+4</td>
<td>+5</td>
<td>+6</td>
<td>+7</td>
<td>+8</td>
</tr>
<tr>
<td>+9</td>
<td>+10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extremely Personal</th>
<th>Moderately Personal</th>
<th>Mildly Personal</th>
<th>Mildly Impersonal</th>
<th>Moderately Impersonal</th>
<th>Extremely Impersonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>+10</td>
<td>+9</td>
<td>+8</td>
<td>+7</td>
<td>+6</td>
<td>+5</td>
</tr>
<tr>
<td>+4</td>
<td>+3</td>
<td>+2</td>
<td>+1</td>
<td>-1</td>
<td>-2</td>
</tr>
<tr>
<td>-3</td>
<td>-4</td>
<td>-5</td>
<td>-6</td>
<td>-7</td>
<td>-8</td>
</tr>
<tr>
<td>-9</td>
<td>-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Extremely Unenthusiasm</th>
<th>Moderate Unenthusiasm</th>
<th>Mild Unenthusiasm</th>
<th>Mild Enthusiasm</th>
<th>Moderate Enthusiasm</th>
<th>Extremely Enthusiasm</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10</td>
<td>-9</td>
<td>-8</td>
<td>-7</td>
<td>-6</td>
<td>-5</td>
</tr>
<tr>
<td>-4</td>
<td>-3</td>
<td>-2</td>
<td>-1</td>
<td>+1</td>
<td>+2</td>
</tr>
<tr>
<td>+3</td>
<td>+4</td>
<td>+5</td>
<td>+6</td>
<td>+7</td>
<td>+8</td>
</tr>
<tr>
<td>+9</td>
<td>+10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extremely</td>
<td>Moderately</td>
<td>Mildly</td>
<td>Mildly</td>
<td>Moderately</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>------------</td>
<td>--------</td>
<td>--------</td>
<td>------------</td>
</tr>
<tr>
<td>12.</td>
<td>Interest</td>
<td>Uninterest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+10 +9 +8 +7 +6 +5 +4 +3 +2 +1 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td>Discourteous</td>
<td>Courteous</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td>Business-Like</td>
<td>Unbusiness-Like</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+10 +9 +8 +7 +6 +5 +4 +3 +2 +1 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Unprofessional</td>
<td>Professional</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Discouraging</td>
<td>Encouraging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Unpleasant</td>
<td>Pleasant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX G

QUESTIONNAIRE FOR THE EXPERIMENTER
1. The purpose of this experiment is ...

2. On the scale below circle the number which you feel best expresses your average obtained rating. Do this without actually scoring or referring back to your data.

<table>
<thead>
<tr>
<th>Extreme Failure</th>
<th>Moderate Failure</th>
<th>Mild Failure</th>
<th>Mild Success</th>
<th>Moderate Success</th>
<th>Extreme Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. On the scale below circle the number which you feel best expresses your satisfaction with your participation in the experiment.

<table>
<thead>
<tr>
<th>Extremely Satisfied</th>
<th>Moderately Satisfied</th>
<th>Mildly Satisfied</th>
<th>Mildly Unsatisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>+10 +9 +8 +7 +6 +5 +4 +3 +2 +1 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. On the scale below circle the number which you feel best expresses your liking for the subjects.

<table>
<thead>
<tr>
<th>Extreme Dislike</th>
<th>Moderate Dislike</th>
<th>Mild Liking</th>
<th>Mild Liking</th>
<th>Moderate Liking</th>
<th>Extreme Liking</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On each of the rating scales listed below circle the number which best characterizes yourself during the experiment.

5. Extremely Honest | Moderately Honest | Mildly Honest | Mildly Dishonest | Moderately Dishonest | Extremely Dishonest |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+10 +9 +8 +7 +6 +5 +4 +3 +2 +1 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6. Extremely Unfriendly | Moderately Unfriendly | Mildly Friendly | Mildly Friendly | Moderately Friendly | Extremely Friendly |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6 +7 +8 +9 +10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

7. Extremely Personal | Moderately Personal | Mildly Impersonal | Mildly Impersonal | Moderately Impersonal | Extremely Impersonal |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>+10 +9 +8 +7 +6 +5 +4 +3 +2 +1 -1 -2 -3 -4 -5 -6 -7 -8 -9 -10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8. Extremely Moderately Mildly Mildly Moderately Extremely Uncooperative Cooperative

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

9. On the average did your $\ddot{S}s$ rate: (DO NOT LOOK BACK OR SCORE YOUR DATA!)

1. Too high? (By how many points? __________)
2. Too low? (By how many points? __________)
3. Just right?
APPENDIX H

WRITTEN INSTRUCTIONS TO S

Neutral Condition
Instructions to Ss

I am not permitted to say anything which is not in the instructions nor can I answer any questions about this experiment.

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 judgements 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 and circle the number. Here is the first photo.
APPENDIX I

INSTRUCTIONS GIVEN TO EXPERIMENTERS BY ASSISTANTS
Please print at the top of the answer sheet the name of the subject and his ordinal number (your first subject will be No. 1, your second No. 2 and so on).

Then read to the subject the attached instructions. Just read the instructions and say nothing else to them. If they ask you any questions about the experiment just read the statement at the bottom of the attached instruction page.

Ask your subjects to tell you their rating for each picture before they write it in order to avoid mistakes.

See that the subject rates every picture on a separate line.
APPENDIX J

INSTRUCTIONS GIVEN TO SUBJECT BY ASSISTANTS
Please give the subject the information sheet and see that he reads it.

Tell the Subject: Today you are asked to rate the pictures under somewhat different conditions. The experimenter will read to you the instructions, please listen to him carefully even if you still remember from last time what you have to do. This will eliminate mistakes. After you have finished your ratings, please come back to this room to answer a short questionnaire.
Social Perception Test Experimental Procedure

Please write at the top of the answer sheet the name of the subject. Then read to the subjects the attached instructions. Just read the instructions to the subjects and say nothing else to them. If they ask you any questions about the experiment just read the statement at the bottom of the attached instruction page. Then show the photographs one at a time.

For your information, we have found from past research that certain types of people tend to rate this picture in a very particular way. On the basis of personality test data and other information that we collected earlier this term from the subjects whom you will be running, you should expect them to give an overall average rating of +5. This is because our earlier testing has shown that all the subjects whom you will be running are in the category that we are calling "success perceiver". Thus, if you will follow instructions and use the proper experimental procedure, they will rate the picture more extremely positive (thus obtaining an average of about +5) than would be the case if your subjects had simply been selected at random.

PLEASE DO NOT DISCUSS THIS PROJECT WITH ANYONE
Instructions to be Read to the Subjects

I am going to read some instructions. We are in the process of developing a test of social perception. 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 recently. To help you make more exact judgements, you are to use this rating scale (Point to the rating scale in front of each subject). 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 recently. A rating of +10 means that you judge the person to have experienced extreme success recently. A rating of -3 means that you judge the person to have experienced mild failure recently, while a rating of +3 means that you judge the person to have experienced mild success recently, etc.

You are to rate each photo as accurately as you can but do not spend an excessive amount of time deliberating about any particular photograph. Just tell me the rating you assign to each photograph and circle that number. If, in the process of giving your answers, you wish to change your original response, feel free to change it by telling me, make an X on the old rating and circle the number of the new rating you choose.

The letters you see in the lower left corner are merely to identify the card you are rating. Ready? Here are the photographs.
If Questions Asked: I am permitted only to re-read the instructions. I cannot say anything which is not in the instructions, nor can I answer any questions about this experiment.
Today you will be participating in a psychological experiment; and shortly you will be assigned to an experimenter who will explain the task to you. Although we are not able to answer any questions until after the experiment is over, we do want to give you a brief description of the purpose of the experiment. This should make participating more interesting and meaningful for you. Also, a growing number of psychological researchers are beginning to realize that they have an ethical responsibility to make the purpose of their experiments known to the individuals who are helping them out by participating in their research. We are interested in studying social perception (i.e., how people perceive other people). More specifically, we want to find the factors which increase or decrease the accuracy of an individual's perception of other people. Certainly, with the lack of understanding in the world today, we do need to find out as much as we can about the reasons for inaccurate social judgement. Prior research by ourselves and others indicate that, typically, poor social perception is associated with psychopathology. That is, people who are not able to accurately perceive how other people are feeling, or what they are experiencing, usually are found to be psychologically maladjusted. Much of our initial research in this area indicates that on the basis of performance on the social perception task, we can pick out from a college population those students who would be judged clinically to be maladjusted. Several other researchers have presented data which support the preceding findings. Morgan and Provino (1963) for example,
report that in a college setting, the Social Perception Test could make rather subtle discriminations between varying degrees of emotional maladjustment and normalcy. The purpose of today's experiment, therefore, is to replicate the previous results, and thus to test further the generality of the finding that people who cannot accurately judge what other people are experiencing tend to be psychologically maladjusted.
APPENDIX M

OBSERVATION SHEET
Please mark each smile of E to S by √, the number of glances exchanged (by √) and measure the time of each period.

<table>
<thead>
<tr>
<th>Instruction period</th>
<th>Number of E's Smiles</th>
<th>Number of Mutual Glances (between E and S)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rating period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total experiment</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX N

ANSWERS TO "QUESTIONNAIRE TO SUBJECTS"
QUESTIONS 2 TO 7, SUMMARY OF RESULTS
The probabilities that there was no difference between the two kinds of Ss (dominant and submissive) regarding their awareness of any kind of Es expectancy are given below:

Question 2. While going through the pictures, did you think that you were supposed to rate them in any particular way?

<table>
<thead>
<tr>
<th></th>
<th>Dominant Ss</th>
<th>Submissive Ss</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>4</td>
<td>5</td>
<td>0.34</td>
</tr>
<tr>
<td>no</td>
<td>5</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Question 3. How hard would you say that you tried to figure out what was a "good rating"? very hard fairly hard not hard at all

<table>
<thead>
<tr>
<th></th>
<th>Dominant Ss</th>
<th>Submissive Ss</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>very hard</td>
<td>2</td>
<td>4</td>
<td>0.243</td>
</tr>
<tr>
<td>fairly hard</td>
<td>8</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>not hard at all</td>
<td>-</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Question 4. Would you say that you wanted to make a good rating?  
very much some didn't care one way or other.

<table>
<thead>
<tr>
<th></th>
<th>Dominant Ss</th>
<th>Submissive Ss</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>very much</td>
<td>5</td>
<td>6</td>
<td>0.315</td>
</tr>
<tr>
<td>some</td>
<td>4</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>didn't care</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Question 5. Did the experimenter want you to rate in a certain way?

<table>
<thead>
<tr>
<th></th>
<th>Dominant Ss</th>
<th>Submissive Ss</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>0</td>
<td>4</td>
<td>0.043</td>
</tr>
<tr>
<td>no</td>
<td>10</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Question 6. Select the rating from +10 to -10 you believed the experimenter wanted you to obtain.

<table>
<thead>
<tr>
<th></th>
<th>Dominant Ss</th>
<th>Submissive Ss</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>selected rating</td>
<td>1</td>
<td>5</td>
<td>0.066</td>
</tr>
<tr>
<td>selected zero or didn't select rating</td>
<td>9</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
### Question 7

On the scale below circle the number which you feel best expresses your average rating of the photographs.

<table>
<thead>
<tr>
<th>Extreme Failure</th>
<th>Moderate Failure</th>
<th>Mild Failure</th>
<th>Mild Success</th>
<th>Moderate Success</th>
<th>Extreme Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Failure</td>
<td>Failure</td>
<td>Failure</td>
<td>Success</td>
<td>Success</td>
<td>Success</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Dominant Ss</th>
<th>Submissive Ss</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>selected rating in the same</td>
<td>1</td>
<td>3</td>
<td>0.67</td>
</tr>
<tr>
<td>direction of E's bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>selected rating in the</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>opposite direction of E's</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Dominant Ss**: 3
- **Submissive Ss**: 2
- **p**: 0.42
References


Auffrey, J. and Robertson, M., "Case History Information and Examiner Experience as Determinants of Scoring Validity WAIS", American Psychological Association, 1972, 7 (p.2), 553-554.


Block, J. and Block, J., "An Interpersonal Experiment on the Reactions to Authority", Human Relations, 1952, 5, 91-98.


Masling, J., "Role-Related Behavior of the Subject and Psychologist and its Effects upon Psychological Data", in Levine, P., (Ed.) Nebraska Symposium on Motivation, Lincoln, Nebraska, University of Nebraska Press, 1966.


Soule, D., "Teacher Bias Effects with Severely Retarded Children", American Journal of Mental Deficiency, 1972, 77, 208-211.


