The Effects of Pre-Experimental Briefing and Experimenter Status on Social Desirability Scale Scores

Paul Ignatius Hettich

Loyola University Chicago

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Part of the Psychology Commons

Recommended Citation


This Dissertation is brought to you for free and open access by the Theses and Dissertations at Loyola eCommons. It has been accepted for inclusion in Dissertations by an authorized administrator of Loyola eCommons. For more information, please contact ecommons@luc.edu.

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License. Copyright © 1969 Paul Ignatius Hettich
The Effects of Pre-Experimental Briefing and Experimenter Status on Social Desirability Scale Scores

By

Paul I. Hettich

A Dissertation Submitted to the Faculty of the Graduate School of Loyola University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

February

1969
ACKNOWLEDGEMENTS

The writer wishes to thank the Loyola University Department of psychology for the use of subjects and laboratory facilities. Special acknowledgment is due Dr. Ronald E. Walker, the dissertation director, whose continued advice, guidance, and support have been invaluable during the past year. Special thanks are also due Dr. Paul J. von Ebers and Dr. Homer Johnson for their many suggestions and critical readings of the proposal and final manuscript. Finally, I wish to thank my wife Mary Ann for her continued support, encouragement and patience shown during my graduate education.
Life

Paul Ignatius Hettich was born on October 24, 1939, in Chicago, Illinois. He attended grade schools in various Midwestern states, and in 1957 was graduated from Subiaco Academy, Subiaco, Arkansas. He attended Creighton University, Omaha, Nebraska, from 1957 to 1958. Subsequently, he transferred to Marquette University, Milwaukee, Wisconsin, from which he was graduated in June, 1961, with a Bachelor of Arts degree in Psychology. In June, 1963, he received a Master of Arts degree in General-Experimental Psychology from New Mexico State University, Las Cruces, New Mexico. He entered active duty in the United States Army in November, 1963, and was stationed at Fort Jackson, South Carolina, where he served as a Personnel Psychologist (1st Lt.) at the Armed Forces Examining Station. In September, 1965, he entered a doctoral program at Loyola University, Chicago, Illinois, where he served as a graduate assistant and Lecturer in Psychology. In February, 1969, he received the Doctor of Philosophy degree in Experimental Psychology.
Chapter I

Introduction

A survey of the literature for the past fifteen years indicates a growing interest in the nature of research. Various approaches have been developed to ascertain which variables present in the research environment, in addition to those controlled and manipulated by the experimenter, can influence the experimental results.

The first purpose of this study is to attempt to clarify a methodological problem existing in experimenter effects research. Rosenthal (1966) has amassed a wealth of evidence indicating that the experimenter and the experimenter-subject relationship are critical variables in psychological research. However, this writer contends that there exists at least one methodological problem in Rosenthal's work which merits investigation. Involved is the failure of researchers to employ a control group which performs the task without having met an experimenter. While the findings of the Rosenthal studies are usually in terms of correlations between subjects' responses on the task and the experimenter variable manipulated, perhaps the same results are possible in the absence of an experimenter. This is not to deny that, for instance, the experimenter's sex or his expectancy cannot influence the results. But without a comparison group consisting of subjects not exposed to the physical presence of the experimenter, the results cannot confidently be ascribed to experimenter effects. If the performance of the proposed control group was not significantly different from that of groups receiving the treatments, then the unique influence of the independent variable (the particular experimenter
effects variable manipulated) would be lessened. Thus, this study attempts to introduce a methodological control, an absent experimenter group, into experimenter effects research.

The experimenter effects variable chosen for manipulation, labeled experimenter's relative status (Rosenthal, 1966), will be defined in this study as the experimenter's formality in dress (casual, formal, or absent), his academic rank (student or faculty), and his manner of behavior (casual or dignified). Thus, one examiner can alternate all three roles, the casually dressed student, the formally dressed, dignified faculty member, and the absent experimenter.

The use of deception in psychological research, according to Kelman (1967), rests on an assumption that subjects' awareness of the conditions which the experimenter is trying to create would affect his behavior in such a way that valid conclusions could not be drawn from the study. Although deception is a necessity for certain research problems, e.g., the Asch conformity situation (Asch, 1952) and stress research, its use has been criticized on methodological and ethical grounds (Baumrind, 1964; Kelman, 1967; MacKinney, 1955; Orne, 1962; and Vinacke, 1954).

Debriefing is, perhaps, the most commonly used method for offsetting the effects of deception and the withholding of information from the subject, besides attempting a resolution of the subjects' suspicions generated by the experiment. It should be noted, however, that debriefing is not synonymous with deception. While debriefing denotes the post-experimental explanation of the purposes, deception involves presenting a subject with false or misleading information about any aspect of the experiment.
Possibly the only feature which both procedures share in common is that their use is based on the acceptance of the assumption of the contaminating effects of a subject's prior awareness of the purposes.

Investigations of debriefing and its effects upon subjects are relatively recent (Brock & Becker, 1966; Walster, Berscheid, Abrahams, & Aronson, 1967; MacKinney, 1955) and many questions remain unanswered. One such question is: Would a subject's prior knowledge of the true purposes of an experiment affect the results? Common sense would generally dictate a "yes" answer in light of Kelman's comment above. Yet experimental evidence supporting this assumption for research in general seems to be lacking.

A simple approach for testing this assumption would be to pre-experimentally brief one group of subjects, i.e., explain the true purpose of the research to subjects prior to their participation, and debrief a second group. Nonsignificant differences between briefed and debriefed groups, assuming that subjects believed the instructions, would suggest that briefing did not affect the results; whereas significant differences would show that briefing is an important factor. If in certain situations briefing and debriefing are found to have similar effects, then the assumption of the contaminating effects of subject's prior knowledge cannot be regarded as universal. Thus, the second purpose of this study is to determine the existence of differences between briefed and debriefed groups of subjects.
Chapter II
Review of the Related Literature

**Experimenter's Relative Status**

Rosenthal (1966) classifies experimenter's relative status as a social psychological attribute of experimenter effects research. In studies of experimenter status the personality of the examiner is not investigated *per se*, but rather his externally defined characteristics. That is, the primary concern is not with such inherent attributes as sex, anxiety level, need for approval, authoritarianism, etc. Instead, the effects of dress, occupational rank, and relevant cues which provide information about his status, e.g., a name plate indicating rank or position, are the independent variables. The inherent personality attributes are present and can interact with the external characteristics being varied. However, the experimenter attempts to hold constant the level of these attributes.

There seems to be no clear definition of experimenter's relative status, but in Rosenthal's review status is described in terms of formality of "dress and manner," officer versus enlisted military rank, "professional, businesslike and less noisy," and student versus faculty. The studies defining status as student versus faculty warrant further comment as status in this experiment most closely approximates this classification.

In a discussion of introspective, observational, and participant levels of defining emotion, McTeer (1953) relates an example of how a laboratory experiment, originally designed for the objective level, inadvertently slipped to the participant level because the status of the experimenter had influenced the results. McTeer had been concerned with the effects upon
"other hand" tension of an electric shock punishment accompanying star tracing. The results revealed the existence of much tension as measured by pressure upon a rubber bulb. A year later an advanced student performed the same experiment but found little or no evidence of tension. After careful examination of all the data McTeer concluded that the differences in the results were such because in the first study the experimenter-subject relationship was of teacher to student, whereas in the second experiment an advanced student-subject relationship existed.

Birney (1958) attempted to replicate studies by French (1955) and Lowell (1951) on achievement motivation which showed a positive relationship between motivation and striving. French and Lowell had both carried a "student" status at the time the studies were conducted. To control for this factor Birney employed two groups in which a student experimenter administered the n Ach TAT to two groups of students, one group of which was from Birney's class, Birney administered the same task to two other groups all of whom were students in his class. Mean n Ach scores were greater for groups run by the faculty examiner than by the student examiner. Birney ruled out alternative explanations of inter-scorer unreliability and possible anxiety aroused in the student-teacher conditions.

In a verbal conditioning experiment, Sarason and Minard (1963) varied both experimenter and subject sex and hostility level, in addition to two situational variables, face to face experimenter-subject contact and experimenter prestige. Subjects in the high prestige condition were contacted by a person "arranging appointments" and greeted by a businesslike, well dressed experimenter whose name was on the door of the room. Low
prestige subjects were told that "a bunch of us are practicing on subjects," were greeted by the words "I guess you're mine," and led to the room which carried a sign on the door "practice subjects." Results relevant to the prestige factor revealed an interaction between hostility and prestige. Only two of the four experimenter hostility-prestige groups manifested learning, the high hostile experimenter-high prestige group and the low hostile experimenter-low prestige group. The interaction between prestige and the personal-impersonal conditions was attributed to a drop in the mean number of personal pronouns for subjects run in the low prestige-impersonal conditions. Only this group showed a decrement in performance.

The three studies reviewed suggest that subjects perform differently under treatments in which the relative status of the experimenter is that of student or teacher. While each of these studies credited the results to the experimenter's status, only Birney's investigation appears sufficiently generalizable to the present experiment in which subjects' social desirability is measured. Assuming that achieving is socially desirable in the American culture, subjects who are high achievers would tend to give more socially desirable responses. Birney found that higher n Ach scores were attained when the teacher administered the treatments. On this basis it could be predicted that more socially desirable responses would be elicited in the presence of a faculty member than in the presence of a student, assuming that subjects believe that the experimenter is a faculty member or a student.

At this point two limitations of the present study necessitate description. First, the findings of this research cannot be generalized beyond the one experimenter who administers all the treatments to all the
subjects. Although this is true, the variables under scrutiny here lack substantial investigation and are not well understood. At this stage in research it seems more appropriate to reduce to a minimum experimenter variance in hopes of focusing on the effects of the variables. A second limitation is the failure to control for the possible interaction of the experimenter's personality characteristics, e.g., anxiety level, need for approval, etc. and the external characteristics defining status, even if the experimenter attempts to hold the internal attributes constant.

Rosenthal (1966) points out:

Not only the kind of person the experimenter 'is' but the things that happen to him before and during the experiment affect his behavior in such a way as to evoke different responses from his subjects. The subject's behavior may have feedback effects on his own subsequent behavior not only directly but also by changing the experimenter's behavior, which then alters the subject's responses (p. 109).

That is, the experimenter may actually respond differently and perceive himself differently in the formal role, for instance, than in the informal role. Subjects could unconsciously or consciously pick up these subtle cues and react differentially. Yet the experimenter's awareness of the effects of his potentially differing role perceptions was itself a means of controlling these varying perceptions. Perhaps the best method for controlling the two limitations discussed above would be to employ multiple experimenters for all conditions. However, it was not the scope of this particular investigation to do so.

Absent Experimenter

A survey of the literature yielded no studies in which the experimenter's total physical absence from the whole of the experiment has been compared
with situations in which the examiner was present. It should be mentioned that Rosenthal is convinced that the experimenter can never be completely eliminated from the experiment because he makes decisions which may unintentionally affect the subject's responses. If this statement is true then the writer prefers to define an absent experimenter as one who is totally physically absent from the experiment until the treatment has been administered. This definition removes the researcher from the situation almost as far as possible. This distinction is important in this study since the writer contends that an absent experimenter condition is a requisite for experimenter effects designs. As the literature did not reveal any studies containing absent experimenter conditions as defined above, a brief review is in order of those in which the examiner was partially absent.

Bernstein (1956) reports how student nurses were administered the TAT under written or oral conditions and under examiner present or examiner absent conditions. There were no significant differences found between oral and written TAT protocols, but the examiner absent stories were sadder, had sadder outcomes, and showed greater involvement on the part of the subject. The results confirmed the hypothesis that the presence of an examiner in a testing situation acts as an inhibitor for strongly emotional material. Yet, Bernstein suggests that the presence or absence of the experimenter may not have been the operative variable; rather, the subject's expectancy of an immediate evaluation of the response. It is not clear whether examiner-absent subjects saw and spoke to the principal investigator, for it states: "Subjects in all conditions were shown into a private office
and were told they would find directions for what they were to do and the necessary materials on the desk." Nevertheless, all subjects did come face to face with someone associated with the experiment who could have been perceived as the experimenter.

In a study by Cassel, Johnson, and Burns (1958), adult job applicants were given the HTP test as part of a job application procedure. To test the hypothesis that an examiner's presence would affect the drawings, an examiner alternated being present and absent while an applicant was completing the test. Drawings made in the examiner's presence were significantly smaller and had an overall significantly smaller number of interpretable features on the house and person. Tree drawings were least affected by an examiner's presence or absence.

Van Krevelen (1954a) administered the MAPS test to twenty normal adults who first dictated a story to the examiner and subsequently wrote a story for the same stimulus after the examiner left the room. That written stories were significantly longer than dictated stories was interpreted as due, possibly, to the subject's ability to think more creatively or more elaborately in the absence of the examiner. The author failed to mention that the results could have been due to the subjects' ability or experiences in expressing themselves better in written rather than in an oral manner.

Van Krevelen (1954b) also administered the Szondi test to nineteen normal adult females under two conditions, one in which the test was self-administered by the subjects, the second in which an examiner administered the test. Subjects' responses were significantly more consistent and produced significantly more plus-minus reactions when the test was self-
Felice (1961) tested the hypothesis that if interpersonal relations disrupt the task performance of schizophrenics then the reduction of interpersonal relations in the testing situation should enhance performance. Subjects were assigned to one of four groups, schizophrenic-interpersonal, schizophrenic-impersonal, non-psychiatric-interpersonal, and non-psychiatric-impersonal, and administered the Ferguson Formboards, the Gorham Proverbs Test, the Concept Sorting Test, and the Mirror Drawing task. In the interpersonal conditions the examiner administered the tests, read the directions to the subjects, and remained in the room interacting with them throughout the session. In the two impersonal conditions the subjects were given initial instructions but, subsequently, were left alone to read the directions and complete the tests. The results relevant to this review showed that: 1. Schizophrenics performed less efficiently in the interpersonal conditions as predicted but only on the Ferguson Formboards; and 2. Performance of the non-psychiatric subjects in the interpersonal conditions was significantly inferior to other groups on the Mirror Drawing task. The latter finding was interpreted as being due to the frustrating nature of the task which appeared to become more stressful to these subjects in the presence of an examiner.

Verbal conditioning was used by Singer (1961) to determine if subjects' responses to the California F scale could be altered, and if so could the change be generalized to the California E scale. Singer also wanted to know if the generalization takes place when the source of reinforcement is absent.
One each of twenty-four pairs of female psychology students, first matched on a previous administration of the E and F scales, were assigned either to an experimental condition where the examiner read the items of the F scale and verbally reinforced pre-democratic responses, or to a control group where the experimenter simply recorded the responses to the statements. Next, the two groups of twenty-four subjects were each divided into an Experimenter Present and an Experimenter Absent condition. In the former condition the experimenter simply recorded subjects' responses to the E scale, whereas in the latter treatment the experimenter excused himself from the room and left the subjects to complete the E scale on their own. The results indicated that: 1. Learning did occur in the conditioning period for the experimental group (who received reinforcement) but not for the control group; 2. Subjects manifested an overall generalization effect by responding more prodemocratically when the experimenter remained in the room; and 3. No overall generalization occurred when the examiner was absent, but there was some generalization during the first ten trials. Singer concludes that the results reflect the dependence of learning on social influence situations.

In none of the six studies reviewed above was the examiner totally physically absent from the whole of the testing situation. However, a pilot study conducted by this writer sought to determine if any significant differences exist on the Marlowe-Crowne Social Desirability Scale (M-C SDS) scores among Formal Experimenter, Informal Experimenter, and Absent Experimenter groups in which half the subjects were briefed (Briefed conditions) while the remaining were debriefed (Debriefed condition). Table 1 summarizes
<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Briefed-Formal (BF)</td>
<td>6</td>
<td>12.83</td>
<td>7.65</td>
</tr>
<tr>
<td>Briefed-Informal (BI)</td>
<td>6</td>
<td>15.00</td>
<td>2.00</td>
</tr>
<tr>
<td>Briefed-Absent (BA)</td>
<td>6</td>
<td>15.17</td>
<td>4.59</td>
</tr>
<tr>
<td>Debriefed-Formal (DF)</td>
<td>6</td>
<td>16.33</td>
<td>3.70</td>
</tr>
<tr>
<td>Debriefed-Informal (DI)</td>
<td>6</td>
<td>15.67</td>
<td>5.87</td>
</tr>
<tr>
<td>Debriefed-Absent (DA)</td>
<td>16</td>
<td>11.44</td>
<td>4.37</td>
</tr>
<tr>
<td>Debriefed-Absent-Set (DAS)</td>
<td>16</td>
<td>12.69</td>
<td>5.54</td>
</tr>
<tr>
<td>Combined Groups</td>
<td>62</td>
<td>13.48</td>
<td>5.36</td>
</tr>
</tbody>
</table>
the results of this pilot study. The purpose of the seventh group will be explained later. The means for the Briefed-Formal (BF), Briefed-Informal (BI), Briefed-Absent (BA), Debriefed-Formal (DF), Debriefed-Informal (DI), and Debriefed-Absent (DA) conditions were 12.83, 15.00, 15.17, 16.33, 15.67, and 11.44, respectively, for male subjects as compared with a mean of 15.06 reported by Crowne and Marlow (1964) for 666 males in Table 2. Although no statistical analysis was performed on the data, dress, academic status, and manner of behavior were not differentiating variables in Debriefed conditions, whereas complete physical absence of the experimenter tends to result in lower performance. Yet, when subjects were briefed, i.e., had prior knowledge of the purposes of the study, a Formal Experimenter condition tended to emit slightly lower scores than an Informal Experimenter or Absent Experimenter conditions which do not differ. The absence of significant differences between DF and DI groups may have been due to: 1. the possibility that the distinction between Formal and Informal experimenters was not sufficiently evident; and/or 2. that differences in formality of dress, academic rank, and behavior would not be reflected on Marlowe-Crowne scores even if the Formal-Informal distinction were widespread. The lower mean score for the BF group may have been brought about because the subjects, having prior knowledge that the experimenter's role was enacted, "bent over backwards" to avoid the influence of the experimenter's formal role. Prior knowledge of the purposes also may have cancelled out any differences due to the experimenter status treatments, thus accounting for the similar mean scores of the BI and BA groups. The puzzling mean score for the DA group may be understood partially from the
<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crowne &amp; Marlowe (1960)</td>
<td>120 males and females</td>
<td>13.72</td>
<td>5.78</td>
</tr>
<tr>
<td>Crowne &amp; Marlowe (1964)</td>
<td>666 males</td>
<td>15.06</td>
<td>5.58</td>
</tr>
<tr>
<td></td>
<td>752 females</td>
<td>16.82</td>
<td>5.50</td>
</tr>
</tbody>
</table>
results of the post-experimental inquiry which revealed that some subjects felt apprehensive about walking into an experimental booth and finding no one there to greet them. Such apprehension may have caused discomfort to some subjects to the point that socially desirable responses were inhibited. Perhaps, if subjects in the DA group had prior knowledge of the experimenter's absence the situation would be more conducive to assessing social desirability.

To test this hypothesis sixteen new subjects, constituting the Debriefed-Absent-Set (DAS) group, were administered the DA condition except that the original "sign-up" folder contained instructions stating that they would not see their experimenter until they completed the task. Table 1 shows that the mean DAS performance of 12.69 was 1.25 units higher than the mean DA scores, indicating that prior knowledge of the examiner's absence may be significant in subsequent task performance for some subjects in the DA condition.

Generalization from the pilot study is limited because five of the seven groups contained an N of six, and because the number of experimenters was one (implying that the findings may be specific to the particular experimenter). In conclusion, the literature of the absent experimenter has been of limited value for the present investigation. Only the results from the pilot study provide information and these results are of restricted value.

Pre-Experimental Briefing

Regarding the status of research on debriefing, Brock and Becker (1966) state: "There is no published research in which debriefing has been independently varied and no theoretical writing was found to aid clear
specification of what should lead to what under which conditions [p. 316]." Support for this statement was found in the writer's search of the literature which revealed only three studies pertinent to debriefing.

In a study allegedly investigating the effect of decision making upon subsequent reading preferences (MacKinney, 1955) students were told that they could select either a multiple-choice or essay type examination for the psychology course in which they were enrolled. Subjects served in one of three groups: 1. a partial catharsis group which received cursory debriefing; 2. a full catharsis group which received complete debriefing; or 3. a control group which merely filled out a thirteen item post-experimental questionnaire. Analysis of the responses to the questionnaire led to the conclusions that there was little evidence that the subjects were disturbed by being deceived. The full catharsis and control groups expressed a significantly higher negative attitude than the partial catharsis group in regard to participating in an experiment without knowing that it was an experiment (item 7). Second, about 75% of the subjects stated that they preferred a complete explanation to a partial explanation or none at all, after the experiment (item 13).

Brock and Becker (1966) wanted to determine how the debriefing of subjects in one experiment influenced their acceptance of the events in a subsequent experiment. They found that the debriefing of subjects after the first experiment did not affect their sensitivity to the second unless there was a similarity between the tasks involved. They suggest that minimal debriefing is desirable in studies employing procedures and tasks similar to those in which the subjects have previously served.
Walster, Berscheid, Abrahams, and Aronson (1967) sought to determine if it would be more difficult to successfully debrief a subject who received false information about a topic he is concerned about, than it would to debrief a subject who received information irrelevant to his concern. There were no significant differences between high and low concerned subjects in reference to the difficulty of debriefing, nor was debriefing successful for some subjects.

With a dearth of studies in this area a host of questions remain unanswered. None of these reports compared debriefed groups with briefed groups. Perhaps the authors accept the assumption that prior knowledge of the examiner's intents are contaminants. The MacKinney study does supply some measure of subjects' attitudes toward knowing what an experiment is about, but it does not report if briefing and debriefing differentially affect subjects' performance.

The pilot study discussed earlier attempted to ascertain the role of prior knowledge of the examiner's purpose on the measurement of social desirability. It was suggested that briefing subjects produces lower scores for the Formal Experimenter condition, higher scores for the Absent Experimenter condition and no differences for the Informal Experimenter condition as compared with similar situations where subjects were debriefed. Again, it should be noted the small N for five of the seven conditions, including all the Briefed conditions, may have been non-representative of the population of subjects.

In conclusion, the literature reviews for experimenter's relative status, the absent experimenter, and pre-experimental briefing, and the
writer's pilot study do not provide a modicum of strong evidence for making specific predictions with confidence regarding the two problems under investigation: Do absent-experimenter groups differ from present-experimenter groups? Does the social desirability level of briefed and debriefed subjects differ? Consequently, only the following hypotheses were formulated:

1. Subjects serving in Formal Experimenter conditions (BF and DF) will receive higher M-C SDS scores than subjects in the Informal Experimenter (BI and DI) treatments, as based on the study by Birney (1958); and 2. On the basis of data gathered from the DA and DAS groups of the pilot study, it is hypothesized that DA subjects will receive lower M-C SDS scores than DAS subjects.
Chapter III

Method

The design of this experiment was a 2 X 4 factorial design as shown in Figure 1 where pre-experimental briefing is varied in two ways. Briefed and Debriefed, and experimenter's relative status is classified as Formal, Informal, Absent, and Absent-Set. It should also be noted that experimenter's relative status may also be classified into Experimenter-Present (E-Present) and Experimenter-Absent (E-Absent). Therefore, in this design Formal, Informal, Absent, and Absent-Set levels of status may be considered as nested within E-Present, E-Absent conditions.

All subjects were male college students enrolled in a daytime General Psychology course at Loyola University during the Spring semester of 1968.

Of ten subjects drawn from the Lewis Towers campus, the data of one subject was discarded because of his failure to follow instructions. Of the remaining nine subjects, two each served in the BF, DF, and BA groups, while one each served in the BI, DI, and DA conditions.

Of the two hundred subjects drawn from the Lake Shore campus pool, who, it should be noted, constituted about 90% of all make students enrolled in General Psychology during the day that semester, the data of nine subjects was discarded because of their failure to follow instructions (N=3); because they served twice, the second time being in an absent experimenter condition (N=4); or because as subjects assigned to DAS or BAS treatments, they saw the experimenter prior to their participation (N=2). The remaining 191 subjects from the Lake Shore campus were distributed through the eight conditions.
<table>
<thead>
<tr>
<th>EXPERIMENTER'S RELATIVE STATUS</th>
<th>Formal E</th>
<th>Informal E</th>
<th>Absent E</th>
<th>Absent E-Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Briefed Ss</td>
<td>BF</td>
<td>BI</td>
<td>BA</td>
<td>BAS</td>
</tr>
<tr>
<td>Debriefed Ss</td>
<td>DF</td>
<td>DI</td>
<td>DA</td>
<td>DAS</td>
</tr>
</tbody>
</table>

Figure 1. 2 X 4 Factorial Design
Thus, the data used in this study were acquired from two hundred subjects who were randomly assigned to one of eight treatments of twenty-five subjects each.

The task subjects performed was the Marlowe-Crowne Social Desirability Scale (see Appendix A), entitled "Personal Reaction Inventory," developed by Crowne and Marlowe (1960). The SDS consists of thirty-three items requiring true or false answers and was designed for use with "normal" individuals. While critics claim that the SDS is not completely independent of psychopathology (Katkin, 1964) and must be distinguished from social approval (Goldfried, 1964), numerous studies have appeared in which this scale is accepted as valid. Crowne and Marlowe's research relating the approval motive to M-C SDS scores was not examined in this study. A social desirability scale was chosen for two reasons. First, Rosenthal cites more than twenty studies supporting his work in which a photo rating task was used. His findings regarding the effects of the experimenter might be strengthened or weakened were experimenter effects research to employ a variety of tasks. Second, since this study attempts to assess subjects' reactions to the experimenter, the instructions, and the whole experimental instrument was desirable. Social desirability, by one definition, may be regarded as "the conscious or unconscious tendency of some subjects to respond in a socially desirable or undesirable fashion [p. 209 Megargee, 1966]." Although subjects may respond similarly regardless of the situation, their social desirability responses were made while reacting to the particular experimental situation. Through the random assignment of subjects to the eight conditions it was presumed that any differences in social
desirability were evenly distributed.

Regarding procedure, all subjects were administered the treatments individually. When subjects signed up, the examiner's name was listed as P. Hettich so as not to indicate his sex and status. For all except the DAS and BAS groups there were no special instructions on the "sign-up" folder. DAS and BAS subjects were run only during pre-designated weeks, one week at a time, during the fifth, ninth, thirteenth, and fifteenth week of the semester, with the following special instructions attached to the sign-up folder.

When you appear for this experiment your experimenter will not be there to meet you. Therefore, you are asked to follow the signs, enter the booth, read the instructions, and complete the task. However, your experimenter will contact you after you have finished and sign your 'requirement' sheet.

As a matter of convenience the examiner dressed either Formal or Informal on testing days, but never both. When formally dressed a die was cast (odds-evens) twice, first to determine if the subject would serve in the BF or DF group, and second, to determine of the experimenter was to be absent or present. Likewise, when informally dressed the die was cast first to determine the pre-experimental briefing condition and secondly, to ascertain the examiner's presence or absence. During weeks when the DAS and BAS subjects were run the die was cast once for assignment to pre-experimental briefing condition. Thus, the assignment of subjects to the eight groups was not technically random since each subject did not have an equal and independent chance to appear in any one of the eight conditions. However, the writer regarded the selection of subjects as random in the practical sense of the term in that the limiting conditions, e.g., formality
of dress, introduction of the examiner as a student or as a faculty member, and "set" instructions, were functions of the procedure and not of any inherent characteristics of the subjects. The writer was not aware of any relevant extraneous variables which were uncontrolled as a consequence of this subject assignment procedure.

In the Formal Experimenter conditions the examiner always wore a white laboratory coat over a suit coat or sweater with a tie and stated that he was Mr. Hettich, Lecturer in the Psychology Department. He attempted to act dignified, but not cold, in his speech and actions. Subjects in the Informal Experimenter treatments were greeted with "Hi!" by the examiner who always wore a casual shirt and slacks, but no laboratory coat, and carried a coffee cup. He introduced himself as Paul Hettich, a psychology student. In neither the Formal nor the Informal roles was the experimenter deceiving the subjects regarding his status. Instructions and subsequent activities for the Informal groups were identical with those of the Formal conditions in the appropriate pre-experimental briefing conditions.

Subjects arrived at the booth with the aid of signs placed in the hallway and on the door. Subjects in the E-Absent treatments saw a sign on the door directing them to enter. Subsequently, they saw a large sign on the table inside the booth instructing them to shut the door, be seated, and begin reading the instructions.

All subjects in the debriefed conditions read (In E-Present conditions he read the instructions with the subjects) the following:

In front of you is a list of 33 statements which I would like you to answer. Please mark your answers on the IBM answer sheet (1st alternative = TRUE, 2nd alternative = FALSE). After
you have finished this task I will sign your 'Laboratory Requirements' sheet. Now, please read the instructions to the task and begin work.

In addition to the instructions presented to Debrief groups, all Briefed subjects read the following:

This experiment has two purposes. First, I would like to see what the effects of clothes and formality of behavior have to do with the way subjects perform on a task. For one group of subjects I dress in a coat, tie, lab jacket, introduce myself as a faculty member, and try to act formal. For a second group I wear casual clothes, use my first name, and try to act informal. For other groups I am not even present until the end of the experiment. These groups read the instructions, perform the task, and don't see me until the end. You are in one of these conditions described above. This is the first purpose of this experiment.

The second intent of this study is to see if describing its purpose has any effects upon the results. I really don't know if it does or doesn't. Half the subjects (and you are one of them) are told the purpose of the experiment before they begin the task. Telling you the purpose is what I'm doing now. The other subjects aren't told the purpose until they finish the task. Thus, if you look at the diagram below you can see that you are a member of one of eight groups of subjects used for this study, and you are a member of one of the Briefed groups.

<table>
<thead>
<tr>
<th>Formal E Briefed Ss</th>
<th>Informal E Briefed Ss</th>
<th>Absent E Briefed Ss</th>
<th>Absent E-Set Briefed Ss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal E Debriefed Ss</td>
<td>Informal E Debriefed Ss</td>
<td>Absent E Debriefed Ss</td>
<td>Absent E-Set Debriefed Ss</td>
</tr>
</tbody>
</table>

You may be wondering about the task you have to complete. We chose this particular task because we believe that it can give us some indication of your psychological reaction to this experiment. The Personal Reaction Inventory (PRI) is intended to measure how well you want to do in this experiment. Okay, let's begin the task.

Subjects in the two Briefed Absent conditions, BA and BAS, found the sheet containing these instructions on top of the SDS and the answer sheet. It should be noted at this point that instructions pertinent to
briefed and debriefed treatments refer only to the SDS and not to the questionnaire. That is, while subjects were informed of the rationale for the SDS either before or after its administration, no subject had knowledge of the post experimental rating scale and questionnaire until the examiner requested him to complete it.

Upon completion of the task, the examiner (entered the room in the Absent Experimenter conditions) instructed the Debriefed subjects that431(134,471),(924,958) the experiment was over and that the purposes of the study would be explained. A copy of the "Explanation" was then presented and the examiner read the explanation with the subject. The instructions on the "Explanation" sheet were nearly identical to those on the "Instructions" sheet used in the Briefed conditions, except that the past tense was used.

To check the effect of the manipulation of the pre-experimental briefing variables and the absent experimenter variable, each subject completed a post-experimental questionnaire and rating scale. On a plus 10 to minus 10 scale subjects rated the examiner on each of the following items: Honest, Professional, Dignified, and Casual. In addition all subjects answered two questions: "How many psychology experiments have you participated in prior to this one?"; and "Do you believe that the experimenter was telling you the truth in regard to the purpose of this experiment?" Subjects could answer the latter question by circling "Yes," "No," or "Don't Know."

All subjects who served in the Absent Experimenter conditions (BA, BAS, DA, and DAS) were asked "In one or two words tell me how you felt about walking into a room and finding no one there."
Subjects who served in the BAS and DAS conditions were also queried: "For some Absent Experimenter subjects there were special instructions in the sign-up folder telling what to expect when they arrive for the experiment. Briefly, what were these instructions?" Appendix B contains the post-experimental rating scale and questionnaire of which the above items constitute a part. The remaining items and questions in this appendix provide for data which is to be incorporated in an additional study.
Chapter IV

Results

Marlowe-Crowne Social Desirability Scale Scores (M-C SDS)

A subject's SDS score was calculated by summing the number of items he answered in a socially favorable manner. While the possible range of scoring extends from zero to 33, the extreme scores obtained in this study ranged from two to 29.

The overall SDS mean and standard deviation of 14.24 and 5.77, respectively, obtained from a sample size of 200, approximates those reported by Crowne and Marlowe (1964) (\(\bar{X}=15.06\), SD=5.58, N=666 males).

The means and standard deviations for each of the eight groups of subjects are presented in Table 3.

Examination of the data showed that each of the means of the Experimenter-Present groups (BF, BI, DF, DI) were higher than each of those of the Experimenter-Absent groups (BA, BAS, DA, DAS).

To simplify their presentation the means and standard deviations were calculated (Table 4) according to the independent variables manipulated: experimenter status and pre-experimental briefing, in addition to showing the E-Present versus E-Absent nested classification. The higher mean scores attained by the E-Present groups become more apparent not only in the E-Present versus E-Absent comparison but also when the Formal and Informal groups are juxtaposed with the Absent and Absent-Set groups. The overall differences between Briefed and Debriefed groups was slight.

Figure 2 plots the relationship between the mean SDS scores and the briefing conditions for the four experimenter status groups. Besides
### TABLE 3

Means and Standard Deviations for M-C Scores

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Briefed-Formal (BF)</td>
<td>25</td>
<td>14.40</td>
<td>6.09</td>
</tr>
<tr>
<td>Briefed-Informal (BI)</td>
<td>25</td>
<td>15.48</td>
<td>5.58</td>
</tr>
<tr>
<td>Briefed-Absent (BA)</td>
<td>25</td>
<td>11.48</td>
<td>4.09</td>
</tr>
<tr>
<td>Briefed-Absent-Set (BAS)</td>
<td>25</td>
<td>14.32</td>
<td>5.38</td>
</tr>
<tr>
<td>Debriefed-Formal (DF)</td>
<td>25</td>
<td>16.04</td>
<td>6.26</td>
</tr>
<tr>
<td>Debriefed-Informal (DI)</td>
<td>25</td>
<td>15.16</td>
<td>5.09</td>
</tr>
<tr>
<td>Debriefed-Absent (DA)</td>
<td>25</td>
<td>13.60</td>
<td>5.20</td>
</tr>
<tr>
<td>Debriefed-Absent-Set (DAS)</td>
<td>25</td>
<td>12.12</td>
<td>6.64</td>
</tr>
<tr>
<td>Combined Groups</td>
<td>200</td>
<td>14.24</td>
<td>5.77</td>
</tr>
</tbody>
</table>
TABLE 4

Means and Standard Deviations According to Variables Manipulated

<table>
<thead>
<tr>
<th>Variable</th>
<th>Condition</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status</td>
<td>Formal</td>
<td>50</td>
<td>15.22</td>
<td>6.22</td>
</tr>
<tr>
<td></td>
<td>Informal</td>
<td>50</td>
<td>15.32</td>
<td>5.34</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>50</td>
<td>12.72</td>
<td>4.76</td>
</tr>
<tr>
<td></td>
<td>Absent-Set</td>
<td>50</td>
<td>13.22</td>
<td>6.13</td>
</tr>
<tr>
<td>E-Presence</td>
<td>E-Present</td>
<td>100</td>
<td>15.27</td>
<td>5.80</td>
</tr>
<tr>
<td></td>
<td>E-Absent</td>
<td>100</td>
<td>12.97</td>
<td>5.50</td>
</tr>
<tr>
<td>Debriefing</td>
<td>Briefed</td>
<td>100</td>
<td>14.01</td>
<td>5.49</td>
</tr>
<tr>
<td></td>
<td>Debriefed</td>
<td>100</td>
<td>14.23</td>
<td>6.02</td>
</tr>
</tbody>
</table>
Briefed  Debriefed

PRE-EXPERIMENTAL BRIEFING

Figure 2. The Relationship Between Pre-Experimental Briefing and Mean M-C SDS Scores As a Function of Experimenter Status
making apparent the higher performance of the E-Present groups, the data suggested the presence of interactions within the E-Present and E-Absent groups.

The SDS scores were analyzed by means of applying a 2 X 2 factorial analysis of variance (McCuigan, 1968) to three classifications of the data: Briefing and Debriefing with E-Present and E-Absent, Briefing and Debriefing with Absent and Absent-Set, and Briefing and Debriefing with Formal and Informal.

The results of the factorial analysis of variance comparing Briefing and Debriefing with E-Present and E-Absent are shown in Table 5 where an F of 8.13 for the main effects status was found significant beyond .01.

To further ascertain the specific source of the main effects of status, a Duncan's Multiple Range test (McGuigan, 1968) was performed on the four conditions of status by combining (the F for the main effects of Debriefing was only .07) the scores of the Briefed and Debriefed subjects. The difference of 2.60 between the means of 15.32 and 12.72 for the Informal and Absent groups, respectively, was significant beyond the .05 level (Rp₄=2.41), as was the difference of 2.50 between the means of 15.22 and 12.72 for the Formal and Absent groups, respectively, (Rp₃=2.33).

A 2 X 2 factorial analysis of variance which compared Briefing and Debriefing with E-Absent and E-Absent-Set was not significant although the obtained F of 3.23 for the interaction effects reported in Table 6 approached the required F (for significance at .05) of 3.92.

The third 2 X 2 factorial analysis of variance comparing Briefing and Debriefing with Formal and Informal is summarized in Table 7 where
TABLE 5

Summary Table: 2 X 2 Factorial Analysis of Variance--Briefing and Debriefing X E-Present and E-Absent

<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>Between Groups</td>
<td>(276.60)</td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Briefing (B)</td>
<td>2.42</td>
<td>1</td>
<td>2.42</td>
<td>.07</td>
<td>NS</td>
</tr>
<tr>
<td>Between E-Status</td>
<td>264.50</td>
<td>1</td>
<td>264.50</td>
<td>8.13</td>
<td>.01</td>
</tr>
<tr>
<td>Interaction: B X SP</td>
<td>9.68</td>
<td>1</td>
<td>9.68</td>
<td>.19</td>
<td>NS</td>
</tr>
<tr>
<td>Within Groups</td>
<td>6378.52</td>
<td>196</td>
<td>32.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>6655.12</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**TABLE 6**

Summary Table: 2 X 2 Factorial Analysis of Variance—Briefing and Debriefing X E-Absent and E-Absent-Set

<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>Between Groups</td>
<td>(105.47)</td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Briefing (B)</td>
<td>1.21</td>
<td>1</td>
<td>1.21</td>
<td>.04</td>
<td>NS</td>
</tr>
<tr>
<td>Between Absente (A)</td>
<td>6.25</td>
<td>1</td>
<td>6.25</td>
<td>.21</td>
<td>NS</td>
</tr>
<tr>
<td>Interaction B X A</td>
<td>98.01</td>
<td>1</td>
<td>98.01</td>
<td>3.23</td>
<td>.10</td>
</tr>
<tr>
<td>Within Groups</td>
<td>2917.44</td>
<td>96</td>
<td>30.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3022.91</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 7

Summary Table: 2 X 2 Factorial Analysis of Variance--Briefing and Debriefing X Formal and Informal

<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>Between Groups</td>
<td>(35.15)</td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Briefing (B)</td>
<td>10.89</td>
<td>1</td>
<td>10.89</td>
<td>.31</td>
<td>NS</td>
</tr>
<tr>
<td>Between E-Present (P)</td>
<td>.25</td>
<td>1</td>
<td>.25</td>
<td>.01</td>
<td>NS</td>
</tr>
<tr>
<td>Interaction: B X P</td>
<td>23.98</td>
<td>1</td>
<td>23.98</td>
<td>.69</td>
<td>NS</td>
</tr>
<tr>
<td>Within Groups</td>
<td>3332.56</td>
<td>96</td>
<td>34.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3367.71</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
none of the F values reached significance.

In summary, the statistical analysis performed have demonstrated the presence of significant differences within the variable of experimenter's relative status. Furthermore, the clearest differences within the status variable were found in comparisons of E-Present with E-Absent groups.
Post-Experimental Rating Scale and Questionnaire

Upon completion of the social desirability scale, subjects in all conditions were asked to fill out a combination rating scale and questionnaire. Subjects were instructed to rate the experimenter on a plus 10 to minus 10 scale for being Honest, Professional, Dignified, and Casual. The ratings were summed, minus ratings subtracted from plus ratings, and a group mean and standard deviation was calculated for each item rated.

The three questions sought to determine: 1. Whether the subjects believed the experimenter's instructions regarding the purpose of the experiment ("Yes," "No," or "Don't Know" were the forced choice alternatives) 2. For subjects in the four E-Absent treatments, how they felt about entering the room and finding no one there (Subjects' free choice responses were classified as "Apprehensive" or "Non-Apprehensive"); and 3. If subjects in the two Absent-Set conditions remembered the special instructions placed in the sign-up folder (Answers were classified as "Set" if the instructions were remembered, and "No Set" if they were not).

The results of the rating scale and questionnaire are presented in Table 8 for the eight treatment groups separately, and in Table 9 for the groups combined according to the variables manipulated. Generally, the mean ratings were located in the upper half of the positive side of the scale and none of the mean ratings were negative. Spaces were left blank for many of the questions in Tables 8 and 9 since certain questions pertained only to the E-Absent groups.

Although the subjects were told either before or after completing the
<table>
<thead>
<tr>
<th></th>
<th>Honest</th>
<th>Professional</th>
<th>Dignified</th>
<th>Casual</th>
<th>Was E truthful</th>
<th>Reaction to empty room</th>
<th>Special Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>Mean</td>
<td>9.16</td>
<td>8.52</td>
<td>7.88</td>
<td>4.04</td>
<td>Y=19*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.64</td>
<td>1.77</td>
<td>2.87</td>
<td>5.07</td>
<td>N=0</td>
<td>DK= 6</td>
</tr>
<tr>
<td>BI</td>
<td>Mean</td>
<td>7.72</td>
<td>5.84</td>
<td>5.12</td>
<td>7.72</td>
<td>Y=15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.59</td>
<td>3.84</td>
<td>3.71</td>
<td>2.72</td>
<td>N= 2</td>
<td>DK= 8</td>
</tr>
<tr>
<td>BA</td>
<td>Mean</td>
<td>7.24</td>
<td>4.48</td>
<td>4.16</td>
<td>7.20</td>
<td>Y=14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.76</td>
<td>4.61</td>
<td>3.89</td>
<td>3.50</td>
<td>N= 2</td>
<td>DK= 9</td>
</tr>
<tr>
<td>BAS</td>
<td>Mean</td>
<td>6.32</td>
<td>5.28</td>
<td>5.52</td>
<td>6.24</td>
<td>Y=13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.65</td>
<td>3.94</td>
<td>3.37</td>
<td>3.25</td>
<td>N= 2</td>
<td>DK= 10</td>
</tr>
<tr>
<td>DF</td>
<td>Mean</td>
<td>9.04</td>
<td>8.52</td>
<td>7.64</td>
<td>2.44</td>
<td>Y=24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.93</td>
<td>2.84</td>
<td>2.69</td>
<td>5.78</td>
<td>N= 0</td>
<td>DK= 1</td>
</tr>
<tr>
<td>DI</td>
<td>Mean</td>
<td>8.92</td>
<td>5.80</td>
<td>6.04</td>
<td>8.36</td>
<td>Y=20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.76</td>
<td>4.59</td>
<td>4.82</td>
<td>2.08</td>
<td>N= 0</td>
<td>DK= 5</td>
</tr>
<tr>
<td>DA</td>
<td>Mean</td>
<td>8.04</td>
<td>6.20</td>
<td>6.16</td>
<td>6.92</td>
<td>Y=17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>2.63</td>
<td>3.51</td>
<td>3.38</td>
<td>3.18</td>
<td>N= 1</td>
<td>DK= 7</td>
</tr>
<tr>
<td>DAS</td>
<td>Mean</td>
<td>7.64</td>
<td>4.88</td>
<td>4.00</td>
<td>7.40</td>
<td>Y=16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.53</td>
<td>4.24</td>
<td>3.61</td>
<td>2.96</td>
<td>N= 2</td>
<td>DK= 7</td>
</tr>
</tbody>
</table>

* Y= Yes  
N= No  
DK= Don't Know  
** Ap= Apprehensive  
NAp= Not Apprehensive  
*** S= Set  
NS= No Set  

1. See Appendix B for complete Questionnaire.
### TABLE 9

Post-Experimental Rating Scale and Questionnaire:

Means Only For Combined Groups

<table>
<thead>
<tr>
<th></th>
<th>Honest</th>
<th>Professional</th>
<th>Dignified</th>
<th>Casual</th>
<th>Was E truthful</th>
<th>Reaction to empty room</th>
<th>Special instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formal</td>
<td>9.10</td>
<td>8.52</td>
<td>7.76</td>
<td>3.24</td>
<td>Y=21.5</td>
<td>N= 0.0</td>
<td>DK= 3.5</td>
</tr>
<tr>
<td>Informal</td>
<td>8.32</td>
<td>5.82</td>
<td>5.58</td>
<td>8.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>7.64</td>
<td>5.34</td>
<td>5.16</td>
<td>7.06</td>
<td></td>
<td>Y=15.5</td>
<td>N= 1.5</td>
</tr>
<tr>
<td>Absent-Set</td>
<td>6.98</td>
<td>5.08</td>
<td>4.76</td>
<td>6.82</td>
<td></td>
<td>Y=14.5</td>
<td>N= 2.0</td>
</tr>
<tr>
<td>E-Present</td>
<td>8.71</td>
<td>7.17</td>
<td>6.67</td>
<td>5.65</td>
<td></td>
<td>Y=19.50</td>
<td>N= 0.50</td>
</tr>
<tr>
<td>E-Absent</td>
<td>7.31</td>
<td>5.21</td>
<td>4.96</td>
<td>6.94</td>
<td></td>
<td>Y=15.00</td>
<td>N= 1.75</td>
</tr>
<tr>
<td>Briefed</td>
<td>7.61</td>
<td>6.03</td>
<td>5.68</td>
<td>6.30</td>
<td></td>
<td>Y=15.25</td>
<td>N= 1.50</td>
</tr>
<tr>
<td>Debriefed</td>
<td>8.41</td>
<td>6.35</td>
<td>5.96</td>
<td>6.28</td>
<td></td>
<td>Y=19.25</td>
<td>N= 0.75</td>
</tr>
</tbody>
</table>

1. See Appendix B for complete Questionnaire.
M-C SDS the true purpose of the research, it seemed useful to assess the degree to which subjects accepted this information, not only to determine if subjects believed or disbelieved the experimenter, but also to ascertain if groups would differ. It may be argued that it is difficult or even impossible to assess the subjects' feelings on this matter in a face to face situation. However, since no other tool was available their response to the question and their ratings of the experimenter's honesty were accepted as measures of the examiner's credibility.

The results of the question "Do you believe that the experimenter was telling you the truth in regard to the purpose of this experiment?" were not amenable to conventional non-parametric statistical techniques (since expected values were less than five per cell a Chi Square test could not be used). However Tables 8 and 9 indicate that subjects maintained, generally, high confidence in the examiner's credibility. Of the nine "No" responses (4 1/2% of the sample of 200), six came from subjects in Briefed conditions. This finding is not unlike that observed in the author's pilot study in which all but one of the eight "No" answers (which constituted 13% of the sample of 62) to the same question came from subjects in Briefed conditions.

Regarding the ratings of the item Honest, application of a 2 X 4 factorial analysis of variance to the ratings, summarized in Table 10, produced a significant F of 5.05 (P < .01) for the main effects of status and a significant F of 3.90 (P < .05) for the main effects of pre-experimental briefing. A Duncan's Multiple Range test, applied to locate the source of the significance, produced the following significant comparisons: BF-BAS
TABLE 10
Summary Table: 2 X 4 Factorial Analysis of Variance—Honest

<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>Between Groups</td>
<td>(172.06)</td>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Status (S)</td>
<td>124.10</td>
<td>3</td>
<td>41.37</td>
<td>5.05</td>
<td>&lt; .01</td>
</tr>
<tr>
<td>Between Briefing (B)</td>
<td>32.00</td>
<td>1</td>
<td>32.00</td>
<td>3.90</td>
<td>&lt; .05</td>
</tr>
<tr>
<td>Interaction: S X B</td>
<td>15.96</td>
<td>3</td>
<td>5.32</td>
<td>.65</td>
<td>NS</td>
</tr>
<tr>
<td>Within Groups</td>
<td>1573.92</td>
<td>192</td>
<td>8.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1745.98</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(9.16-6.32=2.84, \( R_{p6} = 2.37, P < .01 \)); BF-BA (9.16-7.24=1.92, \( R_{p7} = 1.82, P < .05 \)); DF-BAS (9.04-6.32=2.72, \( R_{p7} = 2.34, P < .01 \)); DF-BA (9.04-7.24=1.80, \( R_{p6} = 1.80, P < .05 \)); and DI-BAS (8.92-6.32=2.60, \( R_{p6} = 2.31, P < .01 \)).

Acceptance of the experimenter's statement regarding the purpose of the research might best be represented by ranking the eight groups, from high to low, on the mean rating for Honest and comparing the mean ratings with the number of "No" responses to the question. Such an ordering is revealed in Table 11 where an inverse relationship exists between the rated item and the question. The results of the Duncan's Multiple Range test and Table 11 suggest that not only was the examiner regarded as credible but also E-Present groups, except for the BI group, maintained this attitude to a higher degree than E-Absent groups.

In the DF and BF conditions the experimenter dressed and attempted to act Formal while in the DI and BI treatments he wanted to appear Informal. Was the experimenter actually perceived as Formal and/or Informal in these roles, or were they ineffective manipulations? To answer this question the ratings of three items, Dignified, Professional, and Casual were analyzed. Success in the Formal conditions would be found if the experimenter was rated higher on the Dignified and Professional items than on the Casual. Likewise, success in the Informal conditions would be probable if the ratings of Casual were higher than those of Dignified and Professional.

Regarding the ratings of Dignified, Table 9 reports mean ratings of 7.76 and 5.58 by subjects in the Formal and Informal conditions, respectively. A 2 X 4 factorial analysis of variance was calculated (Table 12) and the main effects of status (\( F=5.90, P < .01 \)) and the interaction effects
TABLE 11

Relationship Between Number of "No" Responses to Question and Mean Rating of Honest

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Rating of Honest</th>
<th>Number of &quot;No&quot; Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>9.16</td>
<td>0</td>
</tr>
<tr>
<td>DF</td>
<td>9.04</td>
<td>0</td>
</tr>
<tr>
<td>DI</td>
<td>8.92</td>
<td>0</td>
</tr>
<tr>
<td>DA</td>
<td>8.04</td>
<td>1</td>
</tr>
<tr>
<td>BI</td>
<td>7.72</td>
<td>2</td>
</tr>
<tr>
<td>DAS</td>
<td>7.64</td>
<td>2</td>
</tr>
<tr>
<td>BA</td>
<td>7.24</td>
<td>2</td>
</tr>
<tr>
<td>BAS</td>
<td>6.32</td>
<td>2</td>
</tr>
</tbody>
</table>
### TABLE 12

Summary Table: 2 X 4 Factorial Analysis of Variance: Dignified

<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>Between Groups</td>
<td>(359.19)</td>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Status (S)</td>
<td>238.23</td>
<td>3</td>
<td>79.14</td>
<td>5.90</td>
<td>.01</td>
</tr>
<tr>
<td>Between Briefing (B)</td>
<td>4.20</td>
<td>1</td>
<td>4.20</td>
<td>.31</td>
<td>NS</td>
</tr>
<tr>
<td>Interaction: S X B</td>
<td>116.76</td>
<td>3</td>
<td>38.92</td>
<td>2.89</td>
<td>.05</td>
</tr>
<tr>
<td>Within Groups</td>
<td>2584.96</td>
<td>192</td>
<td>13.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2944.15</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of status and debriefing (F=2.89, P<.05) were found significant. Subsequently, A Duncan's Multiple Range test conducted on the data of the combined four status groups, yielded significant differences for the comparisons of Formal with Informal (7.76-5.58=2.18, Rp₂=1.88, P<.01).

The mean ratings for Professional were 8.52 and 5.82 for the Formal and Informal groups respectively. The results of a 2 X 4 factorial analysis of variance performed on the ratings are found in Table 13 where an F of 8.41 for the main effects of status was found significant beyond the .01 level. The difference of 2.70 between Formal and Informal groups was found significant beyond .01 by a Duncan's Multiple Range test (Rp₄=2.07).

On the basis of this analysis it appears then that the Formal groups did perceive the experimenter as Dignified and as Professional to a significantly greater degree than did Informal groups.

However, it remains to determine the examiner's informality in the Informal conditions. On the item Casual, Informal groups achieved a mean of 8.04 as compared with the mean of 3.24 obtained by the Formal groups. The 2 X 4 factorial analysis of variance conducted (Table 14) produced an F of 15.06 (P<.01) for the main effects of status. A Duncan's Multiple Range test revealed that the difference of 4.80 between the Informal and Formal groups reached significance at the .01 level (Rp₄=2.09). Thus, there is evidence which indicates that the Informal groups perceived the experimenter significantly more informal than did Formal groups.

In the analyses of the success of the Formal and Informal manipulations, data from Absent and Absent-Set groups were included because of the author's
### TABLE 13

Summary Table: 2 X 4 Factorial Analysis of Variance—Professional

<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>Between Groups</td>
<td>(410.40)</td>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Status (S)</td>
<td>371.38</td>
<td>3</td>
<td>123.79</td>
<td>8.41</td>
<td>.01</td>
</tr>
<tr>
<td>Between Briefing (B)</td>
<td>4.81</td>
<td>1</td>
<td>4.81</td>
<td>.33</td>
<td>NS</td>
</tr>
<tr>
<td>Interaction: S X B</td>
<td>34.21</td>
<td>3</td>
<td>11.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>2852.76</td>
<td>192</td>
<td>14.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3263.16</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 14

Summary Table: 2 x 4 Factorial Analysis of Variance: Casual

<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>Between Groups</td>
<td>(716.86)</td>
<td>(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between Status (S)</td>
<td>661.94</td>
<td>3</td>
<td>220.65</td>
<td>15.06</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Between Briefing (B)</td>
<td>.02</td>
<td>1</td>
<td>.02</td>
<td>0.00</td>
<td>NS</td>
</tr>
<tr>
<td>Interaction: S X B</td>
<td>54.90</td>
<td>3</td>
<td>18.30</td>
<td>1.25</td>
<td>NS</td>
</tr>
<tr>
<td>Within Groups</td>
<td>2812.12</td>
<td>192</td>
<td>14.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3529.18</td>
<td>199</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
contention stated earlier that E-Absent groups act as control groups for the experimenter variable studies and provide a necessary basis for comparing E-Present with E-Absent groups. The significant comparisons reported between Formal and Informal groups for the Duncan's tests do not account for all the variance among groups because significant differences were also found between certain E-Present (Formal and Informal) and certain E-Absent (Absent and Absent-Set) groups for ratings on Professional, Dignified, and Casual. Such significant differences are not reported since in the Absent and Absent-Set conditions subjects could not be affected by the Experimenter's dress and manner until the SDS was completed.

Final questions pertinent to the success of the experimental manipulations center on the Absent and Absent-Set groups. First, did subjects in the BAS and DAS groups actually remember (form a set) the instructions which informed them of the examiner's absence? The data reported in Table 8 show that 22 of the 25 BAS subjects and 24 of the 25 DAS subjects were able to recall the instructions typed on the sign-up folder. A Fisher Exact Probability Test (Siegel, 1956) compared BAS with DAS subjects according to the Set-No Set dichotomy. The probability that these groups differed by chance on the number of subjects forming the set was .25. At "one even more extreme" the probability was .12. Thus, not only was there near unanimous recollection of the instructions, but also there were no significant differences between BAS and DAS groups.

The purpose of introducing the Absent-Set groups into the study was based upon the hypothesis that subjects might feel apprehensive about participating in the experiment in the absence of an experimenter. This
hypothesis received some support in the author's pilot study as mentioned earlier. Apprehension in the present study was measured by a statement placed in the post-experimental questionnaire and rating scale (Tables 8 and 9). To the question "In one or two words tell me how you felt about walking into a room and finding no one there.", responses were dichotomized as "Apprehensive" (Ap) or "Not Apprehensive" (NAP). A Chi Square one sample test (Siegel, 1956) was performed on these scores for each of the four E-Absent groups in order to determine if the answers differed from chance. From Table 15 it is observed that none of the Chi Square tests were significant, indicating that the number of Ap subjects in the BAS and DAS groups did not differ from chance. Had the Absent-Set conditions actually reduced subjects' degree of apprehension then significant Chi Square values should have been found for the BAS and DAS groups.

A Chi Square test for k independent samples (Siegel, 1956) compared mean Ap with NAP scores of the Absent groups (BA, DA) with those of the Absent-Set groups (BAS, DAS) in a further attempt to search for significant differences. However, Table 15 shows that the value of .44 was not significant (df=1, P<.70).

In summary, data from the post-experimental rating scale and questionnaire suggest the following conclusions: 1. Subjects in all groups appraised the examiner as honest and believed his instructions; 2. The experimenter was successful in appearing Formal and Informal; 3. Prior knowledge of the experimenter's absence did not significantly affect subjects' apprehensions.
<table>
<thead>
<tr>
<th>Comparison</th>
<th>Chi Square</th>
<th>df</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA with Expected</td>
<td>3.24</td>
<td>1</td>
<td>.10- NS</td>
</tr>
<tr>
<td>BAS with Expected</td>
<td>1.00</td>
<td>1</td>
<td>.50</td>
</tr>
<tr>
<td>DA with Expected</td>
<td>1.00</td>
<td>1</td>
<td>.50</td>
</tr>
<tr>
<td>DAS with Expected</td>
<td>.004</td>
<td>1</td>
<td>.95</td>
</tr>
<tr>
<td>Absent with Absent-Set</td>
<td>.44</td>
<td>1</td>
<td>.70</td>
</tr>
<tr>
<td>BA with BAS with DA with DAS</td>
<td>2.10</td>
<td>3</td>
<td>.70</td>
</tr>
</tbody>
</table>
Chapter V

Discussion

The hypothesis that Formal (BF and DF) groups would attain higher SDS scores than Informal (BI and DI) groups was not supported. While the DF mean of 16.04 was the highest of the eight groups, a Duncan's Multiple Range test showed that his mean did not differ significantly from either the BI ($\bar{X}=15.48$) or from the DI ($\bar{X}=15.16$) groups which attained the second and third highest means. In fact, the mean of 15.32 for the two Informal groups is slightly higher than that of the two Formal groups ($\bar{X}=15.22$). Yet, analysis of the post experimental rating scale and questionnaire suggested that the intended manipulations of Formal and Informal status were successful, since, as Table 9 shows, the examiner was rated more ($P < .01$) professional and dignified by Formal than by Informal groups, and more ($P < .01$) casual by Informal than by Formal groups.

The evidence used to generate this hypothesis was based on Birney's findings that mean n Ach TAT scores were greater for groups run by a faculty experimenter than by a student experiment. However, a closer look at the Birney study raises a question. Could the differing roles of the examiners have accounted for the dissimilar findings? In the Birney investigation the subjects assigned to the faculty experimenter were enrolled in a course taught by the faculty experimenter. This means that there were at least two aspects of the subject-examiner relationship: subjects' relation to a person perceived as a faculty member; and subjects' relation to a known faculty member who, as subjects' instructor at the time, exercised some control over them outside the research setting. In the present investigation, however,
only the first aspect of the experimenter-subject relationship existed since
the examiner, a part-time faculty member, was presumed to be unknown to
subjects at the time of the study. It is plausible to assume that subjects
would endeavor to act more favorably (assuming that the SDS and the n Ach
TAT both measure to some degree social desirability) to a faculty member
who knows and exercises academic control over them than to one who does not.
Thus, differing roles could have accounted for the differences between
the Birney and the present investigations.

A second alternative explanation for the dissimilar findings concerns
the number of examiners used. Birney assumed the role of experimenter in
the faculty condition while a student acted as the student experimenter.
This procedure permitted both experimenter-role and experimenter-personality
variables to operate simultaneously. Consequently, the findings could have
been peculiar to the examiner's personality rather than the role played, or
an interaction of the two. To prevent role from interacting with personality,
each experimenter should have performed at least once in each role using
students unknown to either experimenter. When one experimenter assumed both
roles in the present investigation such contamination was prevented.

The second hypothesis stated that DA subjects would receive lower SDS
scores than DAS subjects. A re-examination of the circumstances in which
this hypothesis was generated suggests that it was predicated on a faulty
assumption.

In the author's pilot study DA subjects were questioned regarding
their feelings about entering the booth and finding no one present. Many
of them, estimated (the writer had not intended, originally, to seek precise
information regarding this particular question) as more than half but not two thirds, indicated that they had felt "anxious," "strange," "apprehensive," or "odd," just to list some of the actual comments. The remaining subjects evoked such statements as "comfortable," "expected anything," or "not bothered." Comments denoting and connoting apprehension and discomfort suggested to the author that the low DA scores were due partially to the inhibition of socially desirable responses from some subjects. This hypothesis, having formed the basis for introducing the DAS group on the supposition that prior knowledge of the examiner's absence would lower apprehension, received some support in the pilot study. Consequently, the DAS and BAS groups were introduced into the present study to reduce subjects' presumed apprehension.

The faulty assumption can be traced to the writer's original debriefing in the pilot study and his subsequent classification of the DA responses into apprehensive versus non-apprehensive. Chance alone would dictate that in a two choice situation in which one classification is not inherently favored (there was no reason to assume that most subjects would be prone to feel either apprehensive or non-apprehensive), about half the responses would be expected to fall in each category. The finding that between half and two thirds of the pilot study subjects felt apprehensive, was expected by chance. Thus, the data emerging from this investigation indicating that the number of apprehensive subjects in each of the four E-Absent groups was not significantly different (see Table 15) merely confirm chance expectations. Therefore, BAS and DAS groups would not have been needed for this research.
While the above discussion attempts to explain why the hypotheses were not confirmed there remains the problem of interpreting the findings regarding pre-experimental briefing and the E-Present versus E-Absent differences. When the combined Briefed and combined Debriefed groups are compared their data appears similar on all but one measure. The factorial analysis of variance applied to the M-C SDS scores yielded a non-significant result \((F=.07)\) for the main effects of pre-experimental briefing (Table 5). That is, informing subjects as to the true nature of the experiment before they began the task had the same effect as the commonly employed technique of debriefing. This finding may have far reaching implications for research methodology in view of the common laboratory practice of deceiving subjects about the examiner's intentions. While deception was not employed (although slight alterations or omissions could have rendered the instructions deceiving) the data of this study are pertinent to Kelman's statement regarding the use of deception in psychological research.

A basic assumption in the use of deception is that a subject's awareness of the conditions that we are trying to create and of the phenomenon that we wish to study would affect his behavior in such a way that we could not draw valid conclusions from it [p. 6]. Probably most researchers would agree that in many social and personality research investigations subjects' awareness of the researcher's intents would differentially affect subjects' behavior, even when deception is not used. Yet, the present study indicates that Kelman's assumption is not always true. Briefed groups were instructed that sometimes the examiner acted formal, sometimes informal, and sometimes he is absent until the end of the experiment. Briefed groups were instructed also that the
information they were receiving was one of the experimental treatments, and that the task they would perform measures their willingness to participate. Still, these instructions did not cause their scores to differ from the Debriefed groups who did not have prior knowledge of the experimental manipulations. Perhaps Kelman's assumption should be held in abeyance until further research uncovers the types of investigations which are not sensitive to pre-experimental briefing.

Although the existential nature of Kelman's assumptions might be conceded one might question the value of briefing subjects. One answer to this inquiry, suggested by the work of Orne (1962), is that briefing may serve to reduce "demand characteristics." According to Orne, a subject in a psychology experiment usually wants to be a "good subject" and ordinarily complies with the wishes of the examiner. However, a good subject is not necessarily a passive and intellectually dormant individual, Orne continues, for he usually enters the situation with certain knowledge, attitudes, and expectancies, e.g., the knowledge that it is a psychology experiment, the belief that the experimenter may not be truthful with him, attitudes based on campus rumors, his impressions of the experimenter's original solicitation, etc. The subject, accustomed to receiving inaccurate or incomplete information in other experiments, develops his own hypotheses about the true purpose of the research. According to Orne, "the totality of cues which convey an experimental hypothesis to the subject become significant determinants of the subjects' behavior. We have labelled the sum total of such cues as the 'demand characteristics' of the experimental situation [p. 779]."
Consequently, Orne, notes, a subject's behavior will be influenced not only by the independent variable manipulated but also by the perceived demand characteristics of the situation which can operate as contaminating extraneous variables. If demand characteristics were eliminated or at least reduced then contamination would likewise disappear and more confidence could be placed in the findings. It is being suggested that the practice of briefing subjects in certain areas of research may alleviate the problem of demand characteristics. If this were the policy in psychology laboratories subjects might become less distrustful of psychologists (Kelman and Orne have clearly enunciated the distrustful attitudes which subjects tend to have towards psychologists). To reduce suspicion and distrust a few generations of psychology students may have to pass through our laboratories but the pool of naive and trusting subjects should, subsequently, grow rather than, as Kelman maintains, gradually decrease.

This simple suggestion is not without its complexities, however. First, how are demand characteristics to be identified, assessed, and controlled? Second, how can the subjects be convinced of the examiner's honesty? Perhaps the latter question will be elucidated by an examination of the data regarding the experimenter's honesty obtained in the present study. In Table 9 means of 8.41 and 7.61 were obtained for the Debriefed and Briefed groups' ratings of Honest. A factorial analysis of variance produced a significant (P < .05) main effect for briefing indicating that Briefed subjects did rate the examiner significantly lower than did Debriefed subjects. Likewise, in answer to the question inquiring if subjects believed the instructions, six of the nine "No" responses were evoked by Briefed subjects. In the pilot
study seven of the eight "No" answers to the same question were also traced to Briefed subjects, implying that some individuals couldn't believe that they would receive such information. This trend is corroborated by Orne who notes that even if the experimenter is honest with subjects "more often than not he will be distrusted [p. 779]." Kelman seems to concur. He reports that in his role playing research when the experimenter told subjects before the session began that they would only be pretending to administer shock and that they should react as if they were really administering it, some subjects "did not accept as true the information that this was all make-believe and wanted to know when they should show up for the shock experiment to which they had committed themselves [p. 10]."

Are subjects' suspicions so pervasive that they can not recognize truth? If this is so perhaps such techniques of reducing "demand characteristics" and eliminating distrust as briefing will be of little value. However, if the answer to this question is "Yes" then such techniques appear all the more urgently needed in psychological research.

The SDS was chosen for its presumed sensitivity to subjects' reaction to a situation. Although it was but one of many available instruments a cursory analysis of the findings as related to the M-C SDS seems appropriate. Crowne and Marlowe (1960) broadly defined social desirability as the "need of subjects to obtain approval by responding in a culturally appropriate and acceptable manner [p. 353]." Thus, high scores are said to indicate a strong need for approval while low scores imply a lower level. If it is assumed that randomization was successful, i.e., that those subjects high and low in social desirability or need for approval were approximately
equally distributed throughout the eight treatments, the specific problems
relating to the uniqueness of high or low scoring individuals, e.g., as
pointed out in Miller, Doob, and Butler (1965), are presumed to be eliminated.

Since the only significant differences on SDS scores were found in
the E-Present versus E-Absent comparisons, the question arises: How does
the examiner's presence relate to these scores? The means of the E-Present
groups and E-Absent, respectively, were 15.27 and 12.97, as compared to a
mean of 15.06 obtained by Crowne and Marlowe (1964) in a classroom situation.
Applying the Crowne and Marlowe hypothesis, subjects serving in the E-Absent
conditions manifested an overall lower level of social desirability and need
for approval than those in E-Present conditions. Why this difference exists
is not clear. One plausible explanation is that social desirability may be
a trait or characteristic, the normal manifestation of which must occur in
the presence of an individual regarded as the object of the "need" or who
represents someone or some thing that is the object. Recall that the E-
Present mean of 15.27 (N=100) is more like the Crowne and Marlowe obtained
mean of 15.06 (N=666) than it is like the E-Absent mean of 12.97. In the
first two cases there was someone visibly present to the subjects. That
the status of this individual is irrelevant is borne out by the nearly
identical means obtained by the Formal and Informal groups in this investi-
gation. Although Crowne and Marlowe do not mention the status of the
experimenter or test administrator used to obtain their data, these scores
were obtained from a general psychology class on the first day of the
semester, implying that the instructor or an assistant was present. In
both the Crowne and Marlowe and the E-Present situations there was someone
present, who if not the specific object of subjects' "need" for approval was at least representative of someone or something (the institution) that was. If in the E-Absent treatments there was an object of the need for approval or social desirability, it was not clear to the subjects. Perhaps the definition of social desirability must be extended to include absent experimenter situations.

Admittedly, the relation of experimenter's presence to social desirability is not clear at this time. However, some insight into this problem might be gained if the meaning of the E-Present - E-Absent differences are examined apart from the M-C SDS scores as Marlowe and Crowne interpret them.

The higher scores attained by E-Present groups on the M-C SDS could be interpreted in the light of Zajonc's (1965) integration of social facilitation studies, social facilitation referring to the effect of sheer presence of individuals upon the behavior of others. Although Social Facilitation research originated with Triplett's 1897 experiments it died suddenly at the outbreak of World War II. Zajonc dichotomizes the social facilitation research into the audience effects paradigm, the observation of behavior when it occurs in the presence of passive spectators, and the co-action effects paradigm, the observation of behavior when it occurs in the presence of others who are performing the same activity.

In the audience effects paradigm, either subjects perform in both spectator-present and spectator-absent conditions, or else two groups are assigned to one condition each and perform in front of a number of spectators or in front of a single spectator. Zajonc's summary of the literature regarding audience effects studies showed that maze and nonsense syllable
learning tasks were inhibited by the sheer presence of others while pursuit rotor performance, word association, multiplication, and vigilance tasks improved in the presence of others. He simplifies these findings into the generalization that "performance is facilitated and learning is impaired by the presence of others [p. 270]." He further reasons that in the early stages of learning such tasks as nonsense syllables and finger mazes (which are inhibited by the presence of others) there are more incorrect than correct responses. As training progresses the task is mastered and the correct response becomes dominant. On the basis of this analysis Zajonc modifies his generalization to read: "Audience enhances the emission of dominant responses [p.270]."

If the Zajonc hypothesis is to be applied to the findings of the present research it becomes necessary to assume that socially desirable responses were dominant in the experimental setting. This assumption is tenable since the subjects were students who volunteered for this particular experiment in order to fulfill the requirements of a psychology course, having been told in advance by their instructors that cooperation was expected. Although the tasks mentioned in Zajonc's review involve the learning of perceptual-motor skills, or the memorizing of lists, it is also assumed that the development of social desirability responses follow a process of replacement of incorrect responses with dominant correct responses.

In the present study the four E-Present groups did complete the M-C SDS in the presence of one person, the experimenter, who read a book during most of the session while the subject was seated to the front and left.
While it could be argued that the term "passive spectator" used by Zajonc may not be equivalent to the "experimenter reading a book," the differences appear slight and of only minor importance compared to the fact that an audience of one was present. Thus, the findings lend support to the social facilitation hypothesis. Not only were all four E-Present groups higher on the M-C SDS than the four E-Absent groups, but the difference between E-Present and E-Absent was significant ($P < .01$).

Zajonc suggests that the social facilitation hypothesis ties together a number of studies. What is needed next is to examine why the presence of others enhances the emission of dominant responses. Although Zajonc attempts to answer this question with the statement that the presence of others increases the individual's general arousal or drive level, he quickly admits that the supporting evidence is indirect and that negative evidence exists. What is needed, he notes, is the specification and systematic investigation of the parameters involved in social facilitation.

In summary, the social facilitation hypotheses is congruent with the findings of this experiment. However, it does not explain why scores were differentially affected by the presence or absence of the examiner.

Since the social facilitation hypothesis has suggested that task performance and test scores may be a function of the presence or absence of an audience, one might question the value of designs incorporating E-Present and E-Absent groups. Earlier it was argued that the introduction of the Absent Experimenter group into experimenter effects designs would serve as a control and as a basis for comparing the effects of the experimenter variable being manipulated. Were the Absent Experimenter groups' results
similar to those of any of the treatment groups, the operation of examiner
effects could not necessarily be ascribed to the particular treatments.

Although the data of this research have emphasized the importance of
the experimenter presence variable over that of status (Formal and Informal
groups, specifically), this is not typically the case. Rosenthal's program
consistently points to the experimenter as a significant source of variability
and as an extraneous variable requiring control. Such research has been
primarily concerned with identifying the relevant variables and the circum-
stances in which they can be contaminating. One technique employed in this
study for controlling the effects of the experimenter has been generally
ignored by Rosenthal and his associates, namely, the use of the Absent
Experimenter groups. Without an experimenter how can there be experimenter
effects? It was noted earlier that the examiner can never be completely
eliminated from any experiment since the communications to subjects are
made through some media, printed, visual, or auditory, which convey
particular styles and force responses to be made in a certain fashion. But
if the Absent Experimenter groups cannot completely eliminate the experimenter,
their use can certainly minimize his influence. How can the many variables
including experimenter sex, race, experience, expectancy, anxiety level,
status, and modeling be conveyed when the subject comes face to face with
only a carefully written clear and concise set of instructions directing him
to complete a certain task? The present study has demonstrated that subjects
can arrive at a predesignated room by following written instructions and
clearly printed signs, that he can enter the experimental booth, be seated,
and subsequently read the instructions to the task. If subjects can find
their way to the testing area why have a potentially contaminating experimenter present?

The Marlowe-Crowne Social Desirability Scale, which consisted of a brief set of instructions for completing the thirty-three items, is one of many paper and pencil tasks frequently used in psychological research. Could not personality inventories, anxiety tests, attitude scales, and the like also be administered in the absence of an examiner? Likewise, a little electrical and mechanical know-how combined with clear instructions and illustrations could permit subjects to memorize a list of words presented on a memory drum, perform such perceptual motor tasks as rotary pursuit, mirror tracing, and reaction time, all in the absence of an experimenter. Probably the examiner would make himself known upon completion of the session, as occurred in the present study. He would also function to handle apparatus breakdown or subjects' failure to follow instructions, situations which frequently terminate in the discarding of data. Thus, the use of the Absent Experimenter is being proposed in situations where the experimenter is a known and uncontrolled relevant extraneous variable and where subjects can successfully complete the task with only the aid of printed instructions and drawings. What may result is research specific only to situations where there was no experimenter present, but the work if properly conducted would avoid the pervasive problem of individual differences among experimenters and concomitant experimenter influence on results. Reliability could be gained by nearly maximum standardized conditions and at little or no cost to validity. Would research based on the Absent Experimenter be any less valid than research where the experimenter is a potential source of contamination? The
Absent Experimenter control group is not deemed the solution to the problems besetting psychological research, for further examination and investigation will no doubt unravel complexities which characterize any variable which is systematically studied. However, the Absent Experimenter control group may improve the methodology currently used in psychological research.

As this investigation was exploratory and attempted to determine the existence of basic relationships among the levels of status and pre-experimental briefing, certain information was sacrificed for the sake of expediency. First, in a true experimenter-effects study the employment of multiple experimenters is requisite so that the findings may be generalized beyond one examiner. This study does not allow for such generalizations. Since the same experimenter administered all treatments it is possible that the findings may be unique to the examiner. Would the differences between E-Present and E-Absent groups be more or less pronounced had a second examiner been employed? Likewise, could a different examiner act more formal and more informal? Is the similarity between Briefed and Debriefed groups universal with this type of experiment or was the particular examiner a strong influence? These questions could have been answered had two or three more experimenters been employed.

In addition, the influence of experimenter and subject sex on the status and pre-experimental briefing conditions is unknown. Male and female experimenters administering the treatments to male and female subjects could have provided such information.

In addition to the limitations imposed by the use of one experimenter using only male subjects, the present investigation was restricted to one
task, the Marlowe-Crowne Social Desirability Scale. Personality inventories or perceptual-motor tasks could be introduced, possibly in a counterbalanced order, to ascertain their relationship to the status and pre-experimental briefing conditions.

Finally, one limitation was the amount of time that subjects spent in the setting. Since it was the experimenter's intention to elicit behavior reactive to the situation, subjects could have been permitted to spend some time in the booth before having to complete the scale. The average time elapsing between the point when they entered the booth and the moment when they began completing the social desirability scale was about one minute for Debriefed subjects and approximately two minutes for Briefed subjects. This may not be sufficient to enable subjects to adjust to the setting and develop some kind of feeling or overall attitude towards the experimental situation. The use of additional tasks as noted above would lengthen the time that a subject spends in the experiment.
Chapter VI

Summary

This investigation tested two methodological improvements suggested for experimenter effects research through the introduction of Absent Experimenter control groups and Briefed groups. While this situation permits the use of briefing without disrupting SDS scores, status is best differentiated not in terms of Formal and Informal but Experimenter-Present and Experimenter-Absent, perhaps because of social facilitation. Briefing is implicated as a means of reducing "demand characteristics" and subject distrust, while the use of Absent Experimenter groups may become a means of reducing experimenter effects.
References


Appendix A

Marlowe-Crowne Social Desirability Scale

PERSONAL REACTION INVENTORY

Listed below are a number of statements concerning personal attitudes and traits. Read each item and decide whether the statement is true or false as it pertains to you personally. Be sure to answer each item even if it does not pertain to you.

1. Before voting I thoroughly investigate the qualifications of all the candidates.
2. I never hesitate to go out of my way to help someone in trouble.
3. It is sometimes hard for me to go on with my work if I am not encouraged.
4. I have never intensely disliked anyone.
5. On occasion I have had doubts about my ability to succeed in life.
6. I sometimes feel resentful when I don't get my way.
7. I am always careful about my manner of dress.
8. My table manners at home are as good as when I eat out in a restaurant.
9. If I could get into a movie without paying and be sure I was not seen I would probably do it.
10. On a few occasions, I have given up doing something because I thought too little of my ability.
11. I like to gossip at times.
12. There have been times when I felt like rebelling against people in authority even though I knew they were right.
13. No matter who I'm talking to, I'm always a good listener.
14. I can remember "playing sick" to get out of something.
15. There have always been occasions when I took advantage of someone.
16. I'm always willing to admit it when I make a mistake.
17. I always try to practice what I preach.
18. I don't find it particularly difficult to get along with loud mouthed, obnoxious people.
19. I sometimes try to get even rather than forgive and forget.
20. When I don't know something I don't at all mind admitting it.
21. I am always courteous, even to people who are disagreeable.
22. At times I have really insisted on having things my own way.
23. There have been occasions when I felt like smashing things.
24. I would never think of letting someone else be punished for my wrongdoings.
25. I never resent being asked to return a favor.
26. I have never been irked when people expressed ideas very different from my own.
27. I never make a long trip without checking the safety of my car.
28. There have been times when I was quite jealous of the good fortune of others.
29. I have almost never felt the urge to tell someone off.
Appendix A (cont'd):

30. I am sometimes irritated by people who ask favors of me.
31. I have never felt that I was punished without cause.
32. I sometimes think when people have a misfortune they only got what they deserved.
33. I have never deliberately said something that hurt someone's feelings.
Appendix B

Post-Experimental Rating Scale and Questionnaire

RATING SCALE OF THE EXPERIMENTER AND THE EXPERIMENT

Please rate the experimenter in the experiment in which you have just participated on the following scale. Notice that the scale ranges from -10 to +10. If you feel that the experimenter was high on the characteristic in question please rate him appropriately high on the + end of the scale by circling the number of your choice. If you feel that the experimenter was low on a characteristic in question, do the same thing on the - end of the scale.

Be sure to rate the experimenter's behavior in accordance with the way he acted in the particular experimental condition you served in.

1. honest - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
2. friendly - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
3. personal - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
4. casual - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
5. courteous - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
6. business-like - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
7. professional - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
8. behaved consistently - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
9. pleasant - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
10. dignified - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
11. warm - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +

Using the same rating procedure, please evaluate your feeling about the experiment. What was your reaction to taking part in this research?

1. liked - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
2. annoyed - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
3. interested - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
4. suspicious - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +
Appendix B (cont'd):

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

Using the same rating procedure followed earlier, please evaluate your feelings regarding the following statements.

1. I would recommend that a friend take part in this experiment.
   - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +

2. My feeling towards psychology as an experimental science is
   - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +

3. The experimenter showed consideration towards the subject.
   - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +

4. I would participate in another study of this type.
   - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +

5. I feel toward the information gained by this experiment
   - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +

6. My feeling towards being a subject in an experiment without knowing that it was an experiment would be
   - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +

7. The amount of explanation which should follow an experiment should be
   - 10 9 8 7 6 5 4 3 2 1 - + 1 2 3 4 5 6 7 8 9 10 +

How many psychology experiments have you participated in prior to this one?

Do you believe that the experimenter was telling you the truth in regard to the purpose of this experiment? Yes No Don't know

If you served in one of the Absent Experiment conditions:

a. In one or two words tell me how you felt about walking into a room and finding no one there.
Appendix B (cont'd):

b. Did you think that you were being watched through the one way mirror on the door? Yes  No  Don't know  Wasn't aware of mirror

c. For some Absent Experimenter subjects there were special instructions in the sign-up folder telling what to expect when they arrive for the experiment. Briefly, what were these instructions?
APPROVAL SHEET

The dissertation submitted by Paul I. Hettich has been read and approved by members of the Department of Psychology.

The final copies have been examined by the director of the dissertation and the signature which appears below verifies the fact that any necessary changes have been incorporated and that the dissertation is now given final approval with reference to content and form.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[Signature]

Date

Signature of Adviser