FRIEDMAN, Barry Aaron, 1952-
THE EFFECT OF FIRST IMPRESSIONS AND SENSITIVITY ON RESPONSE BIAS AND BEHAVIORAL ASSESSMENT: A SIGNAL DETECTION THEORETICAL APPROACH.

The Ohio State University, Ph.D., 1977
Psychology, industrial

Xerox University Microfilms, Ann Arbor, Michigan 48106
THE EFFECT OF FIRST IMPRESSIONS AND SENSITIVITY ON
RESPONSE BIAS AND BEHAVIORAL ASSESSMENT:
A SIGNAL DETECTION THEORETICAL APPROACH

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Barry A. Friedman, M.A.

* * * * *

The Ohio State University
1977

Reading Committee:
Dr. Milton D. Hakel
Dr. Richard J. Klimoski
Dr. Edwin T. Cornelius

Approved By
Milton D. Hakel
Advisor
Department of Psychology
ACKNOWLEDGEMENTS

My deepest appreciation goes to Dr. Milton D. Hakel for his direction, ideas and insights. My exchanges with Milt have provided countless contributions to my graduate education, for which I am indebted. Special thanks are also due Dr. Richard J. Klimoski, whose rigor and conceptual skills will serve as worthy professional standards. I would also like to thank Drs. Edwin T. Cornelius and Bob Billings for supplying me with the fundamentals of the field from which to build.

Appreciation is also acknowledged to my fellow graduate students, especially those with whom I have shared four years of growth and development. Thanks is also due Tom Davison for his expertise and encouragement with respect to the present research.

A warm appreciation is afforded my parents and family for their encouragement and love. I express my love and deep appreciation to Maureen for her patience and adeptness at coping with my graduate career in addition to her own.
VITA


1973-1974. Graduate Teaching Associate, Ohio State University

1974-1975. Graduate Research Associate, Ohio State University

1975. M.A., The Ohio State University, Columbus, Ohio

1975-1976. Graduate Teaching Associate, Ohio State University

1977. Graduate Research Associate, Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Industrial and Organizational Psychology
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Part I</td>
<td>3</td>
</tr>
<tr>
<td>Signal Detection Theory</td>
<td>4</td>
</tr>
<tr>
<td>Observer Characteristics</td>
<td>10</td>
</tr>
<tr>
<td>Factors Influencing Bias</td>
<td>11</td>
</tr>
<tr>
<td>First Impressions</td>
<td>11</td>
</tr>
<tr>
<td>Observer Sensitivity</td>
<td>12</td>
</tr>
<tr>
<td>Observational Accuracy</td>
<td>12</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>17</td>
</tr>
<tr>
<td>Part II</td>
<td>18</td>
</tr>
<tr>
<td>Order Effects</td>
<td>18</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>20</td>
</tr>
<tr>
<td>Summary</td>
<td>20</td>
</tr>
<tr>
<td>METHODOLOGY</td>
<td>22</td>
</tr>
<tr>
<td>Part I</td>
<td>22</td>
</tr>
<tr>
<td>Subjects</td>
<td>22</td>
</tr>
<tr>
<td>Experimental Design</td>
<td>22</td>
</tr>
<tr>
<td>Procedure</td>
<td>25</td>
</tr>
<tr>
<td>Behavior Training</td>
<td>25</td>
</tr>
<tr>
<td>First Impression Manipulation</td>
<td>27</td>
</tr>
<tr>
<td>Stimulus Material</td>
<td>29</td>
</tr>
<tr>
<td>Final Ratings</td>
<td>31</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>32</td>
</tr>
<tr>
<td>Part II</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Subjects</td>
<td>33</td>
</tr>
<tr>
<td>Experimental Design</td>
<td>33</td>
</tr>
<tr>
<td>Procedure</td>
<td>34</td>
</tr>
<tr>
<td>Rating Procedure</td>
<td>34</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>36</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>STATISTICAL METHODOLOGY</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Rating Experiment</td>
<td>37</td>
</tr>
<tr>
<td>Response Bias</td>
<td>39</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RESULTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part I</td>
<td>46</td>
</tr>
<tr>
<td>Reliability</td>
<td>46</td>
</tr>
<tr>
<td>Manipulation Check</td>
<td>47</td>
</tr>
<tr>
<td>Tests of Hypotheses</td>
<td>47</td>
</tr>
<tr>
<td>Part II</td>
<td>55</td>
</tr>
<tr>
<td>Final Ratings</td>
<td>55</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DISCUSSION</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary of Results</td>
<td>61</td>
</tr>
<tr>
<td>Implications and Future Research</td>
<td>67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REFERENCES</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>72</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPENDIX</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>76</td>
</tr>
<tr>
<td>B</td>
<td>77</td>
</tr>
<tr>
<td>C</td>
<td>78</td>
</tr>
<tr>
<td>D</td>
<td>80</td>
</tr>
<tr>
<td>E</td>
<td>81</td>
</tr>
<tr>
<td>F</td>
<td>82</td>
</tr>
<tr>
<td>G</td>
<td>83</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Call Content with Respect to Frequency of Correct and Incorrect Behavior for Each Call .............. 30

Table 2: Hypothetical Raw Data Matrix for a Single Observer Following Rating the Correctness of 70 Correct and 70 Incorrect Behaviors .................................. 38

Table 3: Cumulative Conditional Probabilities of Observer's Responses for Each Rating Scale Category Calculated from Table 1 Raw Data. ........................................ 38

Table 4: The Effect of Observers' First Impression on Response Bias: Results of a Hotelling's T² Analysis. .... 49

Table 5: Cell and Marginal Means and Standard Deviations for a 2 x 2 Repeated Measures ANOVA. ............. 50

Table 6: The Effect of First Impressions and Sensitivity on Response Bias: Results of a 2 x 2 Repeated Measures ANOVA. ................................................................. 50

Table 7: Frequency of Classification of Each Behavior as High and Low Sensitivity .................................. 52

Table 8: Intercorrelation Matrix Among Behavior Sensitivity Indexes Computed Across Subjects ................ 53

Table 9: Multiple Regression Analyses Using Impressions, Sensitivity and their Interaction to Predict Response Bias on Each Behavior .............................................. 54

Table 10: MANOVA Results: Multivariate F Ratios .......... 56

Table 11: Means and Standard Deviations of Final Performance, Ability and Motivation Ratings Made Following Observation ..................................................... 57

Table 12: Univariate F Ratios and Estimated Proportion of Total Variance Accounted for by Each Variance Component ................................................................. 59
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Possible Stimulus Event-Observer Response Combinations.</td>
<td>5</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Possible Stimulus Event-Observer Response Combinations in a Performance Appraisal Context.</td>
<td>5</td>
</tr>
<tr>
<td>Figure 3</td>
<td>A Graphical Illustration of the Basic Decision Model for the Rating Situation.</td>
<td>7</td>
</tr>
<tr>
<td>Figure 4</td>
<td>A ROC Curve Illustrating the Relationship Between Hit and False Positive Conditional Probabilities and the Criterion.</td>
<td>9</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Experimental Design for Hypothesis El-B</td>
<td>24</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Behaviors Exhibited by Stimulus Person on 22 Simulated Phone Conversations.</td>
<td>26</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Experimental Design for Experiment Two.</td>
<td>35</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Placement of Criteria for Two Hypothetical Observers.</td>
<td>41</td>
</tr>
<tr>
<td>Figure 9</td>
<td>A Receiver Operating Curve (ROC) for Data in Table 3.</td>
<td>43</td>
</tr>
<tr>
<td>Figure 10</td>
<td>An ROC Curve Divided into Geometric Components.</td>
<td>44</td>
</tr>
</tbody>
</table>
INTRODUCTION

The observation and assessment of behavior is important to such theoretical concerns as impression formation and attribution theory, as well as in such applied concerns as performance appraisal and assessment centers. Although much research has been concerned with the reliability and validity of performance ratings, little has been conducted involving the accuracy of behavioral observation. Accuracy is defined as the relationship between a set of measurements obtained with a fallible scale of some sort and a corresponding set of measurements derived from an accepted standard or less fallible scale (Gordon, 1972). The paucity of research addressing the accuracy of observation may partly be a function of the unavailability of an acceptable standard. Considering the increasing emphasis currently placed on behavior (Wernimont & Campbell, 1968), more research addressing factors affecting observational accuracy and the consequences of different levels of accuracy is needed.

Most impression formation studies assume that individuals abstract certain aspects of others' behavior, organize their behavioral observations around certain dispositions, and develop a "picture" of the observed other. Individuals appear to structure much of their observations in meaningful categories. Hastorf, Schneider and Polefka (1970) have noted that this process may also restrict our awareness
of subsets of others' behavior. Impressions and attributions of others' behavior may create stability and meaningfulness to the observer's experience; however, they may do so at the risk of subsequent observational inaccuracy. Although the impression formation and attribution literature have suggested that judges are more influenced by early than late information when forming overall judgments of performance and ability, there exists little agreement regarding how this process occurs (Jones et al., 1972). In addition, little research addresses the notion of observational accuracy in an impression formation context.

This research investigated several variables believed to affect observational accuracy. Two experiments were conducted. **Experiment One** investigated the effect of the observer's first impression of the stimulus person and his/her ability to discriminate between correct and incorrect behavior (independent variables) on response bias (dependent variable). The concept of response bias as used in Signal Detection Theory (Green & Swets, 1966) refers to an observer's tendency to rate behavior in a certain direction independent of the evidence provided. **Experiment Two** was a rating study that examined the effect of first impressions and response mode (interpolated versus summary mode) on final performance ratings, ability attributions and motivation ratings.

It is argued that first impressions and the observer's ability to discriminate between correct and incorrect behavior (sensitivity) results in primacy effects through the influence of response bias. Primacy effects occurs when information presented early in the
assessment process carries greater weight in final judgments than does later information (Carlson, 1971). Experiment Two was designed to investigate the effects of first impressions and response mode on primacy effects and is therefore a rating study. Experiment one was designed to examine one possible process underlying primacy (response bias), and therefore was a study of observational accuracy within a Signal Detection framework.

This chapter presents relevant literature and hypotheses to be tested in experiments one and two. Chapter two offers a methodology for testing those hypotheses. A statistical methodology for indices used in experiment one is offered in chapter three. Results are reported in chapter four, followed by a discussion and implications of the present study.

This chapter is divided into two parts. Parts I and II are concerned with experiments one and two, respectively. Part I begins with an overview of Signal Detection Theory (SDT) followed by a presentation of factors believed to affect response bias. Literature addressing observational accuracy is then reviewed, followed by hypotheses. Part II begins by outlining literature dealing with order effects and hypotheses to be tested in experiment two.

Part I

Experiment one uses a signal detection theoretical approach to understanding the effect of first impressions and sensitivity on response bias. An overview of SDT is presented below.
**Signal Detection Theory.** The overview presented below is intended to familiarize the reader with the basic concepts and terms in SDT. More exhaustive accounts of the theory may be found in Egan (1975), Green and Swets (1966) and McNicol (1972). SDT is a perceptual model which separates decisional from perceptual aspects of response to stimulus magnitude. Typically, subjects observe two classes of stimuli: one in which a signal (e.g., a tone) appears embedded within noise, and another stimulus class in which just noise occurs. The subjects' task is to report whether or not he detected a signal. The two stimulus classes are sufficiently difficult to discriminate that errors are made.

The fourfold table presented in Figure 1 summarizes the possible stimulus-response combinations when the subject's response is dichotomous. Cells A and D represent correct decisions; that is, the subject reported having detected a signal when it actually appeared and reported not having detected a signal when none appeared. These decisions are called "hits" and "correct rejections," respectively. Cells C and B represent two types of error. In cell C, the subject reported having detected a signal was present when none actually occurred. In cell B, the subject reported that no signal was present when one actually appeared. These errors are called false positives and misses, respectively.

After reporting on each of a series of trials in which a signal or no signal is presented to an observer, a raw data matrix like Figure 1 may be constructed. Conditionalizing on the stimulus events, two probabilities capture the information from this data matrix:
<table>
<thead>
<tr>
<th>Observer's Response</th>
<th>Signal Present (&quot;S&quot;)</th>
<th>Signal Absent (&quot;N&quot;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulus Event</td>
<td>Signal Present (s)</td>
<td>Signal Absent (n)</td>
</tr>
<tr>
<td>Observer's Response</td>
<td>A (HIT)</td>
<td>B (MISS)</td>
</tr>
<tr>
<td></td>
<td>C (FALSE POSITIVE)</td>
<td>D (CORRECT REJECTION)</td>
</tr>
</tbody>
</table>

**Figure 1**

Possible Stimulus Event-Observer Response Combinations

<table>
<thead>
<tr>
<th>Observer's Response</th>
<th>&quot;Correct Behavior Detected&quot; (&quot;C&quot;)</th>
<th>&quot;Incorrect Behavior Detected&quot; (&quot;I&quot;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulus Event</td>
<td>Correct Behavior Present (c)</td>
<td>Incorrect Behavior Present (i)</td>
</tr>
<tr>
<td>Observer's Response</td>
<td>A (HIT)</td>
<td>B (MISS)</td>
</tr>
<tr>
<td></td>
<td>C (FALSE POSITIVE)</td>
<td>D (HIT)</td>
</tr>
</tbody>
</table>

**Figure 2**

Possible Stimulus Event-Observer Response Combinations in a Performance Appraisal Context
P(S/s): the probability of an observer reporting having detected a signal given that a signal appeared (hit).

P(S/n): the probability of the observer reporting having detected a signal given that no signal appeared (false positive).

A similar decision-making framework as shown in Figure 1 may be utilized in a number of organizational contexts. For example, a superior rating the correctness of his subordinate's job behavior is illustrated in Figure 2. In this context, the observer's response consists of reporting whether he saw an incorrect or correct behavior. The stimulus event is the subordinate behavior, which may be characterized as incorrect or correct vis-à-vis some acceptable standard. The various types of errors (e.g., hits, false positives) pertain as before. The method by which observers categorize stimulus events into stimulus classes involves a discussion of a likelihood ratio, which is presented below.

The observer must decide to which of two stimulus categories an observed stimulus event belongs. The observer bases this decision on the magnitude or nature of a single number x. This value is known as the evidence variable. Each stimulus event is represented by x, a specific value of a random variable X. It is upon this evidence variable that the observer makes his decision when categorizing the event. Defining one stimulus class containing correct behavior as c and another class containing incorrect behavior as i, Figure 3 illustrates several criteria an observer may adopt when classifying stimuli. Two conditional probabilities exist for each x:
Figure 3

A Graphical Illustration of the Basic Decision Model for the Rating Situation

Adopted from Baker and Schuck (1975).
\( P(x/c) \): the probability that \( x \) occurred given stimulus class \( c \).

\( P(x/i) \): the probability that \( x \) occurred given stimulus class \( i \).

The likelihood ratio of a given behavior with value \( x \) is defined as the ratio of the two conditional probabilities:

\[
L_c(x) = \frac{P(x/c)}{P(x/i)}
\]

According to the SDT model, the observer decides which behavior occurred (incorrect or correct) in a given trial by first calculating a likelihood ratio for the evidence variable \( x \).

The observer adopts a criterion, that value of the likelihood ratio that enables him to decide from which stimulus class distribution his observation arose. If the likelihood ratio is above this value, the observer will rate the behavior as "correct." If the observed likelihood ratio falls below the criterion set by the observer, it is rated as "incorrect." Referring to Figure 3, \( \text{Crit}_1 \) represents a strict criterion in which a relatively high \( x \) must be observed before the observer is willing to classify it as correct. \( \text{Crit}_2 \) is a lax criterion such that a relatively low \( x \) value is needed for the observer to classify it as correct.

The observer's performance with respect to hits, misses, false positives and correct rejections covary directly with the observer's criterion. Considering the strict criterion (\( \text{Crit}_1 \)) in Figure 3, the area to the right of the criterion and under the distribution of correct behavior represents the probability of responding "correct" when the behavior was actually correct (hit). The area under the incorrect behavior distribution and to the right of the criterion
A ROC Curve Illustrating the Relationship Between Hit and False Positive Conditional Probabilities and the Criterion

Adopted from Baker and Schuck (1975).
represents the probability of responding "correct" when the behavior was actually incorrect (false positive).

The tradeoff between the hit and false positive rates and the criterion can be illustrated by means of a Receiver Operating Characteristic, or ROC curve as shown in Figure 4. A ROC curve is constructed by plotting the observer's hit and false positive conditional probabilities on the ordinate and abscissa, respectively. An observer with $\text{Crit}_1$ has a hit rate probability of 0.40 and a low false positive probability. As the criterion becomes more lax, the hit rate increases; however, so does the false positive probability. As the criterion becomes extremely lax ($\text{Crit}_2$), the hit rate increases in small increments relative to the increase in the false positive probability.

SDT suggests several variables that influence the level of criterion an observer adopts when classifying behavior into different stimulus event categories. These variables are of significance to the study of the evaluation of behavior in performance appraisal, training and other organizational contexts. Some of these variables are offered later in this chapter.

Observer Characteristics. SDT is concerned mainly with two observer characteristics: sensitivity and bias. Sensitivity is defined as the observer's ability to simultaneously make correct decisions while avoiding incorrect decisions. Sensitivity may be estimated by measuring the area under the ROC curve depicted in Figure 4. The larger this area, the greater the observer's sensitivity. A computational methodology for deriving sensitivity estimates from raw
response data is presented in Chapter 3.

Bias refers to the propensity of an observer to favor one response over another independent of the evidence he has been provided (McNicol, 1972). Referring to Figure 3, criterion 1 represents a tendency to respond "incorrect" since a relatively high x value must be evidenced before the observer responds that he observed a correct behavior. On the other hand, criterion 2 represents a tendency to respond "correct" since a relatively low x value must be evidenced before the observer responds "correct." A computational methodology for calculating response bias estimates is presented in Chapter 3.

Factors Influencing Response Bias

First impressions. Baker and Schuck (1975) have suggested that the observer's expectancies about the behavior being observed may influence response bias. The degree of correctedness an observer attaches to behavior may in part be contingent upon his expectancy for correct or incorrect behavior to occur. A variable that might effect this expectancy is the observer's first impression of the stimulus person. The observer may expect subsequent behavior to be similar to previous behavior with respect to correctness. This expectation may be strong if the behavior is viewed as indicative of the stimulus person's ability vis-a-vis the task, assuming that ability is viewed as a fairly stable disposition (Jones et al., 1972). A positive first impression may result in an observer interpreting behavior as more correct than would an observer possessing a negative first impression. An observer with a positive first impression may set a lax criterion
such that more behaviors are reported as being correct than in fact occurred. A negative first impression may cause an observer to choose a strict criterion such that more correct behaviors are reported to be incorrect. This phenomena would result in observers with different impressions of a stimulus person reporting the same behavior differently, perhaps in the direction of their first impressions.

Observer Sensitivity. Though its interaction with the observer's first impression, sensitivity may effect response bias. Response bias may be greater for those behaviors for which an observer has difficulty discriminating between correct and incorrect manifestations. The effect of the observer's first impression on response bias may be less for those behaviors to which the observer exhibits a high level of sensitivity.

As previously mentioned, experiment one investigated the effect of first impressions and sensitivity on response bias. Literature addressing the notion of observational accuracy is relevant in this context.

Observational Accuracy

Utilizing a product inspection procedure, Harris (1968) showed that subjects failed to report defects far more frequently than they incorrectly reported defects (i.e., reported defects when none existed). Jacobs and Vandeventer (1968) found that subjects more accurately identified correctly completed samples of Raven's Colored Progressive Matrices Test than incorrectly completed samples. Gordon (1970) labeled the tendency of observers to more accurately observe correct manifestations of an object or behavior as the Differential
Accuracy Phenomena (DAP).

To test the DAP hypothesis, Gordon (1970) had subjects view a series of videotapes depicting a life insurance agent conducting preliminary phone conversations with prospective insurance buyers. Subjects were provided with correct agent behaviors on nine dimensions which served as a standard against which the videotaped behavior would be assessed. Subjects were provided scripts which contained a verbatim account of what the agent would say if he were performing correctly. The subject's task was to observe the tapes and report when the stimulus person deviated from the script. A checklist was provided for these recordings. The number of incorrect and correct behaviors were counterbalanced across all videotapes. Accuracy scores for both reported correct and incorrect behavior were obtained by comparing the subject's responses on the checklist to a key that was used to develop the videotapes. In other words, an external criterion was available which provided a means of ascertaining the accuracy of the observations. The results indicated that the checklist was 88% accurate for correctly performed behavior but only 73% accurate for incorrectly performed behavior (p < .01).

Gordon (1972) also conducted a partial replication of the study outlined above in which observers were provided with favorable or unfavorable information about the agent prior to observation. Gordon reasoned that if an observer was favorably predisposed toward the agent, the agent would be given the benefit of the doubt whenever uncertainty arose regarding the correctness of observed behavior. This phenomena would also result in similar patterns in the data as
would DAP. The observer's impression of the agent was manipulated by
providing a short paragraph to subjects describing the agent's past
performance in his company. This description was of a general nature
and not directly related to the specific behavior exhibited on tape.
Three overall impression conditions were attempted: favorable, un-
favorable and no first impression. A manipulation check revealed
that only the most and least favorable sets were induced. The mean
favorability rating for the most favorable condition was 5.32 and 4.25
for the least favorable condition using a seven point rating scale.
The results indicated that accuracy using the checklist and favor-
ability of the rater's initial impressions were unrelated; however,
the general finding that observers more accurately report correct
than incorrect behavior was replicated. Gordon (1972) noted that
future research should be conducted utilizing more extreme favorability
groups of observers. A more potent manipulation of first impression
could involve varying the initial behavior of the agent in addition
to presenting prior information about the stimulus person. Since the
initial impression and the subsequent stimulus behavior are more
closely related, the manipulation of impression favorability may be
stronger than was observed in the Gordon (1972) study.

Baker and Schuck (1975) reanalyzed the Gordon (1970) data
utilizing a signal detection approach as described earlier in this
chapter. Their results indicated that the DAP phenomena did not exist
for all types of behavior viewed. These authors suggested that the
observer's subjective probability of occurrence for certain behaviors
may, in part, determine whether the DAP will exist. A variable that
may effect this probability is the observer's first impression. It may be that observers form first impressions early during stimulus presentation and actively seek behavior that confirms those initial impressions. In fact, a major finding of the McGill studies (Webster, 1964) was that a bias is established early in interview contexts.

Few studies have investigated the effect of expectancy factors on observational accuracy. Kent, O'Leary, Diament, and Dietz (1974) studied the effect of expectation bias on the observational evaluation of therapeutic change. Subjects observed a "baseline" and "treatment" videotape of a child's classroom behavior. Observers were trained in the application of a nine category code for recording disruptive behavior. Categories included such items as "orienting," "aggression" and "out of chair." Half the subjects were told to expect a decrease of disruptive behavior from the baseline to the treatment tapes. The other subjects were told to expect no change. Both groups viewed identical tapes. Analysis of the behavior recordings revealed no significant main effects for any of the behavioral categories observed.

Shuller and McNamara (1976) gave subjects differential expectations concerning the trait state of a child viewed on videotape. Subjects were told either that the child had a history of aggressiveness, hyperactivity, normality, or were given no information. Observed recorded the behavior on a form containing several content categories such as aggression (hit, throw), hyperactivity (run, jump) and normality (give, touch). All subjects viewed the same stimulus material, and provided subjective ratings of aggressiveness and hyperactivity following observation. The trait descriptions provided to subjects
prior to observation did not effect their objective recordings of the child's behavior. The groups significantly differed with respect to their subjective ratings made after observation. A posteriori tests revealed that the aggressive and hyperactive expectation groups did not differ on any of the subjective ratings; however, both differed from the normal expectation group and no expectation group. The authors noted that the subjects in the different trait state expectation conditions may not have been able to differentiate the two trait states as they were manifested in the behavior of the child. Blunden et al. (1974) and Spivak and Swift (1966) have reported evidence suggesting that the two trait states of aggressiveness and hyperactivity are closely related as they are manifested in childhood behavior. Observers in the different expectation conditions therefore may have had a general "pathology" expectation resulting in no differences between the aggressive and hyperactivity expectation groups with respect to behavioral recordings.

Few studies have addressed the issue of expectation effects on observational accuracy. There exists little research investigating the effect of first impressions derived from relevant behavior. As mentioned previously, Baker and Schuck (1975) suggested that an observer's expectancies about behavior being observed may influence his response bias. The Gordon (1972) study represents a first attempt to examine expectancy factors in this context. The strength and nature of the manipulation in that study was criticized earlier.
The present study (experiment one) attempted to investigate the effect of subjects' expectancies regarding subsequent behavior via manipulating first impressions. Several hypotheses are offered below.

**Hypotheses**

Experiment one was designed to test three hypotheses listed below. As described in Chapter 2, subjects observed a series of videotapes containing a stimulus person performing correct and incorrect behaviors. The first impression of the stimulus person held by observers was manipulated (favorable, unfavorable). A second independent variable, sensitivity, was internally generated from the data. The dependent variable was response bias.

**Hypothesis El-A:** Positive first impressions will result in a higher mean vector of response bias scores than will negative first impressions.

The first hypothesis refers to the effect of first impressions on response bias. A conceptual definition of response bias was offered earlier. On a rating scale whose midpoint is neutral, a bias score less than the midpoint represents a negative bias. Similarly, a bias score greater than the midpoint represents a positive bias.

The second hypothesis refers to the interaction of first impressions and sensitivity on response bias. Sensitivity is viewed as a within-subjects observed partition (Runkel & McGrath, 1972) which is internally generated from each subject's responses separately. That is, a mean split on sensitivity scores (high-low sensitivity) was conducted for each subject in an ipsitive fashion.
Hypothesis El-B: The simple main effect of first impressions on response bias will be greater for behaviors to which observers exhibit low sensitivity than behaviors to which observers exhibit high sensitivity.

The third hypothesis refers to the effect of first impressions and sensitivity on response bias for each of ten behaviors separately. Sensitivity is a between-subjects variable in this context. The hypothesis is stated operationally in multiple regression terminology.

Hypothesis El-C: The interaction between the observer's first impression and sensitivity to a particular behavior will account for a significant amount of response bias score variance after impression and sensitivity are entered into the regression equation.

The operation test of the interaction between impressions and sensitivity is conceptually similar to treating impression and sensitivity as independent variables in an Anova framework, and testing the interaction source of variance for significance.

A presentation of literature and hypotheses relevant to experiment two are outlined below in Part II.

Part II

Experiment two was designed to study the effect of first impressions and response mode on primacy effects in final performance and ability ratings. Literature addressing primacy effects is outlined below.

Order Effects. Several studies suggest that judges are more influenced by early than late information when forming overall judgments of performance and ability (Feldman & Allen, 1975; Anderson, 1965; Jones et al., 1972). The order effect paradigm can simply be described as
comparing responses to one sequence of information with responses to the same information presented in reverse order (Jones et al., 1972). Final judgments between groups of observers receiving different stimulus sequences are then compared. Primacy effects have been demonstrated in the impression formation literature (Anderson, 1965; Anderson & Hubert, 1967; Briscoe et al., 1967; Luchins, 1958), the employment interviewing literature (Carlson, 1971; Farr, 1973; Hollman, 1972; London & Hakel, 1975), and the ability attribution literature (Jones et al., 1972; Feldman & Allen, 1975). A review of these studies is not presented in this document. The reader is directed to Jones et al. (1972) for a review of these studies and the various conditions under which primacy effects are apt to manifest themselves.

Experiment two treats the observer's first impression and response mode as independent variables. Response mode has two levels: interpolated and summary modes. Interpolated ratings refer to the provision of ratings intermittently throughout stimulus presentation, followed by final ratings. Summary ratings refer to the provision of only final ratings. Several studies have noted that the summary mode and interpolated mode of response typically result in primacy and recency effects, respectively (Farr, 1973; Farr & York, 1975; Hendrick & Constantini, 1970; Stewart, 1965). The recency effect occurs when late rather than early information carries greater weight in final judgments (Jones et al., 1972). Farr (1973) had subjects rate the learning ability, social ability, and overall suitability of hypothetical job applicants in a simulated interview context. Recency effects were found when subjects made interpolated judgments. A similar
result was found in a later follow-up study by the same author (Farr & York, 1975). After reviewing the literature on the issue, London and Hakel (1974) noted that the interpolated mode of response has not inevitably resulted in the recency effect.

Hypotheses to be tested in experiment two are outlined below.

The following hypotheses refer to the effect of first impressions and response mode (independent variables) on final performance ratings (dependent variables).

**Hypothesis E2-A:** Subjects possessing favorable first impressions will provide significantly higher performance ratings following observation than will subjects possessing negative first impressions.

**Hypothesis E2-B:** Relative to summary ratings, interpolated ratings will result in significantly higher final ratings when negative information precedes positive information.

**Hypothesis E2-C:** Relative to summary ratings, interpolated ratings will result in significantly lower final ratings when positive information precedes negative information.

**Summary**

Two experiments, each with their respective literature review and hypotheses, were reviewed. Experiment one examined the effect of first impression and sensitivity on response bias utilizing a SDT framework. Experiment two examines the effect of first impression and response mode on primacy effects with respect to final ratings. The two experiments are complementary in that the first attempts to utilize a new approach (SDT) to understand the process underlying primacy effects in final performance and ability ratings.
A methodology for testing the hypotheses just outlined is offered in the next chapter.
METHODOLOGY

This chapter contains two parts. Parts I and II are concerned with experiments one and two, respectively. Each part contains a description of subjects, experimental design, procedures taken and data analytic techniques used.

Part I: Experiment One

Subjects. Eighty students enrolled in the introductory psychology course at the Ohio State University participated as subjects. Subjects received extra credit towards their course grade in return for their participation. The subjects represented a variety of academic majors (e.g., humanities, social and natural sciences), and were predominantly in their freshman or sophomore year in college.

Experimental Design. (a) The experimental design used to test the effect of first impressions on response bias (hypothesis El-A) was a simple one factor fully randomized design. The independent variable was first impression (favorable, unfavorable). Ten response bias scores, one for each of ten behaviors observed, served as the dependent variables.

(b) A second experimental design was generated in order to test hypothesis El-B. This experimental design was a 2 x 2 factorial in which one factor was fully randomized and the other factor was
 internally generated from the data (i.e., an observed partition created by means of some pre-existing property of the subjects; Runkel & McGrath, 1972). The first independent variable was the observer's first impression of the stimulus person presented on videotape. This variable had two levels: unfavorable and favorable first impressions. The second independent variable was the observer's sensitivity to the behaviors, a within-subjects variable. As mentioned in Chapter 1, sensitivity is defined as the observer's ability to discriminate between incorrect and correct manifestations of behavior. Sensitivity scores, P(A), were computed for each observer on each behavior observed. For each observer, the behaviors were dichotomized into those for which the observer exhibited a high degree of sensitivity (high sensitivity) and those for which the observer exhibited a low degree of sensitivity (low sensitivity). This dichotomization was accomplished by means of a mean split of the P(A) scores for each behavior for a single observer. This experimental design is illustrated in Figure 5.

The dependent variable was the observer's response bias, defined as the extent to which the observer favors one type of response over another independent of the evidence he is provided (McNicol, 1972). Since the specific behaviors classified as high or low are apt to vary somewhat among subjects, the dependent variable was an average response bias score for those behaviors classified as high or low sensitivity for each subject.

"Behavior" refers to one of the ten behaviors listed in Figure 6.
### Observer Sensitivity

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfavorable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favorable</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5**

Experimental Design for Hypothesis El-B
(c) The third hypothesis was tested with a correlational methodology, and will be described later in this chapter.

Procedure

Upon entering the laboratory setting, subjects completed a brief biographical form tapping such items as sex, experience with rating scales and with insurance salespersons. This form may be found as Appendix A. All subjects then heard an audiotape containing a description of the experiment and their task. A transcript of these instructions appear as Appendix B. The purpose of the experiment was described as the investigation of how individuals judge the correctness or incorrectness of selling behavior. Subjects were told that they would soon observe and evaluate how well a life insurance agent performed on a series of videotaped telephone calls made to prospective buyers. Subjects were informed that it would be necessary to learn a set of behavioral guidelines that would serve as a standard of correctness against which the videotaped behavior would be assessed. Behavior Training. A short training session was conducted for the purpose of familiarizing subjects with behaviors said to be related to successful agent performance when making phone contacts. These behaviors are listed in Figure 6. Subjects were asked to judge how well the agent performed these behaviors. Examples of correct and incorrect behaviors were presented (Appendix C). To insure adequate understanding of the behaviors, subjects were informed that a short quiz measuring their comprehension of the behaviors would be administered following the training session (Appendix D). The quiz
1. Smile when first speaking to prospect. (SMILE)

2. Use of Standard Introduction: (1) ask the prospect if he is free to talk on the phone, (2) introduce himself, (3) introduce company he represents, (4) the purpose of the call. (STANDARD INTRO)

3. Use of proper responses to prospect objections. (OBJECTION 1, 2)

4. Prompt response to prospect objection. (PROMPTNESS)

5. Keep trying to make the appointment, but NOT after three objections are received from the prospect. (PERSISTENCE)

6. If the appointment is granted: repeat the time, place and day of appointment before hanging up. If the appointment is denied, leave agent name, company and telephone number before hanging up. (CLOSING)

7. Calling the prospect by his/her correct name throughout the phone call. (NAME)

8. Thanking the prospect regardless of call outcome. (THANK)

9. NOT talking about insurance when the prospect asks what sort of policies his company offers. That is, the agent should not sell insurance over the phone. (SELL INSUR)

Figure 6
Behaviors Exhibited by Stimulus Person on 22 Simulated Phone Conversations

---

1 This behavior occurred twice during each call.
served three functions. First, it assessed how well the subjects understood the behaviors. Second, the knowledge that subjects would be quizzed probably increased their attention throughout the thirty minute behavior training session. Finally, the quiz and subsequent scoring provided a second iteration of the behaviors. The quiz contained nine true-false items intended to measure subjects' knowledge of the behaviors. Subjects graded their own and were encouraged to ask questions of the experimenter if they felt unsure regarding the meaning of any of the behaviors.

Prior to viewing the videotapes to be described below, subjects were familiarized with the scales used to report the correctness of the agent's behavior. Instructions presented to subjects regarding scale use appear as Appendix E. The scale consisted of six anchors: "1" representing a report of "totally incorrect" and "6" representing a report of "totally correct," reflecting the subjects' judgment regarding how correctly (or incorrectly) he thought the agent performed each behavior. An overall rating was also provided using the same scale. A sample set of scales for a single call appears as Appendix F. Subjects viewed a single practice tape and practiced using the scales.

**First Impression Manipulation.** The observer's first impression of the agent was manipulated two ways. Subjects first listened to a short audiotape containing a description of the agent prior to viewing the stimulus material. In the favorable impression condition, the agent

---

2All but one subject received a perfect score on the quiz.
was described as a successful performer possessing a thorough knowledge of the behaviors said to be related to successful telephone approaches, which were written by the insurance company that employed him. A transcript of this agent description is shown below:

Steve Cooper, the agent you are about to observe, is an experienced insurance salesperson. Mr. Cooper has demonstrated a high level of achievement and has a proven track record with Sigourney Life Insurance Company. Before the videotapes you are about to observe were made, Mr. Cooper studied the behaviors we did and took a more comprehensive examination than we did to measure his understanding of those behaviors. Mr. Cooper scored higher than most of our agents with comparable experience. Sigourney plans to use the tapes as a training device for new agents first learning the guidelines.

Subjects in the favorable first impression condition listened to a different audiotape describing a poor performer possessing little knowledge of the behaviors:

Steve Cooper, the agent you are about to observe, has been a life insurance salesperson for several months. Since it takes about a year to become a good salesperson, Mr. Cooper is still learning the ropes. Mr. Cooper has not demonstrated much achievement thus far and is just hanging on with Sigourney Life Insurance Company. In order to improve his performance, Sigourney decided to videotape Mr. Cooper attempting to perform the behaviors. Before the taping, he had studied the behaviors said to be related to successful telephone approaches, and even took the same quiz we did. Unfortunately, Mr. Cooper scored lower than most of Sigourney's agents with comparable experience. We made the tapes so that we could learn which, if any, behaviors he has difficulty performing correctly. That way, we can feed the results back to Mr. Cooper so that he knows what he must work on in order to improve.

The second mechanism for manipulating the observer's first impression was by systematically varying the order of the phone calls with respect to the ratio of incorrect to correct behaviors contained
in each call. A more complete description of the stimulus material is presented below.

**Stimulus Material.** Subjects viewed a series of 22 simulated telephone conversations between a life insurance salesperson and a prospective insurance buyer. The same agent appeared in each call, but the voices of the prospects varied. The tapes were constructed such that the agent performed each behavior either correctly or incorrectly during each call. A transcript of a single call is contained in the Appendix G. Subjects viewed the tapes in a large classroom equipped with two large television monitors positioned in front. Each subject had a clear view of the monitors and could hear the audio portion of the tape clearly. Each call lasted approximately two minutes, and were separated by 30 seconds. During this time interval between calls, subjects finished making their behavior ratings and provided a performance rating for that call. It was determined through pre-testing that 30 seconds was adequate for the task at hand.

Table 1 contains a summary of the call content with respect to the frequency and location of correct and incorrect manifestations of each behavior for each of 22 calls. The total number of behaviors performed correctly and incorrectly were equated over all calls. Each type of behavior (see Figure 6) was performed correctly and incorrectly an equal number of times. The number of correct behaviors

---

3The author is indebted to Dr. Michael Gordon for providing videotapes from a previous study. The tapes were deemed inappropriate for the present study; however, they served as models for the tapes constructed for this study.
Table 1

Call Content with Respect to Frequency of Correct and Incorrect Behavior for Each Call

| Behavior | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | Totals |
|----------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Smile    | I | I | C | I | I | I | C | I | I | C | C | I | C | C | C | C | C | C | C | 11 | 11 |
| Standard Intro. | I | I | I | C | I | I | I | C | I | I | C | C | I | I | C | C | C | C | C | C | 11 | 11 |
| Objection 1 | I | I | I | C | I | C | I | C | I | I | I | C | C | C | C | I | I | C | C | C | 12 | 10 |
| Promptness | I | I | I | I | I | I | C | I | I | C | C | I | I | C | C | I | I | C | C | C | 12 | 10 |
| Objection 2 | I | I | I | I | C | I | I | I | C | I | C | C | C | I | I | C | C | C | C | C | 11 | 11 |
| Persistence | C | I | I | I | I | I | I | C | I | I | C | C | I | C | C | C | I | C | C | C | 10 | 12 |
| Closing   | I | I | I | I | C | I | C | I | C | I | I | C | C | C | C | C | I | C | I | C | 12 | 10 |
| Name      | I | I | I | I | C | I | I | I | I | C | C | C | C | C | C | I | C | C | I | C | 11 | 11 |
| Thanks    | C | I | I | I | I | I | C | I | C | I | I | I | C | C | C | I | I | C | C | C | 12 | 10 |

Sell Insur. | I | I | I | C | I | I | C | I | I | I | C | I | I | I | C | C | C | C | C | C | C | 11 | 11 |

Total Correct | 1 | 1 | 1 | 2 | 3 | 2 | 3 | 4 | 3 | 6 | 1 | 6 | 6 | 6 | 7 | 6 | 8 | 7 | 6 | 9 | 9 | 10 | 113 | 107 |

Call Outcome | D | D | A | D | A | A | D | D | A | D | D | A | D | A | A | A | A | A | A | A | A | A | A | A | 11 | 11 |

\(^{a}\)Calls 2 and 11, and 14 and 19 are identical. These calls were repeated in order to obtain an estimate of observers' rate-rate reliability. Prospect voices however, were different.

\(^{b}\)I = incorrect behavior and C = correct behavior.

\(^{c}\)A = appointment to meet with prospect accepted, D = appointment denied.

\(^{d}\)see Figure 6 for behavior descriptions.
was systematically varied. As Table 1 indicates, the number of
correct behaviors increases as one moves from call 1 to call 22. The
initial behaviors viewed by half of the observers were therefore pre­
dominantly incorrect. Subjects in the negative first impression con­
dition viewed the tapes in this order. Subjects in the positive
first impression condition viewed the tapes (calls) in the opposite
order, thereby initially observing predominantly correct behaviors.
Two calls were repeated (calls 2 and 14) in order to determine the
rate-rerate reliability of the observer's responses. The overall
ratings for the first three calls served as a manipulation check of
the first impression induction.

Final Ratings. Following observation and evaluation of all 22 calls,
subjects rated the overall performance of the agent on each behavior
using the same scale used to rate individual call performance des­
cribed earlier. In addition, subjects provided ability attributions
for each behavior and an overall ability rating using a six point
scale ("1" represented very low ability, and "6" represented excellent
ability). Subjects also indicated how motivated they believed the
agent to be, and the extent to which they thought the agent required
additional training in order to perform the behaviors correctly. The
scales used to make these final ratings appear as Appendices H and I.

Following completion of the final ratings, subjects were
thoroughly debriefed regarding the purpose of the experiment and
their role as subjects.
**Data Analysis.** The reliability of the subjects' responses was estimated by means of rate-rerate correlations among ratings made from repeated stimulus presentations.

A check for the first impression manipulation was also conducted. A t statistic using first impression as the independent variable and the average overall performance rating for the first three individual calls as the dependent variable.

The effect of first impressions on response bias was ascertained by computing a Hotelling's $T^2$ statistic using first impression as the independent variable and the response bias scores on each of ten behaviors observed as the dependent variables. Hotelling's $T^2$ statistic is the multivariate generalization of the univariate student's $t$, and can be used to test for equivalence of multivariate mean vectors derived from two independent samples (Overall & Klett, 1972).

The effect of first impressions and behavioral sensitivity on response bias was examined by means of a 2 x 2 repeated measures analysis of variance (ANOVA). The first impression and sensitivity factors were between and within-subject variables, respectively. The dependent variable was an average response bias score in each cell of the design (Figure 5).

The effect of first impressions and sensitivity on response bias for each behavior separately was examined using a multiple regression approach. The regression model used is shown below:

$$Y' = a + b_1 x_1 + b_2 x_2 + b_{1,2} x_1 x_2$$
Where: \( Y' \) = predicted response bias score  
\( x_1 \) = raw score for first impression (1,0)  
\( x_2 \) = raw score for sensitivity (continuous)  
\( x_1 x_2 \) = cross product term representing interaction  
\( b_1, b_2, b_1^*, b_2^* \) = unstandardized beta weights  
\( a \) = constant

Variables were entered into the regression equation in a hierarchical stepwise fashion in order to examine the relationship between the interaction term and the dependent measure after first impressions and sensitivity were entered into the regression equation. The independent variables were entered into the equation in the following sequence: \( x_1, x_2 \) and \( x_1 x_2 \). A separate analysis was performed for each behavior listed in Figure 6.

Part II describes experiment two in a similar format as that of Part I.

Part II: Experiment Two

Subjects. Subjects for experiment two were drawn from the same subject pool as the first experiment. In addition to the subjects employed in experiment one, 80 additional subjects participated in experiment two. As previously mentioned, experiment two was a final performance rating study. Subjects in experiment one provided final performance rating data following observation and were used in experiment two. The total number of subjects in experiment two was therefore 160.

Experimental Design. The experiment is characterized as fully randomized two factorial design. The first independent variable was the observer's first impression: favorable or unfavorable. The second independent
variable was the observer's mode of response. Subjects in one condition rated the stimulus person on each behavior during or after each call \((N = 22)\). Subjects in this condition then provided final performance ratings after observing all 22 calls. This mode of response is referred to as interpolated. It should be noted that the subjects in the interpolated conditions \((N = 80)\) also served as subjects in the first experiment. An additional 80 subjects provided only final performance ratings after having observed all calls. This mode of response is referred to as the summary mode. The experimental design is presented in Figure 7. The dependent variables in this design are the observers' final performance ratings.

**Procedure**

The experimental procedures employed in experiment two parallel those of experiment one. All subjects viewed the same stimulus material (outlined in Table 1) containing a life insurance agent performing ten behaviors on each of 22 trials (calls). All subjects received identical training with respect to the behaviors said to be related to the successful telephone approach described earlier. The first impression manipulation was also induced in the identical fashion as the first experiment.

**Rating Procedure.** Subjects in the interpolated conditions (unfavorable and favorable first impressions) rated the stimulus person's behavior as described previously in Part I (see Appendices E, F, H, I). Subjects in the summary mode of response conditions did **not** rate behavior following each call. Subjects were encouraged however, to
Mode of Response

Interpolated Summary

<table>
<thead>
<tr>
<th>Unfavorable</th>
<th>Favorable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First</strong></td>
<td><strong>Impression</strong></td>
</tr>
</tbody>
</table>

*Figure 7*

Experimental Design for Experiment Two
take whatever notes they felt were necessary in order to rate the overall performance of the agent following all observation. Paper was provided to subjects for this purpose. Inspection of these notes revealed that subjects did not take systematic or exhaustive notes during stimulus presentation. Subjects in the summary response mode conditions then provided final performance, ability attribution and motivational ratings as did subjects in the interpolated conditions (Appendices H and I). Following completion of the final ratings, all subjects were thoroughly debriefed.

Data Analysis. The main and interaction effects of first impressions and response mode on the final ratings were ascertained by means of a 2 x 2 multivariate analysis of variance (MANOVA). MANOVA is a statistical technique which finds that linear combination of the dependent variables that maximizes the variance between groups and minimizes the variance within groups (Overall & Klett, 1972). Univariate F ratios for each dependent variable were also computed.

The following chapter outlines a statistical methodology for calculating response bias and sensitivity indexes as used in SDT (McNicol, 1972).
STATISTICAL METHODOLOGY

This chapter outlines the computational methodology for response bias and sensitivity indexes employed in the present study. The computations are illustrated by means of rating data from a hypothetical observer. The reader is referred to McNicol (1972) for a more detailed account of the computational procedures described in this chapter. The concept of response bias is addressed first, followed by a treatment of sensitivity.

A Rating Experiment. An experiment is briefly described to illustrate the computational procedures involved in deriving response bias scores. The purpose of the experiment was to measure one observer's response bias when observing behavior. The behaviors the observer viewed were constructed such that they were either correct or incorrect when compared to a standard of correctness defined by the experimenter in an a priori fashion. The subject's task was to rate the correctness of the behavior observed during a specified time interval (trial). A rating scale with 6 anchors was used by the observer, in which "1" represented "totally incorrect" and "6" represented "totally correct." Assume that the observer has rated 140 behaviors, half of which were incorrect and the other half were correct behavior. A hypothetical raw data matrix for this single observer is shown in Table 2.
Table 2
Hypothetical Raw Data Matrix for a Single Observer Following Rating the Correctness of 70 Correct and 70 Incorrect Behaviors

<table>
<thead>
<tr>
<th>Rating Scale Category</th>
<th>&quot;totally incorrect&quot;</th>
<th>&quot;totally correct&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 TOTAL</td>
<td></td>
</tr>
<tr>
<td>Stimulus Event</td>
<td>incorrect</td>
<td>correct</td>
</tr>
<tr>
<td></td>
<td>32 13 10 7 5 3</td>
<td>2 3 8 12 15 30 70</td>
</tr>
</tbody>
</table>

Table 3
Cumulative Conditional Probabilities of Observer's Responses For Each Rating Scale Category Calculated from Table 1 Raw Data

<table>
<thead>
<tr>
<th>Rating Scale Category</th>
<th>&quot;totally incorrect&quot;</th>
<th>&quot;totally correct&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6</td>
<td></td>
</tr>
<tr>
<td>Conditional Probability</td>
<td>P(C</td>
<td>i)</td>
</tr>
<tr>
<td></td>
<td>1.0 .50 .31 .21 .11 .04</td>
<td>1.0 .96 .92 .81 .64 .43</td>
</tr>
<tr>
<td></td>
<td>P(C</td>
<td>i) + P(C</td>
</tr>
<tr>
<td></td>
<td>2.0 1.46 1.33 1.02 .75 .47</td>
<td></td>
</tr>
</tbody>
</table>
Two conditional probabilities exist for each rating scale category:

\[ P(C/c): \text{the probability of responding "correct" given that the behavior was correct.} \]
\[ P(C/i): \text{the probability of responding "correct" given that the behavior was incorrect.} \]

The raw data matrix for each observer may be transformed into cumulative conditional probabilities for each rating scale category. This has been done in Table 3. The conditional probabilities, \( P(C/c) \) and \( P(C/i) \), are obtained by cumulating from the highest to the lowest scale category and dividing the cumulative frequency by the row totals (actual number of behaviors in respective stimuli classes). Referring to Table 3, the cumulative probability of the observer rating a behavior as "slightly incorrect" (category 4) given that the behavior was correct equals .81. The cumulative probability of the same response given an incorrect behavior is .21. The probabilities \( P(C/c) \) and \( P(C/i) \) are referred to as the observer's hit rate and false positive rate, respectively.

**Response Bias.** McNicol (1972) has defined a non-parametric measure of response bias which provides a single bias score for each observer. The bias score represents that point along the observer's rating scale at which he is equally disposed to rating a stimulus as a member of one stimulus class or another. When the two stimulus classes are incorrect and correct behavior and the observer uses the rating scale just described, the bias score would indicate that point at which the observer is equally disposed to rating the behavior as correct or incorrect. This point is operationally defined as the rating scale
category at which \( P(C/c) + P(C/i) = 1 \). For the data presented in Table 3, the bias score lies somewhere between scale category 4 and 5. A more exact estimate of bias is obtained by interpolation. Figures 8a and 8b illustrate placement of criteria for two observers with different biases. The bias score for observers X and Y are 4 and 2, respectively. Observer Y has adopted a strict criterion relative to the lax criterion adopted by observer X. The observer's criterion is defined as that magnitude of "correctness" beyond which he responds "correct" and below which he responds "incorrect." As can be seen in Figure 8, observer Y appears to rate the behaviors as less correct (more incorrect) than does observer X. That is, observer Y has a negative bias and observer X has a positive bias.

**Sensitivity.** The calculation of a sensitivity index begins by constructing a Receiver Operating Curve, or ROC Curve. This is accomplished in SDT by plotting the two conditional probabilities, \( P(C/c) \) and \( P(C/i) \) against each other. This has been done in Figure 9. The probability of responding "correct" given an incorrect behavior is placed along the abscissa. The probability of responding "correct" given a correct behavior is placed along the ordinate. The two probabilities associated with each scale category is then plotted on the ROC curve. This was done for the data contained in Table 3, and is shown in Figure 9.

An index of sensitivity is conceptually defined as the area under the ROC curve, and is designated as \( P(A) \). A method for estimating this area is geometrically illustrated in Figure 10.
Figure 8

Placement of Criteria for Two Hypothetical Observers

---

Adopted from McNicol (1972).
The area under the ROC curve may be estimated by the combined areas of the various trapezia and triangle ABF formed by dropping perpendicular lines BF, CG, and DH to the horizontal axis. The combined area of ABF, ACFG, CDGH and DEHI constitute \( P(A) \), the sensitivity index.

A generalized formula for computing \( P(A) \) written in terms of \( P(C/c) \) and \( P(C/i) \), where \( P_{i}(C/c) \) and \( P_{i}(C/i) \) are the coordinates for the \( i^{th} \) point, is given by McNicol (1972):

\[
P(A) = [P_{1}(C/i) - 0] [0 + P_{1}(C/c)] + \\
1/2 [P_{2}(C/i) - P_{1}(C/i)] [P_{2}(C/c) + P_{1}(C/c)] + ... \\
1/2 [P_{i}(C/i) - P_{i-1}(C/i)] [P_{i}(C/c) + P_{i-1}(C/c)] + ... \\
1/2 [P_{N}(C/i) - P_{i-1}(C/i)] [P_{N}(C/c) + P_{i-1}(C/c)] \\
P(A) = 1/2 \sum_{i=1}^{N+1} [P_{i}(C/i) - P_{i-1}(C/i)] [P_{i}(C/c) + P_{i-1}(C/c)]
\]

Note that \( P_{0}(C/i) \) and \( P_{0}(C/c) \) are the coordinates of scale category 6 (lower left in Figure 9) and that \( P_{N}(C/i) \) and \( P_{N}(C/c) \) are the coordinates of rating scale category 1 (upper right in Figure 9).

\( P(A) \) for the rating scale data in Table 3 is calculated below:

\[
P(A) = (.04-0) (.04+.43) + (.11-.04) (.64+.43) + \\
1/2 (.21-.11) (.81+.64) + \\
1/2 (.31-.21) (.92+.81) + \\
1/2 (.50-.31) (.96+.92) + \\
1/2 (1-.50) (1+.96)
\]
Figure 9
A Receiver Operating Curve (ROC) for Data in Table 3

Note: The negative diagonal is composed of points at which $P(c/i) = P(c/i)$. Therefore, the point at which the ROC curve intersects the negative diagonal is defined as the observers' response bias score.
Figure 10
An ROC Curve Divided into Geometric Components

1Adopted from McNicol (1972).
Thus, the sensitivity index for our hypothetical observer is .8823.
The sensitivity index may be interpreted as an estimate of the observer's ability to discriminate between the classes of stimuli he/she observes. The lower the index, the poorer the observer's ability.

Response bias and sensitivity were calculated in the present study in the methods described in this chapter. The next chapter reports the results of experiments one and two.
RESULTS

This chapter is divided in two parts. Parts I and II report the results of experiment one and two, respectively.

Part I

The reliability of the observer's responses and the manipulation check for the first impression induction are addressed. Analyses dealing with the hypotheses presented in Chapter 2 (Part I) are then presented.

Reliability. The rate-rerate reliability of the observers' responses was ascertained by computing correlations among ratings made on repeated stimulus presentations. The correlation coefficients (one for each behavior) were then averaged using Fisher's r to z transformation (Hays, 1963). The average rate-rerate correlation was \( r = .61 \). Although not impressive by traditional psychometric standards, this reliability estimate was considered adequate for the purposes of the present study. Observers in SDT experiments typically vary their criterion over time within an experimental session, leading to different ratings on repeated trials. This phenomena may have attenuated the present reliability estimate.
Manipulation Check. The manipulation check for the first impression induction consisted of computing a t statistic using first impression (favorable, unfavorable) as the independent variable on the average overall performance rating across the first three tapes as the dependent variable. The result indicated that subjects in the unfavorable first impression condition rated the performance of the stimulus person significantly lower than subjects in the favorable first impression condition (t(78) = 16.76, p < .001). The mean rating for the unfavorable and favorable first impression conditions were 2.40 and 5.09, respectively. These data suggest that the first impression manipulation was potent.

Test of Hypotheses.

Hypothesis El-A. The effect of first impressions on response bias was tested by computing a Hotelling's T² statistic using first impression as the independent variable and response bias on each of ten behaviors (Figure 6) as the dependent variables. Means, standard deviations, univariate F ratios and standardized discriminant weights are reported in Table 4. The multivariate F was significant (F = 11.83, p < .001), suggesting that the unfavorable and favorable impression groups significantly differed with respect to their mean vectors of response bias scores. Five univariate F Ratios reached statistical significance at the p < .025 level. The standardized discriminant weights provide an estimate of the relative importance of each variable in the discriminant function. The standardized discriminant weights indicate that the groups differed most with respect to behaviors 1, 5, 6, 7,
and 8 (see Figure 6 for descriptions). Inspection of the means in Table 4 show the predicted direction for each behavior achieving significance.

A reclassification analysis was conducted as a check of the adequacy of the discriminant function for discriminating among the two groups. This was accomplished by calculating a classification score for each subject in the following manner:

\[
C_i = c_{i1}v_1 + c_{i2}v_2 + \ldots + c_{ip}v_p + c_{i0}
\]

Where

\( C_i \) = classification score for \( i \)th observation

\( c_{ip} \) = classification coefficient for \( i \)th observation on \( p \)th variable.

\( v_{ip} \) = raw score on discriminant variable \( p \)

\( c_{i0} \) = constant

Since each group has a separate equation, two scores are computed for each observation. A probability of membership in each group was then computed, and the observation was classified into the group to which it most likely belonged. The percent of observations (subjects) correctly classified (i.e., hit rate) is an estimate of the adequacy of the discriminant function for separating the groups. The hit rate in the present analysis was 83.75% (chance = 25%). These data lend support for hypothesis El-A.

Hypothesis El-B. This hypothesis predicted that the effect of first impressions on response bias would be greater for behaviors to which each observer exhibited low sensitivity than behaviors to which each observer exhibited high sensitivity. A 2 x 2 repeated measures analysis
Table 4

The Effect of Observers' First Impression
On Response Bias: Results of a Hotelling's $T^2$ Analysis

<table>
<thead>
<tr>
<th>Behavior</th>
<th>First Impression</th>
<th>Univar. Disc.</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unfavorable M SD</td>
<td>Favorable M SD</td>
<td>$F(1,78)$</td>
<td>Weight</td>
<td></td>
</tr>
<tr>
<td>1. Smile</td>
<td>2.86 1.01</td>
<td>4.36 1.02</td>
<td>43.81***</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>2. Intro</td>
<td>4.47 .72</td>
<td>4.42 .89</td>
<td>&lt; 1</td>
<td>-.15</td>
<td></td>
</tr>
<tr>
<td>3. Objection One</td>
<td>3.74 .94</td>
<td>3.85 .93</td>
<td>&lt; 1</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>4. Promptness</td>
<td>4.53 .79</td>
<td>4.93 .65</td>
<td>6.14**</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>5. Objection Two</td>
<td>5.21 .70</td>
<td>5.34 .63</td>
<td>&lt; 1</td>
<td>-.30</td>
<td></td>
</tr>
<tr>
<td>6. Persistence</td>
<td>4.32 1.38</td>
<td>5.05 1.18</td>
<td>6.52**</td>
<td>.29</td>
<td></td>
</tr>
<tr>
<td>7. Closing</td>
<td>4.92 .91</td>
<td>4.70 1.13</td>
<td>1.08</td>
<td>-.24</td>
<td></td>
</tr>
<tr>
<td>8. Name</td>
<td>5.40 1.04</td>
<td>5.94 .51</td>
<td>8.77***</td>
<td>.35</td>
<td></td>
</tr>
<tr>
<td>9. Thank</td>
<td>5.21 .95</td>
<td>5.71 .86</td>
<td>6.06**</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>10. Sell Insur.</td>
<td>4.38 1.09</td>
<td>4.45 1.15</td>
<td>&lt; 1</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

$^1$Multivariate $F = 11.83, p < .001$

*p < .05

**p < .025

***p < .01
Table 5

Cell and Marginal Means and Standard Deviations\(^1\) for a 2 x 2 Repeated Measures ANOVA\(^2\)

**Sensitivity**

<table>
<thead>
<tr>
<th></th>
<th>High</th>
<th>Low</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfav.</td>
<td>4.44 (.71)</td>
<td>4.57 (.91)</td>
<td>4.51 (.81)</td>
</tr>
<tr>
<td>Favor</td>
<td>4.66 (.59)</td>
<td>4.92 (.63)</td>
<td>4.79 (.61)</td>
</tr>
<tr>
<td>Total</td>
<td>4.51 (.65)</td>
<td>4.79 (.77)</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Standard Deviations appear in parentheses.

\(^2\) Sensitivity and first impressions are within and between subjects variables, respectively, N = 80.

Table 6

The Effect of First Impressions and Sensitivity on Response Bias: Results of a 2 x 2 Repeated Measures ANOVA

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>Sums of Squares</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>% Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impressions (A)</td>
<td>3.23</td>
<td>1</td>
<td>3.23</td>
<td>4.48*</td>
<td>.03</td>
</tr>
<tr>
<td>Error</td>
<td>56.11</td>
<td>78</td>
<td>.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (B)</td>
<td>1.51</td>
<td>1</td>
<td>1.51</td>
<td>4.61**</td>
<td>.01</td>
</tr>
<tr>
<td>(A X B)</td>
<td>.13</td>
<td>1</td>
<td>.13</td>
<td>.41</td>
<td>.00</td>
</tr>
<tr>
<td>Error</td>
<td>25.62</td>
<td>78</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Percent of the total variance in the dependent variable.

\(*p < .05\)

\(**p < .01\)
of variance (ANOVA) was conducted (first impression X sensitivity). Table 5 contains cell means and standard deviations. The results of the ANOVA are shown in Table 6. A significant main effect for both first impressions and sensitivity was found (p < .03). Inspection of Table 5 reveals that the marginal mean are in the predicted direction. Observers exhibited higher response bias for those behaviors classified as low sensitivity than to those behaviors classified as high sensitivity. In addition, observers possessing unfavorable first impressions exhibited lower response bias scores than observers possessing favorable first impressions. Table 6 includes an estimate of total variance accounted for by each variance component. Although the impression and sensitivity main effects were statistically significant, they only accounted for 3% and 1% of the total variance of the dependent variable.

An additional analysis was conducted to ascertain whether the same behaviors were classified as high or low sensitivity categories for each observer. It appears intuitively reasonable to expect sensitivity to be related to several stimulus attributes such as intensity, duration and meaningfulness. The results of a chi-square (χ²) analysis conducted to test the hypothesis of no association between behaviors and sensitivity across observers appears in Table 7. The association between the sensitivity classification and the behaviors was significant (χ² = 102.15, p < .001). Cramer's statistic was calculated in order

4The use of χ² in this context violates the assumption of independent observations. Table 8 presents intercorrelations among the sensitivity scores across subjects, which suggest that the sensitivities observers exhibited toward the behaviors were not highly related (r = .12). The use of χ² in the present context was for exploratory purposes only.
to determine the strength of the association (Hays, 1963). The value of this statistic was .35, indicating a considerable degree of association (Cramer's statistic varies from 0 to 1). The effects of sensitivity and the behaviors themselves are therefore confounded in the ANOVA, and should be noted when interpreting results.

Table 7

Frequency of Classification of Each Behavior as High and Low Sensitivity

<table>
<thead>
<tr>
<th>Behavior</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>52</td>
<td>16</td>
<td>59</td>
<td>33</td>
<td>42</td>
<td>45</td>
<td>23</td>
<td>46</td>
<td>45</td>
<td>16</td>
<td>377</td>
</tr>
<tr>
<td>Low</td>
<td>28</td>
<td>64</td>
<td>21</td>
<td>47</td>
<td>37</td>
<td>35</td>
<td>57</td>
<td>34</td>
<td>35</td>
<td>64</td>
<td>422</td>
</tr>
<tr>
<td></td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>79</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>799</td>
</tr>
</tbody>
</table>

Hypothesis El-C. A second method of examining the effect of impressions and sensitivity on response bias consisted of separate analyses for each of the ten behaviors observed across subjects. The correlations among the ten sensitivity scores across subjects is presented in Table 8. The average correlation among the sensitivity scores was \( \bar{r} = .12 \) using Fisher's r to z transformation (Hays, 1963). It was therefore considered inappropriate to compute an average sensitivity score for each subject. In fact, the low correlations among sensitivity scores may have accounted for the weak results obtained for testing hypothesis El-B.

Table 9 contains the results of ten multiple regression analyses using first impressions, sensitivity and their interaction to predict response bias on each of ten behaviors. The Pearson correlations between each independent variable and the dependent variables,
Table 8

Intercorrelation Matrix Among Behavior

Sensitivity Indexes Computed Across Subjects\(^1\)

<table>
<thead>
<tr>
<th>Behaviors</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Smile</td>
<td>15</td>
<td>17</td>
<td>13</td>
<td>23</td>
<td>12</td>
<td>-06</td>
<td>14</td>
<td>15</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>2. Intro</td>
<td>15</td>
<td>26</td>
<td>13</td>
<td>22</td>
<td>03</td>
<td>05</td>
<td>08</td>
<td>00</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>3. Objection One</td>
<td>17</td>
<td>26</td>
<td>12</td>
<td>-17</td>
<td>05</td>
<td>18</td>
<td>-20</td>
<td>-12</td>
<td>-20</td>
<td></td>
</tr>
<tr>
<td>4. Promptness</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>09</td>
<td>16</td>
<td>-07</td>
<td>03</td>
<td>07</td>
<td>-08</td>
<td></td>
</tr>
<tr>
<td>5. Objection Two</td>
<td>23</td>
<td>22</td>
<td>-16</td>
<td>09</td>
<td>26</td>
<td>-09</td>
<td>14</td>
<td>08</td>
<td>00</td>
<td></td>
</tr>
<tr>
<td>6. Persistence</td>
<td>12</td>
<td>02</td>
<td>05</td>
<td>16</td>
<td>26</td>
<td>21</td>
<td>01</td>
<td>14</td>
<td>08</td>
<td></td>
</tr>
<tr>
<td>7. Closing</td>
<td>-06</td>
<td>05</td>
<td>18</td>
<td>-07</td>
<td>-09</td>
<td>21</td>
<td>10</td>
<td>07</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>8. Name</td>
<td>14</td>
<td>08</td>
<td>-19</td>
<td>03</td>
<td>14</td>
<td>02</td>
<td>10</td>
<td>04</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>9. Thank</td>
<td>15</td>
<td>00</td>
<td>-12</td>
<td>07</td>
<td>08</td>
<td>14</td>
<td>07</td>
<td>-04</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>10. Sell Insur.</td>
<td>11</td>
<td>17</td>
<td>-20</td>
<td>-08</td>
<td>00</td>
<td>08</td>
<td>24</td>
<td>16</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)Decimals omitted.
Table 9

Multiple Regression Analyses Using Impressions, Sensitivity and their Interaction to Predict Response Bias on Each Behavior

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Impression</th>
<th>Sensitivity</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Beta&lt;sup&gt;b&lt;/sup&gt;</td>
<td>RSQ&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>1. Smile</td>
<td>54&lt;sup&gt;**&lt;/sup&gt;</td>
<td>-21&lt;sup&gt;**&lt;/sup&gt;</td>
<td>29&lt;sup&gt;**&lt;/sup&gt;</td>
</tr>
<tr>
<td>2. Standard Intro</td>
<td>-08</td>
<td>2.36</td>
<td>01</td>
</tr>
<tr>
<td>3. Objection One</td>
<td>03</td>
<td>-1.81</td>
<td>00</td>
</tr>
<tr>
<td>4. Promptness</td>
<td>19&lt;sup&gt;*&lt;/sup&gt;</td>
<td>-.51</td>
<td>04</td>
</tr>
<tr>
<td>5. Objection Two</td>
<td>07</td>
<td>-2.11</td>
<td>01</td>
</tr>
<tr>
<td>6. Persistence</td>
<td>30&lt;sup&gt;**&lt;/sup&gt;</td>
<td>-.05&lt;sup&gt;**&lt;/sup&gt;</td>
<td>09&lt;sup&gt;**&lt;/sup&gt;</td>
</tr>
<tr>
<td>7. Closing</td>
<td>-08</td>
<td>2.95</td>
<td>01</td>
</tr>
<tr>
<td>8. Name</td>
<td>20&lt;sup&gt;*&lt;/sup&gt;</td>
<td>.83&lt;sup&gt;**&lt;/sup&gt;</td>
<td>04&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>9. Thank</td>
<td>31&lt;sup&gt;**&lt;/sup&gt;</td>
<td>.86&lt;sup&gt;**&lt;/sup&gt;</td>
<td>09&lt;sup&gt;**&lt;/sup&gt;</td>
</tr>
<tr>
<td>10. Sell Insur.</td>
<td>11</td>
<td>.11</td>
<td>01</td>
</tr>
</tbody>
</table>

<sup>a</sup> Correlation between that variable and response bias (decimals omitted).
<sup>b</sup> Standardized regression weight with all variables entered into equation.
<sup>c</sup> Squared multiple correlation (decimals omitted).
<sup>d</sup> Increase in RSQ by adding respective variable (decimals omitted).

*p < .05
**p < .01
standardized beta weights and the change in the squared multiple correlation \( R^2 \) by adding each respective independent variable into the regression equation are contained in Table 9. The variables were entered into the equation in a hierarchical sequence beginning with first impressions, sensitivity and finally the interaction term. The Pearson correlation between the first impression and response bias variables reached significance for behaviors 1, 6, 8, and 9. The addition of sensitivity into the prediction equation significantly increased the \( R^2 \) in the case of behaviors 5, 6, 9, and 10. In regard to hypothesis El-C however, the addition of the interaction term did not significantly increase the \( R^2 \) for any of the ten behaviors. It should be noted that the problem of multicollinearity (Darlington, 1968), which occurs in multiple regression when independent variables are highly related, is especially problematic in the present context. The multicollinearity between the two independent variables and the interaction term (a cross product of the former two variables) may have attenuated the increase in the \( R^2 \) observed presently. The data contained in Table 9 provide no support for hypothesis El-C.

The results of the second experiment are reported in Part II below.

**Part II**

The effect of first impressions and response mode on final performance, ability and motivation ratings are addressed in this part. **Final Performance Ratings.** The means and standard deviations of all final ratings provided following observation are offered in Table 11. The results of the 2 x 2 multivariate analysis of variance (MANOVA)
described in Chapter 2, Part II of this document are presented in Table 10.

Table 10

MANOVA Results: Multivariate F Ratios

<table>
<thead>
<tr>
<th>Source of Variance</th>
<th>df</th>
<th>Mult. F</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Impressions</td>
<td>22,127</td>
<td>5.88</td>
<td>.001</td>
</tr>
<tr>
<td>Response Mode</td>
<td>22,127</td>
<td>5.80</td>
<td>.001</td>
</tr>
<tr>
<td>Interaction</td>
<td>22,127</td>
<td>1.10</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

The main effects for first impressions and response mode were both significant (p < .001). The groups significantly differed with respect to their mean vector of final ratings. The interaction between first impression and response mode was nonsignificant.

Inspection of Table 12 for the first impression main effect reveals that final ratings of variables 1, 2, 8, 11, 12, and 14 received the highest discriminant weights (< .30). With respect to the response mode factor, final ratings of variables 5, 7, and 16 received the highest weights. The discriminant weights provide an estimate of the relative importance of each variable in that linear combination separating the groups.

Table 12 contains the univariate F ratios and estimated proportion of total variance accounted for by each variance component. Regarding the first impression factor, 16 of 22 univariate F ratios achieved statistical significance. First impressions accounted for sizeable percentages of total variance with respect to several of the final ratings. Specifically, the first impression factor accounted for 21%,
### Table 11

Means and Standard Deviations of Final Performance, Ability and Motivation Ratings Made Following Observation

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Interpolated</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance Ratings</td>
<td></td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>1. Smile</td>
<td>2.73</td>
<td>1.10</td>
</tr>
<tr>
<td>2. Standard Intro</td>
<td>3.97</td>
<td>.59</td>
</tr>
<tr>
<td>3. Objection One</td>
<td>3.44</td>
<td>.92</td>
</tr>
<tr>
<td>4. Promptness</td>
<td>3.47</td>
<td>.83</td>
</tr>
<tr>
<td>5. Objection Two</td>
<td>4.07</td>
<td>.91</td>
</tr>
<tr>
<td>6. Persistence</td>
<td>4.02</td>
<td>1.26</td>
</tr>
<tr>
<td>7. Closing</td>
<td>3.57</td>
<td>1.08</td>
</tr>
<tr>
<td>8. Name</td>
<td>4.18</td>
<td>1.13</td>
</tr>
<tr>
<td>9. Thank</td>
<td>4.10</td>
<td>.95</td>
</tr>
<tr>
<td>10. Sell Insur.</td>
<td>3.42</td>
<td>1.00</td>
</tr>
</tbody>
</table>

#### Ability Attributions

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Unfavor. M</th>
<th>SD</th>
<th>Favor. M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Smile</td>
<td>2.28</td>
<td>.92</td>
<td>3.47</td>
<td>1.17</td>
</tr>
<tr>
<td>12. Standard Intro</td>
<td>3.57</td>
<td>.59</td>
<td>4.02</td>
<td>.94</td>
</tr>
<tr>
<td>13. Objections</td>
<td>3.21</td>
<td>.74</td>
<td>4.10</td>
<td>.68</td>
</tr>
<tr>
<td>14. Promptness</td>
<td>3.31</td>
<td>.77</td>
<td>4.21</td>
<td>.87</td>
</tr>
<tr>
<td>15. Persistence</td>
<td>3.44</td>
<td>1.05</td>
<td>4.21</td>
<td>1.23</td>
</tr>
<tr>
<td>16. Closing</td>
<td>3.39</td>
<td>.82</td>
<td>3.73</td>
<td>1.10</td>
</tr>
<tr>
<td>17. Name</td>
<td>3.60</td>
<td>1.24</td>
<td>4.00</td>
<td>1.41</td>
</tr>
<tr>
<td>18. Thanks</td>
<td>3.86</td>
<td>.93</td>
<td>4.31</td>
<td>1.11</td>
</tr>
<tr>
<td>19. Sell Insur</td>
<td>3.23</td>
<td>.94</td>
<td>3.44</td>
<td>1.17</td>
</tr>
<tr>
<td>20. Overall</td>
<td>3.10</td>
<td>.60</td>
<td>3.94</td>
<td>1.03</td>
</tr>
</tbody>
</table>

#### Motivation

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Interpolated Favor. M</th>
<th>SD</th>
<th>Summary Favor. M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>21. Motivation</td>
<td>2.86</td>
<td>.74</td>
<td>3.56</td>
<td>.86</td>
</tr>
<tr>
<td>22. Training</td>
<td>2.18</td>
<td>1.20</td>
<td>3.21</td>
<td>1.10</td>
</tr>
</tbody>
</table>
20%, 19%, and 13% of the total variance with respect to the "smile," "response to prospect's first and second objections," and the "promptness of response to the prospect's objections" final ratings, respectively. The first impression source of variance accounted for 31%, 26%, 27% and 23% of the total variance in ability ratings of "smile," "responses to prospect's first and second objections," and the overall ability rating, respectively. Finally, 16% and 21% of the total variance in the motivation and felt need for additional training ratings were accounted for by the first impression factor.

Table 11 reveals that the mean final performance, ability and motivational ratings made in the unfavorable impression conditions were lower than the ratings provided in the favorable first impression conditions for each variable reaching significance (summing across the response mode factor).

The data reported above provide support for hypothesis E2-A. The primacy effect was earlier defined operationally as a phenomena in which early information carries greater weight in overall judgments than does later information (Carlson, 1971; Jones et al., 1972). The data suggest that the primacy effect was in operation in the present context.

Regarding the response mode factor, only 8 of the 22 univariate F ratios contained in Table 11 reached significance. The response mode factor accounted for 44% and 22% of the total variance in ratings of "promptness" and "persistence" (performance). The response mode source of variance accounted for 27% and 18% of the "promptness" and "persistence" ability ratings. Table 10 reveals that the mean final
Table 12

Univariate F Ratios and Estimated Proportion of Total Variance Accounted for by Each Variance Component

<table>
<thead>
<tr>
<th>Performance Ratings</th>
<th>First Impressions</th>
<th>Response Mode</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p² F Variance</td>
<td>% F Variance</td>
<td>% F Variance</td>
</tr>
<tr>
<td>1. Smile</td>
<td>21.55*** 21</td>
<td>.25 00</td>
<td>2.69 01</td>
</tr>
<tr>
<td>2. Standard Intro</td>
<td>1.01 00</td>
<td>.59 00</td>
<td>.18 00</td>
</tr>
<tr>
<td>3. Objection One</td>
<td>17.17*** 20</td>
<td>1.11 00</td>
<td>.16 00</td>
</tr>
<tr>
<td>4. Objection Two</td>
<td>20.84*** 19</td>
<td>5.40* 04</td>
<td>.01 00</td>
</tr>
<tr>
<td>5. Promptness</td>
<td>25.58*** 13</td>
<td>83.61*** 44</td>
<td>.94 00</td>
</tr>
<tr>
<td>6. Persistence</td>
<td>7.94** 07</td>
<td>25.48*** 22</td>
<td>.21 00</td>
</tr>
<tr>
<td>7. Closing</td>
<td>3.89* 03</td>
<td>3.32 02</td>
<td>.21 00</td>
</tr>
<tr>
<td>8. Name</td>
<td>1.34 00</td>
<td>.06 00</td>
<td>1.34 00</td>
</tr>
<tr>
<td>9. Thank</td>
<td>3.32 02</td>
<td>1.41 00</td>
<td>.15 00</td>
</tr>
<tr>
<td>10. Sell Insur</td>
<td>2.59 01</td>
<td>3.10 02</td>
<td>.99 00</td>
</tr>
</tbody>
</table>

Ability Attributions

<table>
<thead>
<tr>
<th></th>
<th>p² F Variance</th>
<th>% F Variance</th>
<th>% F Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Smile</td>
<td>37.15*** 31</td>
<td>.52 00</td>
<td>.08 00</td>
</tr>
<tr>
<td>12. Standard Intro</td>
<td>3.86* 03</td>
<td>1.14 00</td>
<td>1.56 00</td>
</tr>
<tr>
<td>13. Objections</td>
<td>31.68*** 26</td>
<td>9.09** 07</td>
<td>1.01 00</td>
</tr>
<tr>
<td>14. Promptness</td>
<td>47.33*** 27</td>
<td>47.33*** 27</td>
<td>.01 00</td>
</tr>
<tr>
<td>15. Persistence</td>
<td>11.26*** 10</td>
<td>20.46*** 18</td>
<td>.76 00</td>
</tr>
<tr>
<td>16. Closing</td>
<td>4.33* 04</td>
<td>1.64 00</td>
<td>.02 00</td>
</tr>
<tr>
<td>17. Name</td>
<td>.67 00</td>
<td>.01 00</td>
<td>1.12 00</td>
</tr>
<tr>
<td>18. Thank</td>
<td>7.95** 08</td>
<td>.70 00</td>
<td>.01 00</td>
</tr>
<tr>
<td>19. Sell Insur</td>
<td>8.00** 08</td>
<td>3.55 03</td>
<td>2.00 00</td>
</tr>
<tr>
<td>20. Overall</td>
<td>28.98*** 23</td>
<td>18.23*** 13</td>
<td>.55 00</td>
</tr>
</tbody>
</table>

Motivation

<table>
<thead>
<tr>
<th></th>
<th>p² F Variance</th>
<th>% F Variance</th>
<th>% F Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>21. Motivation</td>
<td>15.66*** 16</td>
<td>.287 01</td>
<td>1.27 00</td>
</tr>
<tr>
<td>22. Training</td>
<td>21.76*** 21</td>
<td>3.69* 03</td>
<td>.92 00</td>
</tr>
</tbody>
</table>

¹Estimated proportion of total variance accounted for by that source of variance. (Note that these proportions are read across rows).

²Degrees of freedom = 1,148

*p > .05
**p < .01
***p < .001

Note: Underlined variables indicate standardized weights greater than .30. (Note that relative weights are read down columns).
performance and ability ratings made in the interpolated rating condition were higher than the ratings provided in the summary rating conditions for each rating reaching significance.

The multivariate F for the interaction between first impression and response mode was not significant. The interaction source of variance did not account for more than 1% of the total variance in any of the final ratings. These data provide no support for hypotheses E2-B and E2-C.

The next chapter contains a summary of the results reported above, and a discussion of the findings. Implications of the present studies are also presented.
DISCUSSION

The present research investigated several variables hypothesized to affect observational accuracy. As suggested in the impression formation literature, individuals may form first impressions readily and organize their behavioral observations in order to increase the stability, predictiveness and meaningfulness of their experience (Hastoff, Schneider & Polefka, 1970). It has also been implied that while this process may lead to increases in stability and meaningfulness, it might also contribute to subsequent observational inaccuracies. With this in mind, two experiments were conducted. Experiment one examined the effect of first impressions and sensitivity on response bias within a signal detection theoretical framework. Experiment two investigated the effect of first impressions and response mode on primacy effects in final performance, ability and motivation ratings. It was argued that first impressions and sensitivity result in primacy effects through the influence of response bias. Relevant literature reviews, hypotheses, and methodologies for experiments one and two were offered in Chapters 1, 2 and 3.

This chapter begins by summarizing the results reported in Chapter 4, followed by a discussion of possible methodological shortcomings of the present research. Implications of the findings for various organizational contexts are then presented. Some possibilities
Summary of Results

The results of the present study are outlined below. Results of experiment one are addressed first:

1. Positive first impressions resulted in a higher mean vector of response bias scores than did negative first impressions.

2. The simple main effect of first impressions on response bias was not significantly different for behaviors to which each observer exhibited low sensitivity and those to which each observer exhibited high sensitivity (within subjects).

3. The interaction between the observers' first impression and sensitivity to a particular behavior did not account for a significant amount of response bias score variance after first impressions and sensitivity were entered into the regression equation.

Response bias was previously defined as an observer's tendency to favor a certain response independent of the evidence he is provided. Observers in experiment one provided responses that were in the direction of their first impressions. That is, subjects possessing positive first impressions of the stimulus person reported having observed behaviors that were "more correct" than subjects possessing negative first impressions. Signal detection theory (Green & Swets, 1966) suggests that observers in the positive first impression condition adopted lax criteria relative to the strict criteria adopted by observers in the negative first impression condition. Relative to observers with lax criteria, observers with strict criteria require a greater magnitude of an evidence variable x in order to interpret a
behavior as "correct." The reader is referred to the SDT overview contained in Chapter 1 for definitions of such terms as criteria and likelihood ratios.

It was implied in the present research that observer's first impressions would effect response bias by influencing observer's expectations regarding subsequent observed behavior. These expectancies are referred to as the observers' prior probabilities, denoting a subjective probability of a signal occurrence on a following trial(s). In some SDT experiments, subjects provide their prior probabilities before viewing each of n trials. Green and Swets (1966) state that the observer's likelihood ratio (criterion) is a function of both the observer's prior probabilities and his "payoff matrix". The payoff matrix refers to the observer's perceived utilities associated with correct decisions (gains) and incorrect decisions (costs). The following equation characterizes the relationship between the observers prior probabilities, payoff matrix and the likelihood ratio:

\[
L_c(x) = \frac{p(x/c)}{p(x/i)} = \frac{p(i)}{p(c)} \cdot \frac{G_{cr} + C_{fp}}{G_H + C_m}
\]

where:
- \( p(i) \) = prior probability of an incorrect behavior
- \( p(c) \) = prior probability of a correct behavior
- \( G_{cr} \) = gain associated with a correct rejection
- \( C_{fp} \) = cost associated with a false positive
- \( G_H \) = gain associated with a hit
- \( C_m \) = cost associated with a miss

It should be noted that neither the observers' prior probabilities nor his perceived payoff matrix were measured in the present study. Future research could measure prior probabilities and/or manipulate payoff matrices in order to better link the observers' first impressions and subjective utilities to their resulting level of response bias. Coomb,
Dawes and Tversky (1970) have reviewed studies that describe various methods observers use to revise their subjective probabilities when provided new information (e.g., over trials) during decision-making tasks.

The failure to reject the null hypothesis that sensitivity and first impressions interact to effect response bias may have been due to several factors. Referring to Hypothesis E2-B, sensitivity was an observed partition (Runkel & McGrath, 1972) and not a "true" independent variable systematically manipulated by the experimenter. Sensitivity scores were dichotomized by means of a mean split for each subject's ten sensitivity scores (within-subjects). There may have been individual differences with respect to each observer's sensitivity score variance. The dichotomization may have been inappropriate for those observers whose intrasensitivity variance was low. Treating sensitivity as a between-subjects continuous variable in a regression framework (as was done for hypothesis E2-C) may have been a more powerful method of studying the relationship between sensitivity and response bias. Table 9 reveals that 4 of 10 correlations between sensitivity and response bias (one correlation for each behavior) were significant and in the predicted direction. That is, the higher the sensitivity score for a particular behavior, the lower the response bias score. A more efficient method of investigating the effects of sensitivity on response bias (or in interaction with first impressions) would be to systematically manipulate attribute(s) of the behaviors themselves. Since observers' sensitivity and the nature of the behaviors observed were shown to be related (Table 7), future research might do well to
manipulate behavioral attributes directly. In their research linking ability requirements and task characteristics, in a SDT context, Wheaton, Eisner, Mirabella and Fleishman (1976) have shown that decreased signal duration and signal to noise ratio resulted in decreased observer sensitivity.

A possible methodological shortcoming of the present research should be noted regarding call content. As Table 1 outlines, each of 22 stimulus tapes terminated with the agent either succeeding or not succeeding with respect to inducing the prospect to meet with him to talk about insurance (call outcome). Subjects may have based their overall performance ratings or behavioral reports on subsequent tapes on the call outcome alone. Subjects were instructed to base all ratings on the incorrectness or correctness of the behaviors and not on call outcome, which was characterized as not under the complete control of the agent. Some subjects however, may have still based their judgments on call outcome. A correlation coefficient between call outcome and the number of correct behaviors in that call was computed in order to investigate the extent to which these two variables were confounded. This correlation was not significant \( r = .23, t(20) = 1.05 \), suggesting little relationship between call outcome and the number of correct behaviors in each stimulus tape. The correlation between the overall rating for each call and the call outcome was also non-significant \( r = .29, t(20) = 1.35 \) suggesting that observers performance rating for each call was not influenced by call outcome. Future research should however utilize stimulus material without outcomes to ensure that performance and behavioral attributes (e.g., correctness) are
Additionally, SDT experiments typically involve more stimuli than were presented to subjects in the present study (22). The greater the number of stimulus presentations, the greater the accuracy of the sensitivity and response bias estimates. However, practical constraints are always a consideration when deciding the number of stimuli presentations in a SDT experiment (e.g., boredom).

The results of experiment two are presented below:

4. Subjects possessing favorable first impressions provided significantly higher performance, ability and motivation ratings following all observation than subjects possessing unfavorable first impressions.

The finding that observers with positive first impressions provided significantly higher final ratings than observers with negative first impressions suggests that the primacy effect was operating in the present context. The primacy effect was operationally defined as a phenomenon in which early information carried greater influence than later information in overall ratings or judgments.

5. Subjects in the interpolated response mode conditions provided significantly higher performance and ability ratings following all observation than subjects in the summary response mode condition.

6. The interaction between first impressions and response mode did not significantly effect final ratings provided by rater.

Subjects in the interpolated conditions provided higher final ratings than subjects in the summary conditions. As London and Hakel (1974) suggested, interpolated ratings do not necessarily result in recency effects. Inspection of Table 10 reveals that subjects providing
interpolated responses and possessing positive first impressions provided higher ratings than subjects providing interpolated responses but possessing negative first impressions for those ratings reaching significance (response mode main effect). The same pattern of results appeared in the summary mode conditions.

In the present study, mean differences between subjects possessing negative and positive first impressions were considered to be indicative that primacy effects were operating. Jones et al. (1972) have characterized the order effect paradigm as comparing responses to one sequence of information with responses to the same information presented in reverse order, as was done presently. It should be noted however, that the favorability and order of the stimuli in the present study were confounded. It was therefore impossible to ascertain whether the significant main effect for first impressions on final performance ratings was due to the favorability on the order of the stimuli. It was difficult to ascertain the extent to which primacy or recency effects (order effects) were operating without the addition of two additional experimental groups. One group would provide interpolated responses and the other provide summary responses; however, both groups would receive stimuli in a random or counterbalance order with respect to favorability.

Implications and Future Research

The implications of the present findings for several organizational contexts are many. The observation of behavior is the foundation of such traditional processes as training evaluation and performance appraisal. Observations and subsequent evaluation of trainee
Trainee behavior may be effected by initial contacts within the training context. Initial impressions of the trainee may bias subsequent observation and thus bias the final rating of proficiency given by the trainer. First impressions may also be prevalent in performance appraisal settings, serving to bias observations of job behavior. Within an assessment center context, the present findings may also have several practical implications. For example, should assessors observe each assessee in more than one exercise? The present findings suggest that assessors should not have prior exposure to the assessee before observing behavior in simulated exercises.

The present findings may also have implications for less traditional processes in organizational contexts where observational accuracy is sought. The observation and subsequent rating of tasks along various continua by job analysts is a case in point. Tasks may differ along several dimensions (e.g., things, people, data), or the abilities required to perform them. Job analysts may form wholistic (global) impressions of a job which may effect his/her observations of the tasks with respect to various dimensions.

In SDT, there must exist some justification for classifying stimuli into a particular class. In the present experiment (one), a standard of "correctness" was defined in an a-priori fashion by the experimenter and stimuli were constructed to either conform or not conform to that standard. The criterion for classification may not be as straightforward in other cases. McNicol (1972) offers an example of an experimenter who is interested in differences in sensitivity and response bias between abstract and concrete nouns. Assuming that no
published norms exist regarding the concreteness or abstractness of nouns, the experimenter must pre-test his stimuli to determine which class (if any) a given stimulus belongs. This may be accomplished by expert judgments regarding the degree to which each noun is concrete and/or abstract. Words may then be classified by this method. A similar procedure could be followed for classification of other stimuli in other stimuli classes. For example, tasks may be classified into those characteristic to the high end of the "People" scale (Fine & Wiley, 1972) and those characteristic of the lower end. Those tasks that were reliably rated or classified by job experts would be chosen for the SDT experiment.

Future research should address several issues regarding observational accuracy. The conditions under which accuracy is maximized should be explored, as well as other variables that effect response bias. Several variables have been shown to effect response bias in previous SDT research (Green & Swets, 1966). For example, observers may perceive subjective gains associated with correct decisions and costs associated with incorrect decisions (see Table 1). A cost associated with false positives may in some settings be large relative to the gains associated with hits. The set of gains and costs associated with various decisions are referred to as the observer's "payoff matrix." High costs for false positives and low gains for hits may result in an observer adopting a strict criterion. In the present context, the observer with this payoff matrix may exhibit a tendency to interpret behavior as predominantly incorrect relative to an observer with a different (or no) payoff matrix. It should be noted that payoff
matrices may be used to reduce response bias as well, which may have implications for training programs designed to reduce observational errors.

Another variable that may affect observational accuracy, and hence reduce response bias, is feedback to the observer regarding his past rating behavior. Such feedback could be incorporated into a training program whose objective is to increase observational skills. In a workshop approach in which managers were trained to minimize rating errors in the observation of behavior, Latham, Wexley, and Pursell (1975) provided subjects immediate feedback regarding the accuracy of their ratings made from videotaped behavior. Although the effect of feedback is impossible to ascertain in that study, the workshop approach reduced rating errors relative to control groups. Other studies have investigated the effect of training programs on reducing rating errors (Borman, 1975; Wexley et al., 1973); however, little research has been conducted on training programs designed to decrease observational errors.

A new application of a theory such as SDT may carry with it some benefits as well as costs. For example, SDT does not explain what attributes of behavior in different contexts render some behaviors more detectible than others. Such questions as the conditions under which observers are most sensitive to various behavior, and the nature of individual differences with respect to sensitivity and response bias are not addressed by the model. SDT assumes that observers can reduce their multidimensional stimuli space into a unidimensional decision axis (x); however, the process by which this is accomplished is
unknown. What is needed is a typology of behavioral attributes, such that observers' sensitivities to its elements may be investigated.

The application of SDT to such domains as impression formation however, has several benefits with respect to information gain. The observer is characterized as an active decision-maker instead of a passive receptor of stimuli. Once observers are considered in a decision-making perspective, literature from such areas as decision-making and information processing (e.g., Slovic, 1972), as well as subjective expected utility theory (Edwards & Tversky, 1967) may lend new insight into the observational process of complex behavior.

More research is needed in order to understand the nature of observation with respect to variables that affect accuracy and the conditions under which accuracy is optimized. Observational data serves as a baseline for many organizational functions, necessitating that more research be conducted. As Baker and Schuck (1975) point out, many variables studied within SDT research (e.g., expectancies, payoff matrices and task difficulty) may be important determinants of observer performance. These authors argue that since these variables effect observational accuracy in SDT experiments, they should be controlled in other observational contexts by the use of carefully designed instructions to observers, training of observers, and the explicit manipulation of payoff matrices. In the present study, SDT was considered a useful paradigm in which to study observational and rating behavior.
REFERENCES


Farr, R. S., & York, C. M.  *Amount of information and primacy-recency effects in recruitment decisions.*  *Personnel Psychology,* 1975, 28, 233-238.


APPENDIX A

SUBJECT 0

NOTE: This information will remain strictly confidential and will be used only in the event we need to contact you regarding incomplete responses.

NAME __________________________

SEX (circle one): M F

ACADEMIC MAJOR __________________________

YEAR IN SCHOOL (CIRCLE ONE): Fr Soph Jun Sen

INSTRUCTIONS: Respond to the following items by placing that scale number that best answers the question in the space provided to the right. Use whole numbers only.

YOUR RESPONSE:

1. On how many occasions have you had direct contact with an insurance agent?

   1 2 3 4 5 6

   none one two three four five or more

2. On how many occasions have you been contacted by an insurance agent over the phone?

   1 2 3 4 5 6

   none one two three four five or more

3. On how many occasions have you used rating scales for the purpose of evaluating another person's performance (for example: teacher evaluation rating scales)?

   1 2 3 4 5 6

   none one two three four five or more

4. Do you carry insurance of any type? (circle one)

   YES NO

5. If you answered "yes" to item 4, which type(s) do you carry? (check them off):

   life car
   renters health
   homeowners Other: __________________________

-76-
APPENDIX B

General Instructions to Subjects

This is an experiment concerned with how individuals judge the correctness of behavior they observe. We often are asked to judge how well others perform in many situations (school, work), so this experiment should be of interest to everyone.

Many individuals are contacted daily by insurance agents over the telephone. The agent hopes to meet with the individual for the purpose of selling insurance. For those of us who have been contacted in this manner, all we get to hear is the agent's voice. Today, we will both see and hear one agent contact a series of different prospective insurance buyers on the phone to request appointments. IT WILL BE YOUR TASK TO OBSERVE AND EVALUATE HOW WELL HE PERFORMS ON EACH PHONE CALL. You will be provided guidelines with which to make these evaluations in a few moments.

Shortly after they are hired, insurance agents are provided guidelines outlining the correct way to contact individuals by phone. The purpose of the call is to make an appointment to talk about insurance. How well the agent follows these guidelines determines how successful the agent will be in the long run. New agents are asked to study the guidelines and then try them out by actually calling prospective insurance buyers. If the agent cannot perform the guidelines, he is placed in a training program designed to improve the agent's phone approach. But how does the insurance company know if the agent needs the additional training? The answer is simple: observers like yourselves judge how well the agent performs when he tries to follow the guidelines. If he performs well, he doesn't need the additional training. If he performs poorly on some or all of the guidelines, he is given additional training in those areas he needs it most.

As previously mentioned, you will observe an agent on videotape and rate how well he follows the guidelines suggested by the insurance company. The tapes you are about to see were made with a real insurance agent from Sigourney Life, a company based in the Northeast.
APPENDIX C

GUIDELINES ON MAKING TELEPHONE CALLS

Sigourney Life Insurance Company

1. Smile when first speaking to prospect.

2. Use of Standard Introduction: (1) ask the prospect if he is free to talk on the phone (2) introduce himself (3) introduce company he represents (4) the purpose of the call.

3. Use of proper responses to prospect's objections.

4. Prompt response to prospect's objections.

5. Keep trying to make the appointment, but NOT after THREE objections are received from the prospect.

6. If the appointment is granted: repeat the time, place and day of appointment before hanging up.

   If the appointment is NOT granted: leaving the agent's name, company and telephone number with the prospect before hanging up.

7. Calling the prospect by his/her correct name throughout the phone call.

8. Thanking the prospect regardless of whether the appointment was made or not.

9. NOT talking about insurance when the prospect asks what sort of insurance his company offers (that is, the agent should not try to sell insurance over the phone when the prospect asks about what Sigourney offers).
GUIDELINE NO. 3 : USE OF PROPER RESPONSE TO PROSPECT'S OBJECTIONS

Research by Sigourney has revealed that there are about four main types of objections commonly raised by prospects on the phone. These common objections are listed in column 1. Column 2 contains the correct response. Column 3 contains examples of how agents should not respond.

<table>
<thead>
<tr>
<th>COLUMN 1: COMMON OBJECTIONS</th>
<th>COLUMN 2: CORRECT RESPONSES</th>
<th>COLUMN 3: POOR RESPONSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. CAN'T AFFORD INSURANCE</td>
<td>INFORM PROSPECT THAT MORE INSURANCE MAY NOT MEAN MORE COST</td>
<td>INFORM PROSPECT THAT BECAUSE INSURANCE IS SO IMPORTANT, HE CAN'T AFFORD NOT TO BUY.</td>
</tr>
<tr>
<td>2. ALREADY HAS AN INSURANCE AGENT, LIKE A NEIGHBOR OR BROTHER</td>
<td>LET THE PROSPECT KNOW THAT THIS AGENT IS PROBABLY DOING A GOOD JOB, BUT THAT THE AGENT CAN PROVIDE ADDITIONAL COVERAGE</td>
<td>AGENT SAYS THAT HE CAN DO A BETTER JOB</td>
</tr>
<tr>
<td>3. PROSPECT REQUESTS THAT SOME MATERIALS BE MAILED TO HIM INSTEAD OF HAVING THE APPOINTMENT</td>
<td>INFORM PROSPECT THAT SINCE EACH POLICY IS ADJUSTED TO MEET THE CLIENT'S SPECIAL SITUATION, A HOME VISIT IS NECESSARY.</td>
<td>MAKE UP SOME EXCUSE LIKE THERE ARE JUST TOO MUCH TO MAIL OR PROSPECT IS TOO DUMB TO READ IT ANYWAY.</td>
</tr>
<tr>
<td>4. PROSPECT SAYS HE JUST IS NOT INTERESTED IN LIFE INSURANCE.</td>
<td>IN SO MANY WORDS, ASK THE PROSPECT WHY HE FEELS THAT WAY.</td>
<td>TELL PROSPECT THAT EVERYONE MUST BE INTERESTED IN LIFE INSURANCE FOR THEIR OWN PROTECTION.</td>
</tr>
</tbody>
</table>
Guideline Quiz

Respond to the following "true-false" questions by circling the correct response. BASE YOUR ANSWERS ON THE GUIDELINES ONLY.

1. The insurance agent should smile when first speaking to the prospect.  
2. The Standard Introduction includes: asking the prospect if he is free to talk on the phone, the insurance agent's name and company he represents and the purpose of the phone call.  
3. The Standard Introduction includes names of personal friends that have already bought insurance from the agent.  
4. If the prospect says that he can't afford to buy insurance, the agent should tell him that he really could get the money up if he wanted to.  
5. After receiving objections from the prospect, the agent should keep trying to get the appointment until the prospect finally gives in.  
6. If the agent gets the appointment, he should repeat the time, day and place of the appointment before hanging up.  
7. The agent should call the prospect by his correct name all the way through the phone call.  
8. The agent does NOT have to thank the prospect if he doesn't get the appointment.  
9. When asked by the prospect to talk about insurance SIGourney has to offer, the agent should reveal all the details about the policies he sells as he possibly can.
APPENDIX E

Instructions for Rating Scale Use

SUBJECT NUMBER_________________

INSTRUCTIONS: You will shortly observe several phone calls. The agent making these calls will contact a different prospect on each call. FOR EACH CALL, indicate how well the agent performs the guidelines. It is best if you rate his performance on each guideline as you see it, instead of waiting until the entire call is finished.

Rate how incorrectly or correctly the agent performs each guideline using the rating scale pictured below. DO NOT PLACE YOUR RESPONSE ON THE SCALE ITSELF. PLACE YOUR RESPONSE IN THE SPACE PROVIDED TO THE RIGHT OF EACH GUIDELINE.

The rating scale:

1 2 3 4 5 6
Totally Mostly Slightly Slightly Mostly Totally
incorrect incorrect incorrect correct correct

Example of using the scale:

Suppose you felt that the agent was mostly incorrect when he used the Standard Introduction. You would then place the number "2" (USE WHOLE NUMBERS ONLY) in the space provided next to the appropriate guideline as shown below:

How correctly did the agent use the standard introduction 2

REMEMBER: base your judgement solely on the guidelines we reviewed. Your task is to judge how well the agent performs each guideline. Notice that a rating of "1", "2" or "3" are different degrees of incorrectness. Ratings of "4", "5" or "6" are different degrees of correctness.

After viewing the call and rating how well the agent performed on all ten guidelines, you should fill the page over and provide one more rating. This additional rating indicates your overall impression of how well the agent performed on that call. You will use the scale pictured below for this overall rating:

1 2 3 4 5 6
Totally Mostly Slightly Slightly Mostly Totally
incorrect incorrect correct correct

Example of using this scale: If you thought the agent performed slightly correctly on the call you just observed, you would place a "4" in the blank space provided to the right as shown below:

Considering the entire phone call, how good a call did the agent make?________
Scales Used by Observers to Rate Individual Calls

Rate how incorrectly or correctly the agent performed each guideline using the rating scale pictured below:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Totally incorrect</td>
<td>Mostly incorrect</td>
<td>Slightly incorrect</td>
<td>Mostly correct</td>
<td>Totally correct</td>
<td></td>
</tr>
</tbody>
</table>

How incorrectly or correctly did the agent:

1. Smile when first speaking to prospect
2. Use the entire "standard introduction"
3. Respond to the prospect's first objection
4. Respond to the prospect's second objection
5. Promptly respond to the prospect's objections
6. Keep trying, but not after 3 prospect objections
7. Repeat time, day & place or his name, company and phone number before hanging up
8. Call the prospect by his/her correct name
9. Thank the prospect regardless of outcome
10. Not sell insurance when asked specifically for information over the phone

Use the same scale pictorial above to rate the agent's overall performance on this call.

Considering the entire call as a whole, how good a call did the agent make?
APPENDIX G
Sample Call Content

Call Number 12
Propects Name Kerr

Smile Yes No

Prospect: Hello?

Agent: Mr. Kerr, My name is Steve Cooper and I represent the Sigourney Life Insurance Company. Are you free to discuss some ideas with me now?

Prospect: Sure, Cooper, shoot!

Agent: Mr. Kerr, I'd like to discuss an idea or two with you that several individuals have found to be quite instrumental in meeting their insurance needs. I feel one of these ideas may help you too, and that's why I'd like to arrange a time for us to meet for the purpose of exposing you to an idea or two. May I see you Tues at 10:00?

Prospect: That's just great, Mr. Cooper. I'm always open to new ideas. Why don't you drop some materials in the mail- I'll look over them and give you a call if I'm interested. Sound fair?

Agent: (long pause) I appreciate your concern for some information before we meet, Mr. Kerr, however, the policies we offer serve you best if they are tailored to fit your specific insurance needs. That's why I'd like show you some materials in person. Is Tues at 10:00 a good time to meet?

Prospect: Surely there must be some preliminary information you could supply, Mr. Cooper. Why not mail at least that much?

Agent: (long pause) I'd like to do that, Mr. Kerr, but I feel an obligation to present this information to you personally. That way, I can assess your needs better and suggest some appropriate coverage. Perhaps we can meet Tues at 10:00?

Prospect: Well, if you refuse to mail me any literature, the least you can offer me some insight into those ideas of yours? Are they term or whole life?

Agent: Mr. Kerr, I'd rather save the details of those ideas for our meeting, so that you understand them more clearly. In general, however, the idea that especially should interest you concerns a new concept in term insurance which has a more attractive payment schedule than any of its kind.

Prospect: Well... I'm not so sure... maybe we shouldn't meet after all.

Agent: I understand your hesitancy, Mr. Kerr, but I'm sure one of those ideas will benefit you over the years. Would Tues at 10:00 be a convenient time to meet and discuss this matter?

Prospect: Let's just forget the meeting, but I'll keep you in mind if the need arises.

Agent: That's good, Mr. Kerr, it's been good talking with you, and I'd appreciate your thinking of Sigourney in mind- just give Steve Cooper at 485-1128. Until then, Goodbye and have a good week, Mr. Kerr.