THE AMBIGUITY MODEL: A DIMENSIONAL ANALYSIS OF EXPERIMENTER EXPECTANCY EFFECTS

A Dissertation

Presented to

the Faculty of Graduate Studies

University of Manitoba

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Morris Shames

October, 1971

ABSTRACT

This investigation was designed to test the ambiguity model in experimenter expectancy effects. This model postulates that the subject attends to the experimenter in particular and to extra-task cues in general for direction when faced with an ambiguous judgment situation. Ambiguity results from complexity of judgment where the subject is not able to consistently encode a number of dimensions employed in making rating judgments over a series of stimuli. This model was investigated, first, by providing subjects with extremes in dimensional judgment conditions, i.e., unidimensional and multidimensional stimuli. Secondly, ambiguity was tested as a graduated phenomenon suggesting that an increased magnitude of expectancy effects was linearly related to graduated increments in dimensionality of judgment.

Through pilot research, 10 chromatic stimuli were selected on the basis of the best approximate match of their means and standard deviations with those of 10 achromatic stimuli in order to construct task stimuli which were matched in all respects save for extremes in dimensionality. This assumed dimensional disparity was empirically established by means of paired-comparisons data which yielded indices of intrasubject intransitivity and rank-orderings which yielded indices of intersubject intransitivity. Then, in a first study, these two sets of stimuli were employed in a traditional expectancy paradigm with positive and negative expectancies induced in both stimulus conditions. The obtained results supported the ambiguity model since expectancy effects were found only under the ambiguous, multidimensional,

judgment conditions imposed by the chromatic stimuli.

For purposes of the second study, it was necessary to construct an experimental task appropriate for the test of the graduated ambiguity model. Through pilot research, the most commonly employed dimensions of success-failure, in judging person perception stimuli, were empirically determined. Then, by presenting subjects with one-, three-, or five-dimensional definitions of success-failure, graduated conditions of dimensionality were established. These levels of dimensionality were instructionally presented in a traditional expectancy paradigm with positive and negative expectancies induced across all three conditions. Generally, the results did not support the ambiguity model which related an increment in dimensionality to increased magnitude of expectancy effects. However, it was suggested that this failure was less a matter of conceptualization than operationalization. Postexperimental questionnaire data indicated that subjects perceived the unidimensional definition as more difficult to use than did subjects in the other conditions. Thus, the unidimensional condition might more properly be considered the highest level of dimensionality, i.e., unrestricted dimensionality. Moreover, the quantitative manipulation of dimensions appeared to be less appropriate in this study than in Study I since the person perception task, despite the specification of dimensional criteria, does not readily fall into a task-ability category. Complexity, in terms of the number of dimensions encoded in a single judgment, is thus less appropriate for the person perception task because of the confounding aspects of this self-quality,

attitudinal type of task. In all, then,this may have yielded an inadequate testof the graduated ambiguity model.

Generally, the ambiguity model was considered heuristic in explicating those conditions which underlie the subject seeking out extra-task, directive cues when faced with an ambiguous judgment situation. This, then, facilitates the transmission of the experimenter's expectancy. More specifically, it was suggested that the ambiguity model be considered in terms of an all-or-nothing encodability of dimensions into consistent judgments over a series of stimuli since a graduated, ambiguity model might be an inappropriate conceptualization. However, further investigation into these parameters is necessary to elucidate this conjecture.

ACKNOWLEDGEMENTS

Drs. Hugh McGinley and Ed Seidman are accorded thanks for their efforts in the development of this research. Especial thanks are due Drs. John Adair and Marion Aftanas without whose incisive criticism, perspicuous advice, and general guidance this research would not have grown to fruition.

Mrs. Lynne Dobbs' good nature and competence in typing this manuscript have added immeasurably to this study.

My wife, Terry, deserves unlimited thanks for having countenanced life with me during the creation, evolution, and consummation of this research. Mostly, however, I thank her because she was a wellspring of support and an excellent sounding-board during the development of this work.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTORY OVERVIEW

The research subject has long been a source of concern to experimental psychology. Indeed, Orne (1962) documents that early in the history of experimental psychology, there was the suggestion that the subject of the psychological experiment was too often regarded as a stupid automaton when, in fact, he was more likely an active participant in psychological research (Pierce, 1908). It was proposed, further, that the subject responds in a willing and cheerful manner, most often with the full intention of subtly assisting the experimenter in the successful execution of his research.

More recently, Riecken (1962), in his now-classic prolegomenon for experimental social psychological research, asserted that the subject's attitude and the effects of the experimental environment upon him were crucial determinants of the outcome of research. In conceptualizing the subject's role in experimental research, he postulated that:

> "...the subject has more than one problem. One is the "task" that the experimenter sets. Another is what we may, for convenience, call his "deutero-problem", meaning that his personal problem is defined by the three aims mentioned above: attainment of reward, divination of the experimenter's true purpose and favorable self-presentation.....the deutero-problem plays some role in all experimental situations and for all subjects, and may, on some occasion, be more important than the "task" or "treatment" in explaining results" (p. 34).

Similarly, in considering the subject's role in the experimental situation, Orne (1962), too, suggested that the psychological experiment

constitutes a special form of social interaction since the subject must respond to the experimental variables under investigation and one other set of variables which he refers to as the demand characteristics of the experimental situation. Included under this schema are the effect of the physical context of the experimental situation upon the subject's behavior and his attempt to divine the experimenter's true purpose in conducting the research. The subject, it is suggested, attempts generally to be "good" inasmuch as he is motivated to validate the experimental hypothesis largely in the interest of making a contribution to and advancing the aims of science. This model, then, suggested that subjects influence the outcome of research -- apart from the experimental variable being investigated and in a manner often unanticipated by the experimenter.

Rosenberg (1965) has postulated a schema which is somewhat supportive of this position. He, too, has suggested that subjects are motivated to be "good", but primarily because they enter the experimental situation with some degree of evaluation apprehension; "that is, an active, anxiety-toned concern that he win a positive evaluation from the experimenter, or at least that he provide no grounds for a negative one" (p. 29). To serve this end, the subject develops hypotheses related to his gaining a positive and for avoiding a negative evaluation and this, apart from the experimental variables being investigated, exerts systematic effects upon the outcome of research.

These formal schemata suggested that the subject indeed influences the outcome of research in such extraneous ways as to impugn the

research process itself. The conduct of valid research, then, was contingent upon a consideration of the systematic influences exerted by the subject upon the outcome of research.

This reasoning is equally applicable to the role of the experimenter. Until recently, however, the prevalent model of the experimenter -- as Friedman (1967) points out -- has been that:

> "...experimenters are created equal; that they have been endowed by their graduate training with certain interchangeable properties; that among these properties are the anonymity and impersonality which allow them to elicit from the same subject identical data which they then identically observe and record" (Pp. 3-4).

Rosenthal (1966) pointed out the invalidity of this kind of model, suggesting instead that:

> "Some of the complexity of man as we know it from his model, the research subject, resides not in the subject himself but rather in the particular experimenter and in the interaction between subject and experimenter" (p. vii).

There are excellent reviews of the plethora of research which has empirically demonstrated that different experimenters differentially affect the outcome of their research (Rosenthal, 1964; 1966; 1969a). It has been found that experimenters who differ in their psychological or biological attributes may, as a result of this difference, affect their subjects' performances. Experimenters with expectations apropos of the outcome of their research have been found to make errors of observation and interpretation when dealing with their data and influence their subjects' behavior during data collection. Often these errors and the influence exerted are in conformity with their expectations resulting in what has formally been called the

experimenter expectancy effect. It has been suggested that much of this influence is the unintended transmission of subtle cues from the experimenter to the subject, an intricate feedback loop of covert communication (Rosenthal, 1967a).

However, the experimenter expectancy effect has not been consistently demonstrated and not all of the experimental tasks employed have been uniformly susceptible to this influence. Based on analysis of the properties of some of these tasks, the present study seeks to determine those conditions under which the influence of the experimenter is either facilitated or impeded. To this end, an ambiguity model of experimenter influence, defined in terms of the dimensionality of judgment required from the subject, is proposed. Thus, a complex task requiring a multidimensional judgment imposes a sufficiently ambiguous stimulus condition as to facilitate experimenter influence while a simple judgmental task obviates the subject's need to seek out additional, unintended but directive information from the experimenter.

CHAPTER II

HISTORICAL REVIEW OF EXPERIMENTER INFLUENCE

The earliest systematic concern with the influence of the experimenter upon his research is not found in experimental psychology. Inasmuch as there is a similarity between the survey researcher and the experimenter, survey research provides some of the earliest relevant research into the influence of the researcher upon his research. As early as 1929, Rice found that the data obtained from charity-case respondents were remarkably consistent with the expectations the interviewers held for them, despite their skill and training in the interview process. Harvey (1938) deliberately manipulated the expectations of his interviewers and then cautioned them to avoid any form of bias in the formation of their impressions of their interviewees; yet bias which was consonant with the induced expectations still occurred. To alleviate the problem of interviewer bias, Mosteller (1944) formulated a procedure for its cancellation through the use of a heterogeneous sample of biases.

Survey researchers have also recognized the potentially contaminating effects of error, especially intentional error, in the conduct of surveys (Cahalan, Tamulonis & Verner, 1947). There has been, then, a continual examination of the extraneous influence of the interviewer in the interview setting.

In contrast to survey research, experimental psychology did not initiate a systematic inquiry into experimenter influence until

considerably later. There were, however, several early, noteworthy investigations. For example, Cason and Cason (1925) rightly determined that the experimenter's sex could effect the subject's response; however, only a sample of two experimenters, one of each sex, was employed in this study. In another early paper, Rosenzweig (1933) pointed out that among other potential sources of error in the experimental situation, the experimenter could cause 'suggestion-error' where subjects react to the experimenter's personality. Robinson and Rohde (1946) found that the experimenter's biosocial attributes, such as the Jewishness of his name and appearance, can influence the outcome of his research -- especially when the subject matter of the research deals with things Jewish. In addition, Lord (1950) demonstrated that variations in Rorschach response resulted from the experimenter's psychosocial attributes, the "warm" or "cold" behavior exhibited during the administration of the Rorschach test.

Another experimenter influence, the experimenter's expectancy, was investigated by Stanton and Baker (1942). Experimenters' expectations were manipulated by providing them with a scoring key, half of which was correct and half incorrect, for a nonsense geometric figures task. The results the experimenters obtained were consonant with these manipulated expectations. This was a landmark study in that it empirically demonstrated the effects of manipulating the experimenter's expectancy, a most important experimenter influence. Two subsequent replications of this study, however, failed to obtain similar results (Friedman, 1942; Lindzey, 1951).

Recent Research into Experimenter Influences

The most recent and systematic investigation into experimenter influences in behavioral research was initiated by Rosenthal (1964; 1966; 1969a) who, in carefully analyzing the problem, categorized experimenter influences with a view toward gaining further understanding of the processes involved. He has categorized the research into: (1) Effects of the experimenter which do not directly influence the subject's behavior,[1] and (2) Effects of the experimenter which do exert an influence on the subject's behavior.

Effects of the experimenter which do not influence the subject's behavior. It has been found that although the experimenter may not influence the subject's behavior directly, indeed, he may even have no contact with the subject, he may yet systematically influence the outcome of research when cast in the role of observer, or when called upon to analyze and interpret data already collected. Of primary concern here has been the treatment of systematic error -- both intentional and unintentional.

There have been some compelling researches demonstrating observer error in experimental research. Rosenthal and Halas (1962), using academically advanced and experienced experimenters, found for the most part statistically significant differences between experimenters

[1]The reader should recognize that this category is only a matter of theoretical perspective, since veridically, as Orne (1962) and Riecken (1962) postulate, the experimenter's very presence in an experimental setting always exerts some influence on the subject's subsequent behavior.

in terms of their observations of planaria responses. These, it was alleged, were attributable to immanent differences in the experimenters themselves rather than to any particular experimental manipulation. Significant differences in observation occurred too when the experimenter's expectations were instructionally manipulated (Cordaro & Ison, 1963). In the first part of this study, all experimenters were led to expect that half of the planaria they would observe would be "high-responders" and the remaining half, "low-responders". In a second part of this study, one-half of the experimenters were led to believe that they would be observing "high-responding" planaria only, while the other half of the experimenters anticipated observing "low-responding" planaria only. In actual fact, all of the planaria were drawn from the same population. The statistically significant difference between groups of experimenters in terms of their observations of planarian response was in accord with the expectancies induced in the experimenters.

In addition to errors of observation, there is evidence to suggest that experimenters make recording and computational errors in their collection and analysis of data, often in conformity with their expectations (Rosenthal, Friedman, Johnson, Fode, Schill, White, & Vikan, 1964). Using a person perception task where subjects were required to rate photographs of faces on a continuum of success-failure, it was found that the incidence of recording error by experimenters was not generally high, but where it did occur it tended to favor the experimenter's expectation for his subjects. The experimenter's

expectation regarding the success-failure judgments that subjects would make were instructionally manipulated by the principal investigators. It was also found that computational error was distinctly related to recording error, but the consistency of direction was found to be less clear. More recently, there have been attempts made to assess the relative contribution of observer/recorder error to the experimenter expectancy effect (Johnson & Adair, 1970; 1971). It was generally found that both, the experimenter's expectancy and observer/recorder error, contribute jointly to the experimenter expectancy effect although the authors suggest that a greater portion of the variance was attributable to the experimenter's expectation.

The bias of the experimenter's expectation is often manifest in his interpretation of the data he collects. Rosenthal (1966) pointed this out by describing an especially lucid study by Cahen (1965). It was found here that different prospective schoolteachers, with different background information on the children employed in this study, interpreted their academic readiness tests differently -- most often in accordance with the background information with which they were furnished.

The research reviewed thus far has considered only unintentional error. However, the possibility of intentional error, although difficult to determine in most cases, should also be considered. In one study, Rosenthal and Fode (1963a) found instances of experimenters deviating from the programmed procedure outlined for them in order to prod their subjects to run the maze as desired. In another, more convincing study, experimenters running rats they supposed were "Skinner Box-bright" and

others running rats they expected would be "Skinner Box-dull", under seven conditions of operant learning, intentionally erred during the conduct of the research by fabricating some of their data (Rosenthal & Lawson, 1964).

There is, then, considerable support for the postulation that the experimenter influences the outcome of his research even without the exertion of direct influence upon his subject's behavior. In fact, even the major critics of the research into experimenter influences concede the importance of this type of experimenter influence (Barber & Silver, 1968). These authors suggested that:

> "...the experimenter brings his results in line with his expectancies or desires by intentionally misjudging, misrecording, or misreporting his results, or by intentionally administering verbal or non-verbal reinforcement, or by intentionally transmitting paralinguistic or kinesic cues, the mediating behaviors are very difficult to differentiate from those that are commonly termed 'cheating'" (p. 21).

Effects of the Experimenter which influence the subject's behavior. In general, the research included under this category has focused upon the experimenter's biosocial and psychosocial attributes, and upon the experimenter expectancy effect (EEE), often referred to as the experimenter bias effect (EBE). Biosocial effects are those influences which are exerted on subjects' behavior as a result of the experimenter's biological characteristics, such as sex, age and race. Similarly, as a result of the experimenter's psychological attributes, such as anxiety and intelligence, subjects' behavior during the conduct of the experiment is accordingly influenced. Thus, systematic differences in subject behavior often results from corresponding differences in the

experimenter's biosocial and psychosocial attributes as well as his expectancy, a phenomenon which will be discussed in a later section.

The biosocial attributes of the experimenter have received considerable research attention (Rosenthal, 1966; 1969a). One of these, the experimenter's sex, has been scrutinized within such varied experimental contexts as a marble-dropping task (Stevenson, Keen, & Knights, 1963), a marble-sorting task (Stevenson & Allen, 1964) and a person perception task (Friedman, 1964; Katz, 1964; Rosenthal, 1963; Rosenthal, Friedman, & Kurland, 1966; Rosenthal, Persinger, Mulry, Vikan-Kline, & Grothe, 1964a) . Generally, it was found that significant differences in subjects' responses resulted, attributable mainly to the sex of the experimenter. Moreover, this was found to interact with the sex of the subject so that dyads of heterogeneous sex, especially male experimenters running female subjects, constitute that condition which is optimal both for performance in experimental tasks and for the transmission of the experimenter's expectancy.

Another biosocial attribute, the experimenter's age, has been found to influence the outcome of research (Ehrlich & Riesman, 1961), but as in the case of the experimenter's sex, it interacts with the subject's age. The experimenter's race, too, has been found to exert an influence on the outcome of research (Sattler, 1970), but the findings, as in the case of studies utilizing physiological measures (Bernstein, 1965), have not always been consistent. Generally, however, the experimenter's race interacts with the race of the subject in partially determining the subject's responses (Hyman, Cobb, Feldman,

Hart, & Stember, 1954; Summer & Hammonds, 1966) in that dyads of the same race, such as Negro experimenters with Negro subjects, yield the most honest responses from subjects. It was also found that the experimenter's race interacts with the subject's prejudice since Smith and Dixon (1968) demonstrated that white, prejudiced subjects conditioned with white experimenters but not with Negro experimenters, in a Taffel (1955) procedure of verbal conditioning. In addition to race, the experimenter's religion was found to interact with the subject's prejudice inasmuch as Hyman et al.(1954) found that a random sample of subjects responded less anti-Semitically to Jewish than to Gentile interviewers.

These data, then, suggest that the experimenter's biosocial attributes may be critical for the research process, especially since they often interact with the subject's biosocial attributes. Based on these data, Rosenthal (1966) has suggested that the most general conclusion to be drawn is that "subjects tend to respond in the way they feel to be most proper in the light of the investigator's (biosocial) attributes". (p. 61).

Research into the experimenter's psychosocial attributes tends to be somewhat more equivocal. For example, based upon investigations into the effects of the experimenter's anxiety on the data he collects from his subjects (McGuigan, 1963; Rosenthal, Kohn, Greenfield, & Carota, 1965; 1966; Rosenthal, Persinger, Vikan-Kline, & Mulry, 1963), Rosenthal (1966) suggested that:

"...it seems safe to conclude that the experimenter's anxiety

> level (or perhaps adjustment level) may affect the subject's responses for a variety of tasks; but the nature of the effect is not predictable on the basis of our current knowledge" (p. 63).

This conclusion is equally applicable to that research which investigated the experimenter's need for social approval (Marcia, 1961; Mulry, 1962; Rosenthal, Kohn, Greenfield, & Carota, 1965; 1966; Rosenthal, Persinger, Vikan-Kline, & Mulry, 1963).

Other psychosocial variables, such as the experimenter's hostility (Barnard, 1968; Sarason, 1962; Sarason, 1965; Sarason & Minard, 1963), his authoritarianism (Mulry, 1962; Rosenthal, Persinger, Mulry, Vikan-Kline, & Grothe, 1964a), his intelligence (Mulry, 1962) and his dominance (Ehrlich & Riesman, 1961; Symons, 1964) have been found to exert some influence on subjects' responses during the conduct of research. However, since these variables often interact complexly with other variables, such as the experimenter's prestige and age, it is difficult to draw a univocal inference with respect to their precise contribution.

Less ambiguous have been the findings that "warm-behaving" experimenters tended to elicit more competent and pleasant responses from their subjects than do "cold" experimenters (Luft, 1953; Reece & Whitman, 1962; Spires, 1960). With respect to prestige, high status experimenters tend to elicit behavior which is consonant with their expectations for their subjects, but these responses generally tend to be less pleasant in nature than those elicited by warm experimenters (Prince, 1962; Rosenthal, Kohn, Greenfield, & Carota, 1966; Sarason & Minard, 1963). In addition, it was found that situational variables

such as experience as an experimenter (Brogden, 1962; Turner & Coleman, 1962) and the experimenter's previous contact with the subject (Jourard, 1968; Stevenson, Keen, & Knights, 1963) clearly exert some influence on the outcome of research.

The experimenter's biosocial and psychosocial attributes, as well as relevant situational contingencies, have all been shown to bear some relationship to the execution of his research. However, the precise function describing these variables is difficult to specify since experimenter influences do not always yield systematic differences in the performance of subjects (McGuigan, 1963). Further research is necessary for the clear specification of those cases where experimenter influences yield systematic differences in the performance of subjects.

Experimenter Expectancy Effects

The biasing effect of the experimenter's expectancy has been extensively researched (Rosenthal, 1969a), with somewhat less equivocal results than was found with research into the experimenter's biosocial and psychosocial attributes. Experimenter's expectancies are usually induced by means of the principal investigator instructing his experimenters with respect to the performance they should expect from their subjects. EEE is then operationally defined in terms of obtained differences in performance between groups of subjects which have been run by experimenters with differing expectancies. Thus, by virtue of his expectation for the subject, among his other attributes, the experimenter may directly, although unintentionally and subtly, influence his subjects' behavior -- an empirical finding which has roots in a

curious, historical event.

Pfungst (1911) documented the case of a seemingly gifted horse, Clever Hans, who was assumed to be mathematically brilliant. Upon investigation, however, it was found that he was more astute than brilliant since his unerring responses to the mathematical problems posed were the result of his ability to perceive subtle and unintended cues communicated unintentionally by his questioners. The leap from this historical phenomenon to the unintentional communication of subtle cues by experimenters to subjects during the conduct of research is logically valid and, in fact, has been made consistently (Rosenthal, 1964; 1966; 1969a).

At the outset a paradigmatic study was employed in the systematic investigation of EEE (Rosenthal & Fode, 1963b). This was a person perception task consisting of 10 photographs of human faces. Subjects were to estimate the "success-failure" experienced by these stimulus-objects by means of a 10-point, algebraic rating scale ranging from EXTREME FAILURE (-10) to EXTREME SUCCESS (+10) in step-intervals of one. The principal investigator instructed one-half of the experimenters to expect an approximate, mean -5 rating (failure) from their subjects, while the other half of the experimenters were led to anticipate an approximate, mean +5 rating (success) from their subjects, based presumably upon similar, previous research on personality tests. Thus both groups of experimenters were treated precisely alike save for this instructional manipulation. The results obtained showed a highly significant difference between the data collected by the two

groups of experimenters, empirically demonstrating the effects of the experimenter's expectancy on the outcome of his research.

The earliest of these investigations consistently and dramatically supported EEE (Fode, 1960; 1965; Rosenthal & Fode, 1961; 1963b). This suggested that the variations in the data obtained in these studies were apparently due to the variations instructionally induced in the experimenters. A proliferation of research into EEE and its interaction with other nontrivial determinants of subject behavior ensued, with the person perception task being employed as the easily-manipulable experimental context. Although this prodigious body of research has not been uniformly consistent in its support of EEE, a summary of these studies -- some 57 in all -- presents fairly convincing evidence in support of EEE, under conditions of the person perception task (Rosenthal, 1969a, pp. 223-228).

More uniform support for EEE has been obtained in studies employing animal subjects. In a study mentioned previously, Rosenthal and Fode (1963a) led half of their experimenters to believe that they would be running "maze-bright" rats while the other half were told that their rats would be "maze dull". In actual fact, all of the rats were drawn from the same albino population, but a significant difference in the performance of the two groups of rats was nonetheless obtained. The authors interpreted this as due largely to the experimenter's outcome expectations. Similarly, Rosenthal and Lawson (1964) found support for EEE using seven distinct criteria of operant learning with rat subjects. In this study, the expectancies induced were related to

"Skinner Box-brightness" and "Skinner Box-dullness". Even more dramatic support for EEE was obtained in an experiment with lesioned and unlesioned rats (Burnham, 1966), where it was found that the effects of the experimenter's expectancy were at least as great as those of the lesions themselves. The observed and recorded responses of planaria were also found to be consonant with the experimenter's expectancies (Cordaro & Ison, 1963).

Although this research with both human and animal subjects constitutes fairly reasonable evidence for the general nature of EEE, Barber and Silver (1968a) have challenged this generality, based on their assessment of 31 studies which investigated EEE.[2] They suggested that the majority of these studies, nineteen in toto, did not clearly demonstrate expectancy effects, and where they did, often inadequacies in the analysis or design were at fault. Rosenthal (1968) rejoined this criticism, suggesting first that EEE is a highly general phenomenon, based on the combined probabilities of the 31 studies in question and, secondly, that the null hypothesis decision procedure employed by Barber and Silver in their review is not a sufficiently

---

[2]It should be emphasized that Barber and Silver's (1968a) conclusions do not constitute an essential quarrel with the validity of the experimenter expectancy phenomenon per se. In fact, some of their conclusions are supportive of those drawn by Rosenthal (1966; 1969a). Predominantly, these point out that errors of observation, recording and interpretation, mostly, account for the systematic, extraexperimental exertion of influence upon the outcome of research. There is also agreement that paralinguistic and kinesic cues are, in part, the mediating mechanisms for the transmission of expectancies. Their point, therefore, appears to focus primarily upon the generality and ready demonstrability of EEE, not its validity.

valid tactic for the purpose of rejecting the generality of EEE. This disputation has been ongoing (Barber, 1969; Barber, Calverley, Forgione, McPeake, Chaves, & Bowen, 1969; Barber & Silver, 1968b; Rosenthal, 1967a; b; c;,1968; 1969c). Apropos of the generality of EEE, Rosenthal (1969b) has pointed out that expectancy research has been carried out in diverse research areas and in more than 40 tasks employing experimental formats other than the person perception task, 66% evinced significant expectancy effects ($p<.10$). In contrast, 36% of more than 60 studies employing the person perception task have done so. There is then, to this extent, reasonable support for the generality of EEE.

Expectancy effects in diverse research areas. This research into EEE in areas other than person perception is useful in that it relates to the question of the generality of EEE, but more importantly, as Levy (1969) points out, it might contribute to the development of a theoretical schema elucidating the nature of EEE. An understanding of those conditions that either facilitate or impede EEE, as occurs in some tasks and not in others, would in turn shed light upon the nature of EEE and, perhaps its mediation. The signal importance of this research, then, lies in an analysis of the susceptibility of some research areas to EEE and the relative immunity of other areas.

In the area of projective testing, expectancy effects have been found with the Rorschach test (Masling, 1965; 1966) and the Holtzman Inkblot test (Marwit, 1968; Marwit & Marcia, 1967). Although there has been one study in this area which has failed to yield EEE (Strauss,

1968) projective tests, in general, appear to be somewhat susceptible to the influence of the experimenter.

Expectancy effects have also been found under conditions of the structured laboratory interview. It was found that experimenters who were led to expect a high degree of awareness in a verbal conditioning study obtained data from their subjects which were in accord with this expectancy; and these data differed significantly from those obtained by experimenters who were led to expect a low degree of awareness from their subjects (Rosenthal, Persinger, Vikan-Kline, & Fode, 1963b). In addition, Rosenthal (1969a) described a study by Jenkins (1966) where it was found that factual information about a stimulus person could be communicated by an experimenter to a subject, thus guiding the subject's perception of the stimulus person. Similarly, expectancy effects were found when experimenters were led to expect either more or less hallucinatory experience from their subjects when interviewed (Raffetto, 1968); and when experimenters were led to expect either high or low levels of aggression from their subjects in a Milgram (1963) type of experiment (Timaeus & Lueck, 1968a). Another study found that the degree of certainty of having to take a test as a function of the degree of preparatory effort expended was more successfully predictable from a knowledge of the experimenter's expectancy than from the actual degree of effort expended by the subject in preparation for the test (Cooper, Eisenberg, Robert, & Dohrenwend, 1967). One study, employing the Taylor Manifest Anxiety Scale under conditions of group administration, failed to yield significant expectancy effects (Pflugrath, 1962)

but, taken as a whole, the research suggests that the laboratory interview is somewhat susceptible to the influence of the experimenter.

The generality of EEE has been extended to reaction time studies as well. Employing visual stimuli, Wessler (1968) found that the experimenter's expectancy influences the subject's reaction time performance. The experimenter's expectancy also influenced subject's reaction times in a word association test (Silverman, 1968). Although McFall (1965) found no EEE with the person perception task, using reaction time as the dependent measure, reaction time studies in general appear to be somewhat susceptible to the influence of the experimenter.

Less susceptible have been the studies in the area of human learning and ability. In a review of this research, Rosenthal (1969a) pointed out that expectancy effects have been obtained in studies of rote verbal learning and mathematical reasoning (Hurwitz & Jenkins, 1966), marble-dropping (Johnson, 1967), verbal conditioning (Kennedy, Cook, & Brewer, 1968) and intelligence testing (Larrabee & Kleinsasser, 1967). There were, however, studies in verbal conditioning (Kennedy, Edwards, & Winstead, 1968), intelligence testing (Wartenberg-Ekren, 1962; Getter, Mulry, Holland, & Walker, 1967), color-recognition (Timaeus & Lueck, 1968b) and dot-tapping (Wessler, 1969) where EEE were not obtained.

Of all research areas, tasks eliciting psychophysical judgments have been most resistant to the effects of the experimenter. Although expectancy effects were obtained with such tasks as line length

estimation (Horst, 1966) and tone length discrimination (Zoble & Lehman, 1969) none were found in studies of line length estimation (Wessler, 1969), estimation of the frequency of light flashes (Müller & Timaeus, 1967) and numerosity estimation (Adair, 1968; Shames & Adair, 1967; Weiss, 1967).

The experimenter expectancy effect and its relationship to the experimental task. The epistemic yield in respect of the development of a theoretical schema for EEE might be increased by an analysis of those studies which consistently failed to obtain significant expectancy effects. Common factors and properties can be derived from those tasks that have been uniformly most resistant to EEE when considered in relation to those experimental tasks which readily yield significant expectancy effects. This strategem readily permits the isolation of properties underlying those conditions which impede EEE, thereby extending the theoretical framework which describes the conditions prerequisite for expectancy transmission.

The person perception task has been most extensively used in EEE research with the incidence of failure to obtain significant expectancy effects at 64% of the total number of studies conducted (Rosenthal, 1969b). However, since there was a relatively substantial proportion of this extensive research which did manifest EEE and many components comprise the person perception task, it is difficult to determine -- in terms of this task per se -- those factors or single factor responsible for either the impedition or facilitation of EEE. Rosenthal (1969a) readily admits to this conclusion and points out,

additionally, that in some cases factors such as the deficiency of experimental design may have been responsible for the failure to obtain significant EEE.

An examination of other experimental tasks indicates that the experimenter's expectation failed to influence the outcome of research when the Wechsler Adult Intelligence Scale was employed (Getter, Mulry, Holland, & Walker, 1967; Wartenberg-Ekren, 1962). Similarly, the Taffel verbal conditioning task has demonstrated some resistance to the effects of the experimenter's expectancy (Kennedy, 1969; Kennedy, Edwards, & Winstead, 1968). Other tasks which have shown some degree of immunity to EEE have been the Stroop Test of Color Recognition (Timaeus & Lueck, 1968b); a tapping task and line length estimation task (Wessler, 1969); reaction time with the familiar person perception task (McFall, 1965); the Rorschach test (Strauss, 1968); a group administration of the Taylor Manifest Anxiety Scale (Pflugrath, 1962); and a person perception task eliciting I.Q. estimations of the stimulus persons (McGinley, McGinley, & Shames, 1970).

These researchers have suggested somewhat different interpretations for their failure to find significant EEE. For example, Wartenberg-Ekren reasoned that the Block Design subtest of the W.A.I.S. requires more attention and absorption on the part of the subject than does the person perception task, resulting in a lessening of attention paid to experimenter behavior during the conduct of the research. Since the experimenter is not attended to, the biasing effect of his expectancy cannot be readily communicated to the subject.

Wessler has also proposed several interpretations of his failure to find significant EEE, each apparently tied to the particular experimental task employed. Resulting from his failure to find EEE in a task of reaction time to paired visual stimuli, he concluded that:

> "...in order for bias to occur, the task should be ambiguous with regard to the criterion response. Faced with such ambiguity, the subject is dependent upon the experimenter for information about what response will be considered most acceptable" (Wessler, 1966, p. 2173B)

In yet another instance where hand tapping and line length estimation tasks were employed (Wessler, 1969), he reasoned that "the more obvious the correct response is to S (subject), the less susceptible S's performance is to E (experimenter) expectancy effects, probably because S need not seek information about how his performance will be evaluated by E" (p. 66).

The majority of tasks reviewed thus far have either a mixed history where EEE has and has not been found or the task has only been used once in an expectancy paradigm. Numerosity estimation tasks, on the other hand, have uniformly failed to evince significant EEE in a sample of studies (Adair, 1968; Müller & Timaeus, 1967; Shames & Adair, 1967; Weiss, 1967). This task in particular and indeed the area of psychophysical judgment, in general, appear to be somewhat more resistant to the effects of the experimenter's expectancy. In reviewing this research, Rosenthal (1969a) pointed out that:

> "On the whole, the area of psychophysical judgment, particularly when the judgment is of numerosity, seems less susceptible to the effects of the experimenter expectancy than the areas considered so far" (p. 214).

This task, then, provides a good point of departure for an analysis of those properties which underlie those conditions which impede EEE.

The Ambiguity Model

Based on research involving the psychophysical judgment of numerosity, Shames and Adair (1967) suggested that the failure to obtain EEE was related to the experimental task employed. It was postulated that a "fact-centered" task like numerosity estimation requires the subject to map a purely numerical, estimative judgment onto a numerical relational system i.e., a numbered rating scale; while an "attitude-oriented" task like the person perception task elicits a more complex response where the subject must map an abstract judgment of "success-failure" onto a concrete numbered rating scale. Interpreted in terms of ambiguity, this suggests that the more ambiguous person perception task, which requires multidimensional judgment, imposes upon the subject a need to seek out cues external to the stimulus material. In contrast, the simpler, unidimensional judgment elicited by the numerosity estimation task does not impose such a need. In addition, it was suggested that the structure of the person perception task was more amenable than the numerosity estimation task to the transmission of "attitudinal" cues, such as facial and gestural cues, paralinguistically and kinesically.

Some support for the ambiguity model in EEE has also been found in other studies. Although significant expectancy effects were not obtained, a study by Weiss (1967), where stimulus ambiguity was varied by variation in the tachistoscopic exposure of stimuli in a numerosity

estimation task, yielded a trend suggesting that the magnitude of EEE increased with increasing stimulus ambiguity, i.e., faster exposure times. Stronger support for this model was found in a study where significant EEE were obtained (Nozick, 1968). The person perception task was presented tachistoscopically with the attendant finding of EEE only under the most ambiguous condition, i.e., with the most rapid exposure time (.1 second). There was, however, no increasing function relating ambiguity to magnitude of EEE -- only EEE under the most ambiguous condition.

There is, then, support for the ambiguity model in EEE. For a broad class of tasks encompassing psychophysical and attitudinal judgments, a category of laboratory behavior which Riecken (1962) has termed the "task-ability situation", it is more molecular and useful an approach to conceive of ambiguity in terms of the judgment enjoined on the subject by the task he is required to perform. Thus, a task which evokes a multidimensional judgment from the subject, it is suggested, is more complex and, hence, more ambiguous than a task which elicits a unidimensional judgmental response. There is some support for this position in a conjecture by Nunnally (1967) who speculates that:

> "In unidimensional scaling, the tasks are intuitively clear to investigators, and subjects apparently have no difficulty in understanding them, e.g., instructions to rate the handsomeness of men on a seven-step scale or to indicate which of two weights is heavier. In contrast, in multidimensional scaling, the tasks are not so intuitively clear to investigators, and it has been this author's experience that subjects are sometimes a bit confused by them.....Because of the complexities of the tasks required of subjects in some forms of multidimensional scaling, one

> wonders if the subject can do what is requested and whether or not the results make any sense" (pp. 422-423).

The first method of multidimensional scaling has been of relatively recent origin (Torgerson, 1951). Since then, there has been a sustained interest in multidimensional scaling models and related procedures. Torgerson (1958), for instance, has classified multidimensional scaling models into two main categories, spatial models and distance models, with the added suggestion that:

> "The models for scaling psychophysical distance (multidimensional scaling) are for the most part directly analogous to corresponding unidimensional scaling methods.....In each case, however, a somewhat more complex judgment is required of the subject" (p. 260).

Ekman and Sjöberg (1965) have categorized multidimensional methods into distance models and the somewhat radical content models, which deal with data as scalar products and not as distances. This work suggests the relative importance that is attached to the conceptualization of judgmental dimensionality.

As Torgerson (1958) has suggested, the primary test for unidimensionality in the judgment methods is the transitivity requirement on the ordering relation. He states, further, that transitivity is not merely a matter of stipulation but can be subjected to empirical test. Thus, when dealing with dimensionality of judgment, as in the case of elaborating the ambiguity model in EEE, there is the advantage of being able to test for dimensionality in terms of transitivity requirements rather than postulating it.

The literature concerning dimensionality of judgment is

particularly applicable to numerosity estimation tasks and psychophysical judgment in general -- that area of research which has most uniformly failed to obtain significant EEE. The focus of this study, then, is to investigate the relationship between EEE and the ambiguity model defined in terms of judgmental dimensionality. Ambiguity considered in this manner ultimately resides in the judgment-to-be-made, not necessarily in the physical stimulus itself. When conditions of ambiguity exist, it is suggested that the subject seeks out cues external to the experimental task itself because the information input is too complex to be processed in a confident, unequivocal manner. Under such conditions, the experimenter's expectancy can effectively, albeit unintentionally, be transmitted.

Rationale and Statement of the Problem

Rosenthal (1969a) has persuasively demonstrated, by means of a prodigious body of research, the biasing effect of the experimenter's expectancy. To this end, many tasks within diverse research areas have been employed although the person perception task has been used most consistently. This body of research has empirically demonstrated the generality of EEE and has been of heuristic value for the examination of mediating variables. Although there is some sound evidence favoring the generality of EEE, this position has not been uncritically accepted (Barber & Silver, 1968a).

In examining EEE research, it was found that research with psychophysical judgment, more than any other research area, has evinced some degree of immunity to the influence of the experimenter's

expectancy (Adair, 1968; Miller & Timaeus, 1967; Shames & Adair, 1967; Weiss, 1967; Wessler, 1969). Based upon these studies and one in which EEE were found only with the most rapid tachistoscopic exposure of the person perception task (Nozick, 1968), it was inferred that an ambiguity model may generally explain the influence of the experimenter's expectancy for most task-ability situations (Riecken, 1962). This conceptualization of ambiguity, however, was treated in general terms in that it was assumed that the more ambiguous the task, the more likely it was that the subject would find need to attend to the experimenter for unintentionally transmitted directive cues.

In the present study, the ambiguity model is elaborated in somewhat more precise terms. Based on Torgerson's (1958) and Nunnally's (1967) speculation of complexity in multidimensional judgment, it is argued that a task which elicits a complex, multidimensional judgment from the subject is ambiguous in relation to one which evokes a simple, unidimensional judgment. Therefore, a task eliciting a multidimensional judgment should prove less immune to the biasing effect of the experimenter's expectancy than one eliciting a unidimensionality judgment, largely because there is a greater likelihood of subjects seeking out cues external to the stimulus material under this more complex condition.

To test this conceptualization, two phenomenologically equivalent tasks were constructed by means of pilot research. These tasks, however, were dimensionally distinct physically since one consisted of 10 diverse, chromatic stimuli and the other consisted of 10 graduated,

achromatic stimuli. Both sets of stimuli were to be rated on a 20-point, algebraic rating scale ranging from EXTREMELY COLOUR POOR (-10) to EXTREMELY COLOUR RICH (+10) in step-intervals of one. Chromatic stimuli can be sensorily discriminated on at least three specifiable dimensions, hue, value and chroma[3] and, in fact, a set of chromatic stimuli have been multidimensionally scaled (Torgerson, 1951). The achromatic stimuli, on the other hand, vary only in one physical dimension, value, since they are physically devoid of chroma and hue.

To empirically ensure that this physical-dimensional difference between sets of stimuli enjoins on the subject a corresponding judgmental distinction, pilot research was conducted wherein two independent groups of subjects were required to scale both sets of stimuli -- one set of stimuli per group -- by the method of paired-comparisons. An intrasubject index of intransitivity was obtained by the method of circular triads analysis (Kendall, 1955, p. 148). Since this method yields only an index of consistency of judgment for each subject individually, multidimensionality of judgment across subjects could conceivably be masked by different subjects using different judgmental dimensions; but each using a single dimension across the series of stimuli and using it, moreover, consistently. It was precisely for this reason that another group of subjects were required to rank-order both sets of stimuli and complete a questionnaire which was germane to

---

[3]The terms value and chroma correspond to the more familiar terms brightness and saturation, respectively.

the dimensionality of these stimuli. From these data, an intersubject index of intransitivity was obtained, by means of Coombs' (1960) unfolding analysis and Kendall's Coefficient of Concordance, W. In this way it was possible to exhaustively establish, on an empirical basis, the dimensional disparity between both sets of stimuli.

With the establishment of the dimensional disparity between these two sets of stimuli, they were then employed in a classical expectancy context in Study I. It was hypothesized that:

Hypothesis I: Significant expectancy effects will be found for the chromatic set of stimuli while none will obtain under achromatic stimulus conditions.

A second study also tested the relationship between EEE and dimensionality. However, this study was extended to describing a function relating the magnitude of EEE to graduated levels of dimensionality, unlike Study I which dealt with dimensionality on a purely all-or-nothing, dichotomous basis. This graduation of levels of dimensionality is somewhat analogous to the technique of varied tachistoscopic exposure. In addition, this study attempted to elucidate the conceptualization of dimensionality, as it is employed herein, by localizing dimensionality in the subject's judgment -- not strictly in the stimulus-object itself.

The traditional person perception task was employed in Study II for two reasons: (1) It has proven to be relatively successful, over sustained use, in the demonstration of EEE, and (2) because it is a physically multidimensional task, the dimensionality of judgment related

to this task can be manipulated readily as to create unidimensional and multidimensional judgmental requirements instructionally. Thus, dimensions of "success-failure" were predetermined by means of pilot research. By this means, the first, second, third, fourth, and fifth most common dimensions of success-failure were determined from subjects' responses to a questionnaire after having rated the photos. In Study II, then, the traditional expectancy paradigm was employed with three levels of dimensionality. The first level of dimensionality took the form of instructionally specifying a one-dimensional criterion to be employed by subjects in making ratings of success-failure. The second level of dimensionality entailed a criterion of three dimensions and the third level specified five dimensions of success-failure to be used conjunctively. It was predicted that:

Hypothesis II: No significant EEE will obtain under the unidimensional condition of the person perception task, but will emerge under both conditions of multidimensionality. Thus, no significant difference is expected only between positive and negative levels of expectancy under unidimensional conditions.

Hypothesis III: The magnitude of expectancy effects will generally increase with an increment in levels of dimensionality although a strictly linear relationship is not anticipated.

CHAPTER III

CHROMATIC-ACHROMATIC STIMULI AND THE EXPERIMENTER EXPECTANCY EFFECT

In order to provide a test for the ambiguity model in EEE, it was requisite that apposite experimental tasks be constructed. This entailed the empirical establishment of dimensional disparity between two sets of stimuli. To this end a series of pilot researches were conducted for the purpose of selecting 10 chromatic stimuli which were judgmentally matched with 10 achromatic stimuli. These two experimental tasks were, thus, matched save for their differential variation in terms of dimensionality, the achromatic stimuli being unidimensional and the chromatic stimuli multidimensional, as was empirically determined by means of a series of pilot researches.

Pilot Research

Prior to the empirical establishment of dimensional disparity it was necessary, first, to construct tasks which were otherwise judgmentally matched. To this end the principal investigator administered 67 stimuli in all, 10 graduated achromatic stimuli taken from a Kodak Gray Scale and 57 diverse, chromatic stimuli taken from Canadian Industries Limited (C.I.L.) color samples, to a total number of 96 females from the Introductory Psychology course at the University of Manitoba. Female subjects (Ss) were employed because it was intended that females would be used throughout the entire research for the purpose of maximising EEE (Rosenthal, Persinger, Mulry, Vikan-Kline, & Grothe, 1964b).

In the first condition 48 Ss were required to rate all 67 stimuli on a 20-point, algebraic rating scale ranging from EXTREMELY COLOUR POOR (-10) to EXTREMELY COLOUR RICH (+10). As shown in Appendix A, this scale ranged in step-intervals of one. The stimuli were presented in eight counterbalanced orders in order to compensate for both chromatic-achromatic order effects and serial order effects within both sets of stimuli. By designating the chromatic stimuli CR and the achromatic stimuli ACR, the counterbalanced presentation orders can be represented as: (1) ACR 1-10 : CR 1-57, (2) ACR 1-10 : CR 57 -1, (3) CR 1-57 : ACR 1-10, (4) CR 57-1 : ACR 10-1, (5) ACR 1-10 : CR 57-1, (6) ACR 10-1 : CR 1-57, (7) CR 1-57 : ACR 10-1, and (8) CR 57-1 : ACR 1-10. Six Ss were randomly assigned to each presentation order.

These same counterbalanced orders were employed in a second condition. Twenty-four Ss, females from the Introductory Psychology course at the University of Manitoba, were required to rate the same 67 stimuli as in the first condition on a 20-point, algebraic rating scale ranging from EXTREMELY LOW COLOUR VALUE (-10) to EXTREMELY HIGH COLOUR VALUE (+10). As illustrated in Appendix A, this rating scale ranged in step-intervals of one.

In a third condition there were 24 Ss, females from the Introductory Psychology course at the University of Manitoba. This was a replicate of the second condition save for the definition of colour value in terms of lightness or brightness, an approach which was designed to modulate S's judgments of the stimuli (Appendix B). It was, thus, assumed that the criterion introduced in this condition would attenuate

the multidemensionality of judgment which otherwise inheres in judging the colour value of visual stimuli. It was also assumed that the variability of judgment would thus be restricted.

Means and standard deviations were determined for all of the stimuli in the three conditions -- a total of 201 distributions. From a best approximate match of means and standard deviations, 10 chromatic stimuli were drawn from the pool of 57 chromatic stimuli and matched with the 10 achromatic stimuli. In addition, it was stipulated that only the one rating scale which yielded the widest range of judgments of the three scales would be selected for subsequent use in Study I. This was designed to meet the requirement of intrastimulus disparity which controls for the difficulty posed by the closeness-of-items factor on the judgmental continuum when Ss make paired-comparisons of the stimuli in a later stage of the pilot research. This then meets one of the provisos of the method of circular triads which was used to analyze the paired-comparisons data. The rating scale employed in the first condition, "color richness", was selected for use in Study I because it best met these criteria. Thus, 10 chromatic and 10 achromatic stimuli, originally phenomenologically matched but different with respect to the assumed physical dimensions of chromaticity, were selected for later use under expectancy conditions.

In order to empirically establish the otherwise assumed, physical-dimensional disparity between the two sets of stimuli, two subsequent stages of pilot research were conducted. The first of these, a paired-comparisons assessment of both sets of stimuli, was conducted for the

secondary purpose of determining intrasubject intransitivity. The second stage of pilot research, a rank-ordering session for both sets of stimuli, was conducted for the primary purpose of obtaining a measure of intersubject intransitivity.

Paired-Comparisons Pilot Research

Method. One hundred females from the Introductory Psychology course at the University of Manitoba served as Ss in the paired-comparisons stage of the pilot study. Fifty Ss were randomly assigned to the chromatic stimulus group and another 50 to the achromatic stimulus condition. All Ss were read the following instructions:

> "I am going to read you some instructions. I am not permitted to say anything which is not in the instructions nor can I answer any questions about this experiment. OK?
>
> "We are in the process of developing a battery of human engineering tests, one of which entails the judgment of colour. This test is designed largely to test your ability to discriminate colours.
>
> "I will show you a series of pairs of stimulus cards. For each pair I want you to judge which stimulus you perceive as richer in colour than the other. The simulus cards may appear more than once in the complete presentation series and, therefore, I would like you to judge the stimulus cards as accurately as you can. Furthermore, you will later be assigned to repeat this same series of judgments so, again, the stress is on accuracy. Now, would you simply indicate the richer colour stimulus of the pair by pointing to it with your finger.
>
> "All ready? Here is the first pair of stimulus cards." (No further instructions may be given although all or part of the instructions may be repeated)."

That part of the instructions that indicates that S will be required to repeat this same task at a later date was introduced for the purpose of assuring a higher level of attention than otherwise from S during her performance of the task. This was necessary for a more valid test of

of the data by means of circular triads analysis.

A counterbalanced presentation of stimulus pairs for each of both groups of Ss was employed. For one-half of each group of Ss, the stimuli were presented in the order indicated in Appendix C. The other half of both groups of Ss had this order of presentation reversed for the purpose of effecting counterbalance.

Results. One measure of dimensional disparity between these two tasks, a measure of intrasubject intransitivity, was obtained by means of circular triads analysis of paired-comparisons data collected in a main phase of the pilot research. The results are somewhat supportive of the assertion of dimensional disparity since the mean ζ-indices calculated over 50 Ss in both sets of stimulus conditions were, as indicated in Table 1, .9572 for the achromatic stimuli and .9065 for the chromatic stimuli. Thus, a higher level of consistency was obtained in the judgment of achromatic stimuli. Since the closeness of the stimulus items on the judgmental continuum was somewhat controlled by selecting a fairly wide range of spaced stimuli and since the attention span of Ss was controlled by means of a deception, the differences between the sets of stimuli, as characterized by the coefficients of consistence, are attributable to differences in dimensionality. Figure 1 describes this difference graphically.

A χ^2 test for two independent samples (Siegel, 1956) was performed to test the significance of the difference between the data collected in both stimulus conditions. In order to apply this test the numerical properties of the ζ-indices were not retained. Instead these data were

TABLE 1

ζ-INDICES* CALCULATED FOR THE CHROMATIC AND ACHROMATIC SETS OF STIMULI BY MEANS OF THE METHOD OF CIRCULAR TRIADS

Achromatic Stimuli		Chromatic Stimuli	
.925	1.000	1.000	.875
1.000	1.000	.925	.825
.850	.975	1.000	.925
1.000	1.000	1.000	.850
.575	.925	1.000	.775
1.000	1.000	.900	.975
1.000	.725	.900	1.000
.975	1.000	.6875	.950
1.000	1.000	.4375	1.000
.600	1.000	1.000	1.000
1.000	.9375	.950	.825
1.000	1.000	.925	.875
1.000	.975	.825	.800
1.000	1.000	.950	.850
1.000	1.000	.950	1.000
.825	1.000	.825	.950
1.000	1.000	1.000	.800
1.000	1.000	1.000	1.000
1.000	.975	.900	.975
.950	.950	.925	.825
.9125	.975	1.000	1.000
1.000	.850	.850	.875
1.000	.875	.975	.975
1.000	1.000	.900	.925
.9125	1.000	.825	.825
N = 50		N = 50	
Mean ζ-Index = .9572		Mean ζ-Index = .9065	

*This index is also referred to as zeta, ζ, Kendall's (1955) Coefficient of Consistence in paired comparisons.

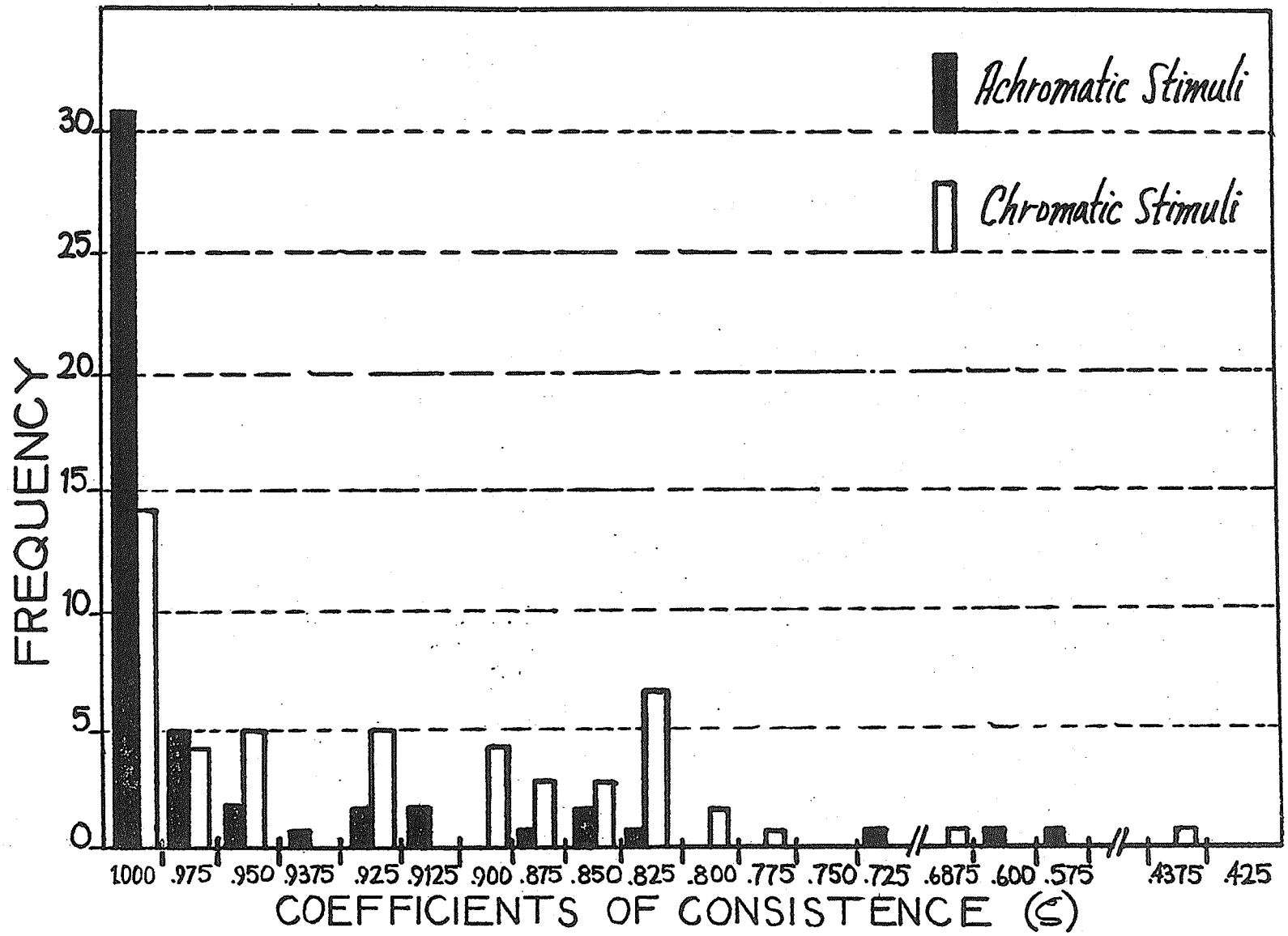

Fig. 1. A comparison between Coefficients of Consistence (C) for the achromatic and chromatic sets of stimuli.

transformed into a dichotomous distribution based on ξ-indices of 1.000 and "non-1.000" ξ-indices. This tactic is justified in that the only rigorous statement that can be made about the coefficient of consistence is that $\xi = 1.000$ implies unidimensionality, given that the factors, carelessness-of-judgment and closeness-of-items have been controlled. Any deviation from 1.000 suggests a corresponding deviation from unidimensionality and, hence, multidimensionality. Thus, a comparison of the chromatic and achromatic stimulus sets in terms of unidimensionality and non-unidimensionality yielded a X^2 of 12.84, which was significant at the .001 level with 1 degree of freedom. These data taken as a whole, then, imply a difference in dimensionality between the achromatic and chromatic sets of stimuli, in terms of intrasubject judgment.

More important, however, is the analysis of dimensional disparity between these two sets of stimuli in terms of intersubject judgment. The index of consistence, ξ, imposes a limiting value on the determination of dimensionality since it only deals with transitivity on a within-subject level. Thus, although perfect unidimensionality might be indicated by ξ averaged over subjects, a stimulus series could be judged multidimensionally across subjects. In order to investigate and compensate for this possibility, data were collected by means of rank-ordering the chromatic and achromatic sets of stimuli, analyzing these by means of Coombs' (1964) unidimensional unfolding analysis. This, it was determined, would yield an intersubject index of dimensionality.

Rank-Order Pilot Research

Method. Fifty females from the Introductory Psychology course

pool at the University of Manitoba served as Ss in the rank-order phase of the pilot studies. The experimental procedure was similar to that employed during the paired-comparisons phase with the exception that each and all Ss were required to rank-order both sets of stimuli. Twenty-five Ss were presented with the achromatic stimuli first and another 25 Ss were required to rank-order the chromatic stimuli first for the purpose of effecting a counterbalance. The instructions used in this study were designed to accommodate the rank-ordering task (Appendix D). In addition, Ss were required to complete a questionnaire which indirectly assessed their perception of the dimensionality of both sets of stimuli (Appendix D). As in the preceding stages of the pilot research, the scaling criterion for rank-ordering the stimuli was "colour-richness".

Results. The rank-orders of both sets of stimuli, shown in Appendix E, suggests that the achromatic stimulus set more adequately fulfills Coombs' (1964) conditions for unidimensionality than does the chromatic set of stimuli. In order to satisfy conditions of unidimensionality, only 45 different I scales are permissible with sets of 10 stimulus items, an I scale being the individual's preference-ordering of the stimuli. With 50 Ss, 49 different I scales were obtained for the chromatic stimulus set, indicating a violation of the condition for unidimensionality. On the other hand, only four different I scales were obtained for the achromatic set of stimuli -- 46 of the 50 I scales being exact replicates of each other. The achromatic stimuli, then, fulfilled this condition for unidimensionality. In addition, Coombs' model

demands that only two stimuli from the set can be terminal and initial stimuli throughout all 50 I scales in order to satisfy conditions of unidimensionality. This condition was patently violated by the data collected in the chromatic stimulus condition since eight of the 10 stimuli in the set were used as initial and terminal stimuli throughout the 50 I scales. In the achromatic stimulus condition, however, three terminal and initial stimuli were used throughout the 50 I scales, stimulus 8 being employed as the initial stimulus consistently over the 50 rank-orders and stimulus 5 employed in 46 of the 50 rank-orders. Again, then, conditions for unidimensionality are generally more adequately satisfied by the achromatic set of stimuli than by the chromatic stimulus set.

Additional support for this assertion of intersubject judgmental unidimensionality in the achromatic stimulus set and multidimensionality in the chromatic set of stimuli was obtained by the application of Kendall's (1955, p. 98) Coefficient of Concordance, W, to the rank-orders of both sets of stimuli. The obtained W for the achromatic set of stimuli was .932 with the average value of r_s over all possible pairs of rank-orders, .930. This extremely high rate of intersubject consistency of judgment stands in direct contrast to the W = .237, and the average r_s of .221, obtained for the rank-orders of the chromatic stimuli.

Yet another source of support for dimensional disparity between the two sets of stimuli was obtained by means of questionnaire data (Appendix D). In response to item 5, which asked S which of the two sets of stimuli she found more difficult to order, 88% found the

chromatic stimuli more difficult to rank-order than the achromatic stimuli. Item 8 asked Ss if they felt that they could repeat their rank-orders precisely, one month later. For the achromatic set of stimuli, 92% of Ss felt confident that they could; while under chromatic stimulus conditions, only 14% of Ss felt that they could precisely repeat their rank-order a month later.

All of these analyses, taken together, suggest that the achromatic stimuli elicit unidimensional judgments from Ss in correspondence with the physical unidimensionality of the task itself. The chromatic stimuli, on the other hand, did not elicit unidimensional judgments from Ss both in terms of intrasubject and intersubject judgment. There is, then, an empirical basis for distinction between these two tasks. Thus, these pilot researches served to construct two tasks which would afford a test of the ambiguity model in EEE in terms of dimensionality.

STUDY I

Method

Subjects (Ss). One-hundred females from the Introductory Psychology course subject pool at the University of Manitoba served as Ss. It was stipulated that any S with a known color defect in their vision, as measured by the A/O H-R-R Pseudoisochromatic Plates (1957), could not participate in the study.

Experimenters (Es). Twenty male volunteers from advanced undergraduate Psychology courses at the University of Manitoba served as Es. An attempt was made to use only those volunteers who were not familiar with EEE research. For the purpose of data analysis, it was stipulated

that each E run at least three Ss per cell.

Materials. Ten stimuli, ranging in value from black through to white, were appropriated from a Kodak Gray Scale (35.6 cm) for use as the achromatic set of stimuli. These were ten 1 1/2 x 1 1/4 inch cardboard tones of gray which were mounted on slide-like 3 x 5 inch white cards for the purpose of manual presentation to Ss. A chromatic set of stimuli, consisting of ten 2 1/2 x 1 3/4 inch coloured arborite chips taken from a C.I.L. colour slide and mounted on 3 x 5 inch white cards, was also constructed for use in this study.

Design. There was a random assignment of Es to each of the four groups in this study, with five Es per cell each running five Ss. There was an approximate random assignment of Ss to treatment groups inasmuch as E's 45-minute experimental time period was partitioned into five equal time segments and Ss could sign for whichever time interval they preferred. In toto 20 Es ran a total number of 100 Ss in both the achromatic and chromatic stimulus conditions across positive and negative levels of expectancy. This study, then, consisted of a fixed-effects 2 x 2 factorial design (Hays, 1963), with two levels of expectancy in each of two stimulus conditions.

Procedure. All Es were greeted by the principal investigator and were read the following set of instructions which were designed for the twofold purpose of describing E's role in the study and, principally, the induction of either a positive (+5) or negative (-5) expectancy in E:

"You have been asked to participate in a research project which is

in the process of developing a battery of human engineering tests. Previous studies have shown that colour can be reliably related to people's affect and, thus, the aim of the test we are designing is to finally measure this relationship. An additional reason for your participation in this project is to give you practice in duplicating experimental results. In Physics labs, for example, you are asked to repeat experiments to see if your findings agree with those already well-established. However, there is also the problem in psychological research of different examiners getting somewhat different data on the same tests as a function of individual differences in both subjects and experimenters. Therefore, to standardize the tests, it is better methodological procedure to use groups of experimenters.

"You will be asked to run a series of subjects and obtain from each a series of ratings of stimulus cards. After you have run each subject, you are to send her down the corridor to me where she will take the Ishihara Color Blindness Test. This is merely a precaution to ensure that all subjects possess normal color vision. The experimental procedure has been typed out for you and is self-explanatory. DO NOT DISCUSS THIS PROJECT WITH ANYONE UNTIL your instructor tells you that you may.

"According to preceding research in this area, the type of subjects you will be using have averaged a (+5) (-5) rating. Therefore, you can expect the subjects you will be running to average approximately a (+5) (-5) rating as well.

"Just read the instructions to the subject. Say nothing else to them except hello and goodbye. If for any reason you should say anything to a subject other than that which is written in your instructions, please write down the exact words you used and the situation which forced you to use them.

"Your subjects will be coming in at about _____and there is a note outside asking them to wait in Room _____ until you get them. When you have finished your task with them, send them down the corridor to me in Room _____."

"GOOD LUCK! "

These instructions were used with all four groups of Es in Study I, the only manipulation being the induction of either a positive (+5) or negative (-5) expectancy across stimulus conditions. Thus, the task was structured for E as both a laboratory exercise to see if he could

replicate "well-established" findings and, primarily, as his contribution to the construction of a human engineering test.

The rating scale employed in this study was the same as that employed in the first condition of the pilot research (Appendix A). This scale was typed on an 5 x 8 inch white card and was placed on the table, immediately in front of S, during the conduct of the experiment. Following the principal investigator's instructions to E, the general procedure followed by each E was: (1) Greet each S in the waiting area, take her to the experimental laboratory and ask her to be seated at the appropriate side of the table. (2) Obtain the factual information from S as required by the Data Recording Sheet (Appendix F). (3) Read the following standardized instructions to S:

> "I am going to read you some instructions. I am not permitted to say anything which is not in the instructions nor can I answer any questions about this experiment. OK?
>
> "We are in the process of developing a battery of human engineering tests, one of which entails the judgment of color. This test is designed largely to assess your perception of color. I will show you a series of cards and for each one I want you to assess the "richness" or "poorness" of color.
>
> "To help you make more exact judgments you are to use this rating scale in front of you. As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the particular color to be EXTREMELY COLOR POOR while a rating of +10 means that you judge the color to be EXTREMELY COLOR RICH. A rating of -1 means that you judge the color to be MODERATELY COLOR RICH while a rating of +1 means that you assess the color as being MODERATELY COLOR RICH.
>
> "You are to rate each color as accurately as you perceive it. Just tell me the rating you assign to each color on the cards I will show you. All ready? Here is the first card. (No further explanation may be given although all or part of the instructions may be repeated).

(4) Present the appropriate set of stimulus cards to S for approximately

five seconds each. (5) record S's rating for each card on the Data Recording Sheet. (6) Send S to the principal investigator for the purpose of taking a test for color vision defectiveness and to obtain official credit for her participation in this study.

Results and Discussion

The chromatic and achromatic stimuli were employed under traditional expectancy conditions in Study I. Each E obtained ratings of the 10 stimuli, in either the chromatic or achromatic stimulus conditions from his Ss. The mean of these 10 ratings per S was computed, as shown in Appendix G, and wasused as the unit dependent measure. Thus, there were 25 observations per cell. These data were analyzed by means of a fixed-effects analysis of variance for a 2 x 2 factorial design, with planned comparisons between levels of bias for the two stimulus conditions. A summary of the analysis is shown in Table 2.

The analysis of variance yielded a significant main effect for the two different stimulus sets ($F = 11.162$, $p < .01$ with 1, 96 df) and a significant expectancy x stimuli interaction ($F = 5.971$, $p < .025$ with 1, 96 df). A planned comparison between positive (+5) and negative (-5) expectancy treatment conditions for the chromatic stimulus condition yielded an F of 14.281, which was highly significant ($p < .01$, 1.96 df). In contrast, a similar comparison for the achromatic stimulus condition failed to attain significance ($F = .457$, 1, 96 df). These results, then, were consistent with the experimental hypothesis suggested in Study I, since significant EEE was obtained under the multidimensional judgmental requirements imposed upon S by the chromatic set of stimuli, while none

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE WITH PLANNED COMPARISONS FOR THE MEAN OF Ss' RATINGS OF THE CHROMATIC AND ACHROMATIC SETS OF STIMULI OBTAINED BY Es WITH POSITIVE AND NEGATIVE EXPECTANCIES

Source	df	MS	F
Expectancy	1	23.52	2.497 N.S.
Stimuli	1	109.61	11.162**
Interaction	1	58.64	5.971*
Comparisons:			
1. Levels of expectancy for Chromatic stimuli	1	139.24	14.281**
2. Levels of expectancy for Achromatic stimuli	1	4.49	.457 N.S.
3. Remainder	1	48.04	----
Within	96	9.82	
TOTAL	99		

** Significant at .01 level.

* Significant at .025 level.

N.S. = Non-significant.

obtained under the unidimensional conditions imposed by the achromatic stimuli. There is, thus, empirical support for the ambiguity model in EEE. A graphic representation for this support, in terms of a grand mean per cell, is shown in Figure 2.

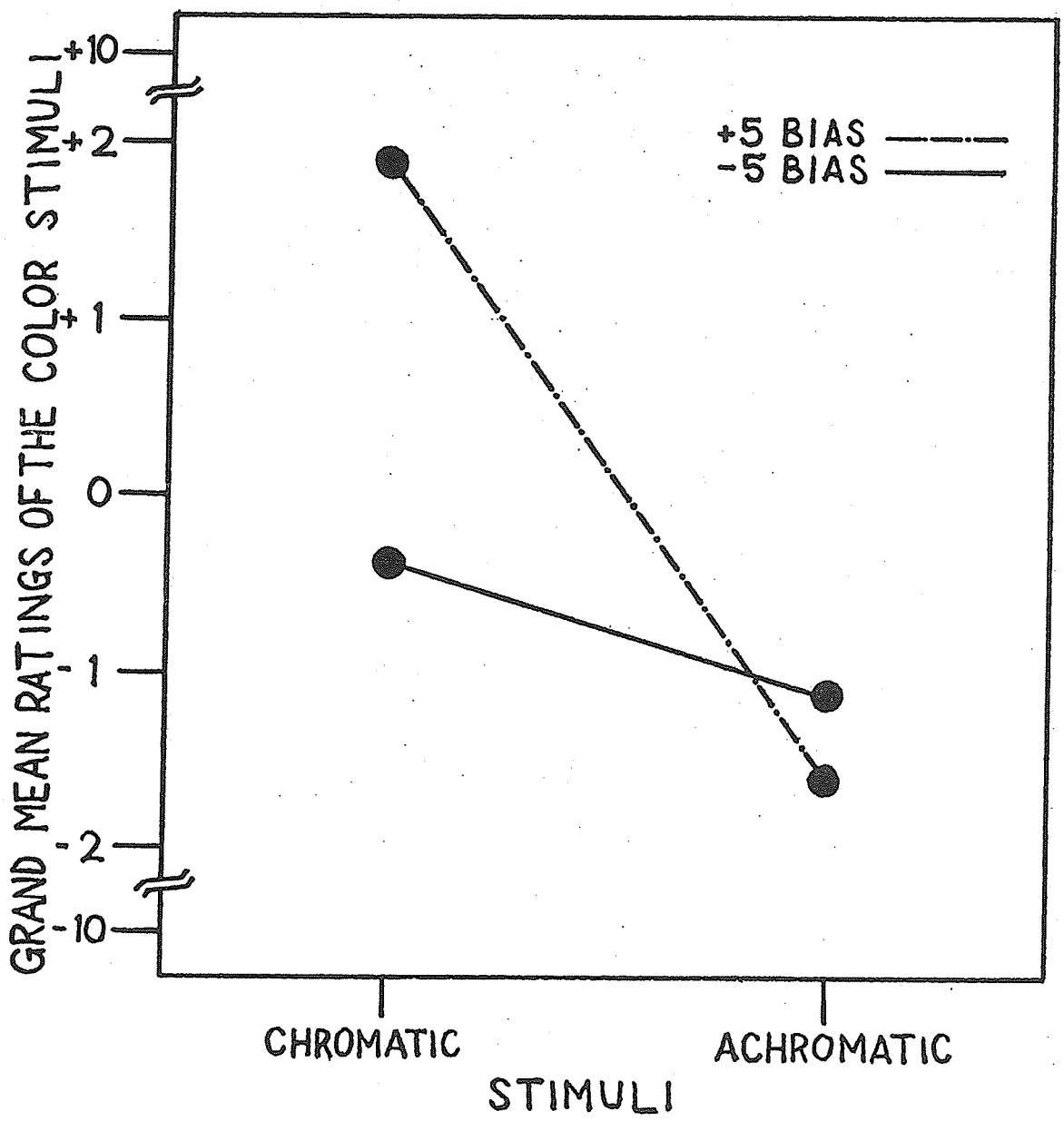

Fig. 2. Grand means calculated for each level of expectancy (bias) across achromatic and chromatic conditions, based upon subjects' mean ratings.

CHAPTER IV

AMBIGUITY IN EEE AND THE PERSON PERCEPTION TASK

Pilot Research on the Dimensionality of the Person Perception Task

Method. This research was undertaken for the purpose of constructing the experimental task which was employed in Study II. Fifty females were drawn from the Introductory Psychology course subject pool at the University of Manitoba to serve as Ss. As in all of the previous pilot research, the principal investigator administered the task. Ss were asked to rate the person perception stimuli, photographs of faces, on a continuum of "success-failure" (Rosenthal & Fode, 1963b). In fact, these stimuli were those used in a previous study where expectancy effects were found (Shames & Adair, 1967). In order to empirically establish levels of dimensionality with this task Ss were required to complete a questionnaire after rating the stimuli (Appendix H). This questionnaire was designed to seek out those dimensions most commonly employed when making photo-rating judgments of "success-failure". The relative weighting of these dimensions was determined by means of a frequency count. These dimensions were then employed to formulate the criteria which were used in the instructions to Ss in Study II. Thus, level one, the unidimensional level of Study II, furnished only one criterion statement for rating the stimuli -- the most commonly used dimension as ascertained in the present pilot research. Level two employed three criterion statements -- the first, second and third most

commonly used dimensions in the pilot study. Finally, the third level of dimensionality consisted of five such criterion statements, based on the dimensions determined by the questionnaire in this stage of the pilot research.

Results. The dimensions of "success-failure" which were employed as definitional criteria in Study II were determined by means of questionnaire data obtained during a pilot research session. The data obtained from Questionnaire 2 was used primarily for this purpose since Questionnaire 1 did not yield sufficiently molecular information regarding the dimensions of "success-failure". Thus, as shown in Figure 3, the first to the fifth most commonly used dimensions of "success-failure" were determined by a simple frequency count from the 1st to 8th ordinal ranks for the eight, predetermined criteria of success-failure. This procedure afforded the construction of a person perception task with three, empirically-determined, graduated levels of dimensionality, ranging from a unidimensional criterion, a three-dimensional criterion, to a five-dimensional criterion of success-failure. Thus, the one-dimensional criterion employed "confidence" as the criterion definition of "success-failure"; the three-dimensional condition used "confidence", "fulfillment" and "intelligence" of the stimulus-objects as definitional criteria, all of which, it was stipulated, were to be used combinationally in making a "success-failure" judgment. Finally, the criteria employed in the five-dimensional condition were the "confidence",-"fulfillment",-"intelligence",-"happiness"- and "prosperity"-appearing aspects of the stimulus-objects, the photographs of faces.

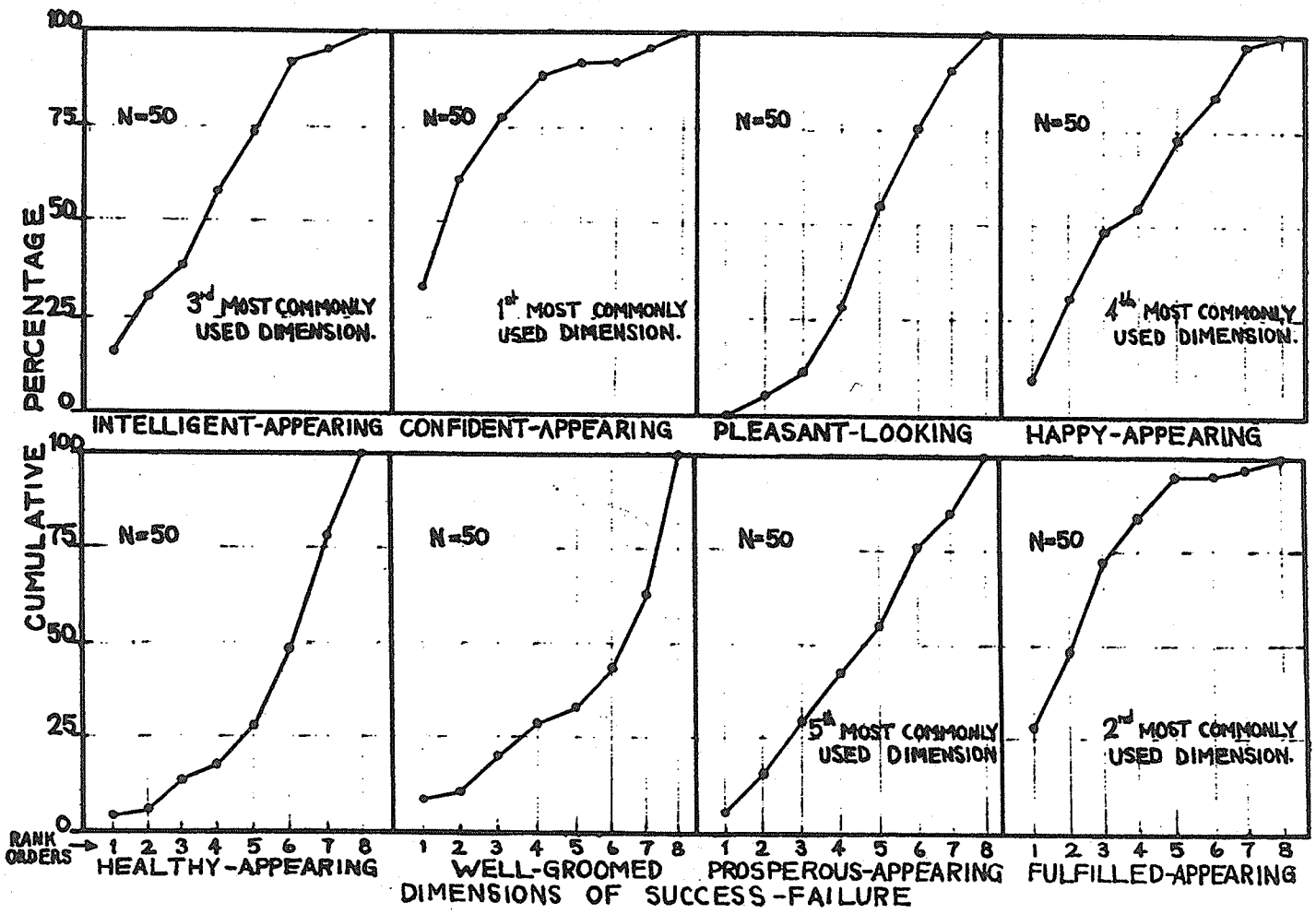

Fig. 3. Dimensions of success-failure determined by a frequency count across ordinal ranks.

STUDY II

Method

Subjects (Ss). One hundred and eighty females from the Introductory Psychology course at the University of Manitoba served as Ss. An attempt was made to limit participation only to those Ss who have not had previous experience in EEE research.

Experimenters (Es). Eighteen male volunteers from advanced undergraduate Psychology and Sociology courses at the University of Manitoba were paid at the rate of $2.00 per hour to serve as Es. The same limitation on participation as in Study I, no previous experience in EEE research, applied to the Es in this study.

Materials. The ten photographs of men's faces used in the pilot research were used as the experimental task in conjunction with the dimensionality data collected during pilot research. These stimuli were mounted on 3 x 5 inch white cards. Judgments were made by means of the Empathy Test Rating Scale (Rosenthal & Fode, 1963b). As illustrated in Appendix I, this is a 20-point, algebraic rating scale ranging from EXTREME FAILURE (-10) to EXTREME SUCCESS (+10) in step-intervals of one.

Design. The general procedure of assignment of Es and Ss to treatment groups was similar to that employed in Study I. A total of 18 Es tested 180 Ss in a 2 x 3 factorial design, fixed-effects model, with three Es per cell, each running 10 Ss. There were two levels of expectancy employed across three levels of dimensionality.

Procedure. The principal investigator read the following instructions to E for the purpose of defining E's role in the study and

for purposes of instructional induction of either a positive (+5) or negative (-5) expectancy in E:

"You have been asked to participate in a research project which is in the process of developing a test of empathy. The primary reason for your participation in this project is to standardize the results of this study. There is the problem, in psychological research, of different examiners getting somewhat different data on the same tests -- often as a function of individual differences in subjects and examiners too. Therefore, in order to standardize tests of the type we are running, we find that it is better methodological procedure to use groups of experimenters. Secondarily, your participation in this project will either familiarize you with psychological experimentation or will give you practice in duplicating experimental results to see if your findings agree with those that have already been well-established -- a procedure which is commonly used in Physics and Chemistry labs.

"You will be required to run a total of 10 subjects and obtain from each ratings of photographs. These are the photographs in question (....point to the photographs....). After you have run each subject, send her down the corridor to me where she will fill out a personal opinion inventory and I will give her official credit for her participation in this experiment. The experimental procedure you will follow has been typed out for your (....point to sheet....) and is self-explanatory.

"At the beginning of each school term, all subjects -- students in the Introductory Psychology course -- are required to fill out a number of tests, many of which are personality inventories. Based on these personality data, and previously-run research which is now well-established, it has been found that subjects of the type you will be running have averaged approximately (+5)(-5) in their ratings of the photographs. Therefore, you can expect that the subjects you will be running will also average approximately (+5) (-5) in their ratings of these photographs.

"Just read the instructions to the subjects. Say nothing else to them except hello and goodbye. If for any reason you should say anything to the subject which is not written in your instructions, please write down the precise words you used and the general situation which forced you to use them.

"Your subjects will begin coming in at about _____ and there is a note outside asking them to wait outside until you come and get them. You will be running 10 subjects in toto, one after another -- every five minutes approximately. When you have finished the experimental task with the subject, send her to me in Room _____ for

experimental credit.

"GOOD LUCK! "

All Es followed the same general procedure outlined earlier. A copy of this general experimental procedure is shown in Appendix J. Each of the three groups of Es read a different set of standard instructions to their Ss, depending upon the dimensionality condition in which they participated. Those three sets of instructions are shown in Appendix K.

Results and Discussion

As in Study I, the dependent measure in this study was the mean rating for each S of the 10 photographs of faces. Thus, as shown in Appendix L, there were 30 observations per cell, based upon the data collected by three independent Es from 10 Ss each. An analysis of variance for fixed-effects was used to analyze the data collected in this 2 x 3 factorial design. As shown in the summary of this analysis, in Table 3, significance was obtained only for the expectancy factor ($F = 4.913$, $p < .01$ with 1, 174 df). An analysis of the planned comparisons between levels of expectancy across the three levels of dimensionality yielded a significant difference between positive (+5) and negative (-5) levels of expectancy for unidimensional conditions ($F = 6.107$, $p < .025$ with 1, 174 df), a non-significant difference between levels of expectancy for the three-dimensional condition ($F = .739$ with 1, 174 df) and, again, a significant difference between levels of expectancy for the five-dimensional condition ($F = 5.595$, $p < .025$ with 1, 174 df). These data, then, generally do not support the ambiguity model as

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE WITH PLANNED COMPARISONS FOR THE MEANS OF Ss' RATINGS OF THREE DIMENSIONALLY DIFFERENT SETS OF STIMULI OBTAINED BY Es WITH POSITIVE AND NEGATIVE EXPECTANCIES

Source	df	MS	F	p
Expectancy (Rows)	1	28.20	9.827	.01
Stimuli (Cols.)	2	.13	.045	---
Interaction (R x C)	2	4.91	1.710	---
Comparisons:				
1. Levels of expectancy for unidimensional stimuli.	1	17.73	6.107	.025
2. Levels of expectancy for 3-dimensional stimuli.	1	2.12	.739	---
3. Levels of expectancy for 5-dimensional stimuli.	1	16.06	5.595	.025
Within	174	2.87		
TOTAL	179			

operationally defined in Study II. As Figure 4 illustrates, no support was obtained for Hypothesis II since significant EEE were obtained under unidimensional conditions and, additionally, none obtained under three-dimensional conditions. The data do not generally support Hypothesis III as well since there is not a consistent increment in EEE with a concomitant increment in levels of dimensionality, although there is an increment in the magnitude of EEE from the three-dimensional to the five-dimensional conditions. The single source most responsible for confusion in the

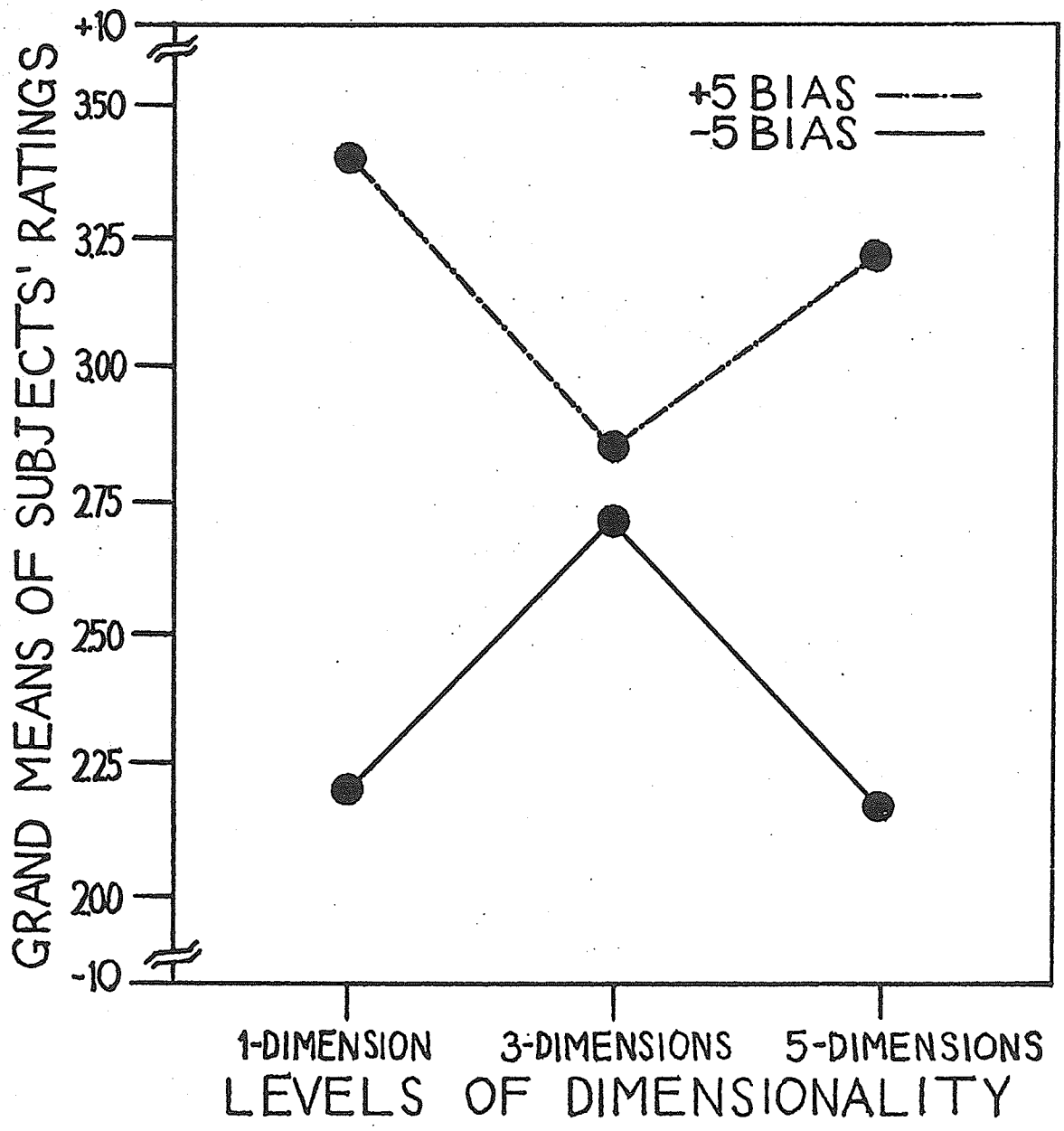

Fig. 4. Grand means calculated for each level of expectancy (bias) across each level of dimensionality, based upon subjects' mean ratings of the person perception stimuli.

data is the unidimensional condition which -- completely contrary to prediction -- evinced the greatest degree of EEE.

Related to these data was a questionnaire (Appendix M) which was administered to each S by the principal investigator upon her completion of the person perception task. The purpose of this questionnaire was twofold: to derive S's subjective impression of the experimental task by means of this written report and to determine S's awareness of the contingencies of the experiment. A summary of these responses, Table 4, points out that there is a marked deviance from the general pattern of response in the unidimensional condition as compared with the remaining two other conditions. Moreover, Table 5 indicates that the negative (-5) expectancy group was most accountable for this deviance which was characterized by inconsistency of response across items and the determination that the unidimensional condition, in terms of these responses, was the most difficult judgment task among the three levels of dimensionality. It was, of course, expected that the unidimensional condition would be considered the least difficult judgmental task among the three conditions of dimensionality.

TABLE 4

PROPORTIONS OF "YES-NO" RESPONSES TO QUESTIONNAIRE ITEMS TAKEN ACROSS LEVELS OF DIMENSIONALITY

Questionnaire Items		1-Dimension	3-Dimension	5-Dimension
1. Did you find this judgment task difficult?(Why?)	YES	78.3%	76.6%	76.6%
	NO	21.7%	23.4%	23.4%
2. Did you use the definition of success and failure provided precisely as you were instructed to do? (Why?)	YES	88.3%	71.6%	66.6%
	NO	11.7%	28.4%	33.4%
3. Did you find it difficult to translate success and failure into the terms provided by the definition given? (Why)	YES	65.0%	61.6%	50.0%
	NO	35.0%	38.4%	50.0%
4. Did you find it difficult to translate success and failure into numerical terms--as provided by the EMPATHY TEST RATING SCALE? (Why?)	YES	716%	73.3%	61.6%
	NO	28.4%	26.7%	38.4%
5. Did you find it easy to integrate all of the information provided for the purpose of making an empathy judgment? (Why?)	YES	60.0%	38.4%	40.0%
	NO	40.0%	61.6%	60.0%
6. What was the purpose of the experiment?		(Unrestricted response)		
7. Do you think you contributed to the purpose of the experiment? (If yes, how? If No, Why?)	YES	60.0%	51.6%	58.3%
	NO	40.0%	48.4%	41.7%

TABLE 5

PROPORTIONS OF "YES-NO" RESPONSES TO QUESTIONNAIRE ITEMS FOR EACH LEVEL OF EXPECTANCY ACROSS LEVELS OF DIMENSIONALITY

Questionnaire Item	Expectancy		1-Dimension	3-Dimension	5-Dimension
1. Did you find this judgment task difficult? (Why?)	+5	YES	83.3%	76.6%	73.3%
		NO	16.7%	23.4%	26.7%
	-5	YES	76.6%	76.6%	80.0%
		NO	23.4%	23.4%	20.0%
2. Did you use the definition of success and failure provided precisely as you were instructed to do? (Why?)	+5	YES	90.0%	70.0%	56.6%
		NO	10.0%	30.0%	43.4%
	-5	YES	90.0%	76.6%	76.6%
		NO	10.0%	23.4%	23.4%
3. Did you find it difficult to translate success and failure into the terms provided by the definition given? (Why?)	+5	YES	73.3%	63.3%	53.3%
		NO	26.7%	36.7%	46.7%
	-5	YES	56.6%	60.0%	46.7%
		NO	43.4%	40.0%	53.3%
4. Did you find it difficult to translate success and failure into numerical terms--as provided by the EMPATHY TEST RATING SCALE?	+5	YES	73.3%	73.3%	50.0%
		NO	26.7%	26.7%	50.0%
	-5	YES	70.0%	73.3%	73.3%
		NO	30.0%	26.7%	26.7%
5. Did you find it easy to integrate all of the information provided for the purpose of making an empathy judgment? (Why?)	+5	YES	46.7%	43.4%	36.7%
		NO	53.3%	56.6%	63.3%
	-5	YES	73.3%	33.4%	43.4%
		NO	26.7%	66.6%	56.6%
6. What was the purpose of the experiment?			(Unrestricted response)		
7. Do you think you contributed to the purpose of the experiment? (If Yes, how? If No, why?)	+5	YES	63.3%	53.3%	60.0%
		NO	36.7%	46.7%	40.0%
	-5	YES	56.6%	50.0%	60.0%
		NO	43.4%	50.0%	40.0%

CHAPTER V

DISCUSSION

The pilot research for the first study established that the physical-dimensional disparity between the chromatic and achromatic sets of stimuli enjoined upon S correspondingly disparate judgments. The principal empirical support was obtained by means of rank-ordering the chromatic and achromatic sets of stimuli. In addition, an analysis of these data by an application of Coombs' (1964) unfolding analysis and Kendall's (1955) Coefficient of Concordance yielded a substantial difference between both sets of stimuli. These provided indices for the inference of intersubject judgmental disparity which were additionally reinforced by indices of intrasubject intransitivity provided by paired-comparisons of both sets of stimuli and a concomitant analysis by means of Kendall's (1955) Coefficient of Consistence determined by circular triads analysis.

When these two sets of stimuli were employed in an expectancy format a significant difference between levels of expectancy was found for the chromatic set of stimuli while none obtained under achromatic stimulus conditions. Interpreted in terms of the ambiguity model, significant expectancy effects obtained under the more ambiguous multidimensional conditions imposed by the chromatic stimuli; but none were found under the less ambiguous, unidimensional conditions imposed by the achromatic stimuli. These results, then, supported the proposal that significant expectancy effects would be found only under the judgmental

conditions imposed upon S by the physically-multidimensional chromatic stimuli.

These data also supported the validity of the strategem suggested earlier: the use of research involving psychophysical judgment as a point of departure for an analysis of those conditions which impede and, complementarily, facilitate EEE. Generally, this area of research has been most uniformly resistant to the effects of the experimenter's influence (Adair, 1968; Müller & Timaeus, 1967; Shames & Adair, 1967; Weiss, 1967; Wessler, 1969). This provided a context for the manipulation of ambiguity in terms of the number of dimensions involved in S's judgments.

These results suggest that the dimensional complexity of the task determines whether or not S is faced with an ambiguous judgmental situation. As in the ambiguous, chromatic, stimulus condition in the present study, previous research (Nozick, 1968; Shames & Adair, 1967) has indicated that Ss will attend to extra-task stimuli for directive cues. Under these conditions, it has been shown that the experimenter's influence affects the outcome of his research. In the case where relatively little ambiguity existed in the judgmental situation, as in the achromatic stimulus condition, the experimenter presumably did not take on significance as a source of directive cues.

The ambiguity model, then, based on the results of Study I, appears to be a useful and heuristic conceptualization for research into EEE, at least for "task-ability" situations in which judgmental conditions are structured dichotomously in terms of extremes of dimensionality. Under the dichotomous extremes of unidimensionality-multidimen-

sionality, the following inference is clear: unidimensional stimulus-objects which elicit corresponding unidimensional judgments from S are more resistant to the effects of E's expectancy than are multidimensional stimulus-objects which elicit multidimensional judgments from S.[4]

Study II was also designed to examine the ambiguity model in EEE. Ambiguity in this study was conceived in terms of dimensionality manipulated by means of instructions rather than the physical attributes of the stimulus-object itself. Moreover, the model was extended from a dichotomous, all-or-nothing schema to graduated levels of ambiguity to test the prediction that there is a linear relationship between increased levels of dimensionality and an increment in the magnitude of EEE. By employing the person perception task dimensionality could be graduated by furnishing S with one-, three-, or five-dimensional criteria for their "success-failure" judgments.

Planned comparisons between the positive (+5) and negative (-5) expectancy groups for each of the three levels of dimensionality, yielded significant differences for the unidimensional and five-dimensional conditions, but none for the three-dimensional condition. These results failed to confirm the prediction that significant EEE would obtain under

---

[4] It is here emphasized that dimensionality is being considered primarily in terms of the judgment enjoined upon S. It is, then, possible that a physically multidimensional stimulus-object could elicit a unidimensional judgment from S and, conversely, a physically unidimensional stimulus-object may -- however unlikely -- elicit a multidimensional judgment from S. It is precisely for this reason that care was taken to empirically establish a dimensional difference between sets of stimuli rather than postulate this dimensional difference solely on the basis of the known physical attributes of the stimulus-objects.

all but the condition of unidimensionality. There was a qualified measure of support for the prediction of an increment in the magnitude of EEE with an increment in levels of dimensionality since the magnitude of EEE increased from the three-dimensional to the five-dimensional condition; but these data tend to be confounded by the unidimensional condition which evinced the greatest magnitude of EEE-- completely contrary to prediction.

There were differences between the experimental tasks employed in Study I and Study II which, it is suggested, bear some relationship to the outcome of these studies. It was noted earlier that the tasks employed in these studies generally fall into that category of laboratory behavior called "task-ability" situations. This classification aptly applies to a psychophysical judgment task such as the chromatic-achromatic sets of stimuli where an objective ability is summoned. Therefore, a quantitative consideration of dimensions is appropriate for the test of the ambiguity model in EEE, since an increment in the number of dimensions to be employed in making a colour-richness judgment thus taxes the upper limit of S's ability to do so consistently and, therefore, makes him more dependent upon E for subtle guidance in an otherwise ambiguous situation.

The person perception task, on the other hand, despite the delineated dimensions provided for judgment, is a more "attitudinally-oriented" task (Shames & Adair, 1967) than the chromatic-achromatic stimulus task. It is, then, a "self-quality" situation calling for an individual projective judgment rather than a task-ability situation

(Riecken, 1962). This suggests that the quantitative manipulation of dimensions may not have been wholly appropriate for the purpose of testing the ambiguity model in terms of graduated levels. Because this task does not tax S's ability, but rather calls for an attitudinal judgment, there may in fact be no upper limit, in terms of numbers of dimensions, which consistently create an ambiguous judgmental situation for S. There are perhaps other factors besides the dimensionality of judgment in an "attitudinally-oriented" task that bear upon S's judgments and presumably contaminate the possibility of testing the ambiguity model with the person perception task in wholly dimensional terms.

The postexperimental questionnaire data bear upon the outcome of Study II. Firstly, the general response indicated that it was unlikely that S was aware that she was participating in an expectancy paradigm. More crucial, perhaps, is the relationship between the remaining questionnaire items and the results of the planned comparisons analysis of the data. In this respect, the unidimensional condition is the single source most responsible for the failure to support the experimental hypotheses. Correspondingly, the pattern of response of Ss in this condition to the questionnaire items demonstrates a marked deviance from the general pattern of response in the other dimensionality conditions. Completely contrary to expectation, a greater proportion of Ss in the unidimensional condition generally found their judgment task to be more difficult than the other Ss.

Moreover, they found it more difficult to translate success-

failure into those terms provided by the unidimensional condition than did Ss in the other two, presumably more difficult, dimensionality conditions. Even more interesting was the lack of consistent responses by Ss in the unidimensional condition vis-a-vis Ss in the other dimensionality conditions. In the three-dimensional and five-dimensional conditions, a preponderance of Ss uniformly found the judgment task to be a difficult one. In the unidimensional condition, however, Ss inconsistently responded finding first that their judgment task was more difficult than easy. When they were asked essentially the same question again, phrased in reversed form, they then found that their judgment task was more easy than difficult. This was the only condition in which such inconsistency of responses was found.

This, then, suggests a possible explanation of those data collected under conditions of unidimensionality. It is suggested that using the traditional judgment criterion of "success-failure" without any further explicit definition may constitute the highest level of dimensionality, i.e., unlimited dimensionality. Therefore, in a graduated series, it may be expected that such a condition would be even more multidimensional than the five-dimensional condition. The questionnaire data supported this speculation in that Ss generally, found the unidimensional condition to be more difficult a judgmental task than the other two conditions. It is highly doubtful that Ss used the criterion "confidence" in a consistent, functionally unidimensional way over the stimulus series when making their success-failure judgments. In view of Ss' performance in this condition and the questionnaire data, it is

reasonable to speculate that this criterion itself may be multidimensional since "confidence" can, itself, be multifariously interpreted. In this sense, then, it would be equally ambiguous for Ss to judge the "confidence" of a stimulus-person as it would be his "success-failure".

In addition, it may be reasonable to speculate that EEE, in this study, may be an all-or-nothing phenomenon which cannot be veridically related to graduated levels of dimensionality. In this respect Nozick (1968) found EEE only under the condition of greatest ambiguity, not a graduated relationship. Research into the parameters of the ambiguity model, defined in terms of dimensionality, may then be a necessary next step since, as was the case in the three-dimensional condition, several dimensions may be employed as functionally unidimensional until some critical point of ambiguity is reached. It has, in fact, been suggested that Ss make only unidimensional judgments irrespective of the number of physical dimensions they are dealing with.[5] The ambiguity model would then suggest that one or more dimensions may be functionally encoded as a unidimensional judgment, but at some critical number of dimensions S is no longer able to consistently employ such a judgment over a series of stimuli. At this point S is susceptible to the influence of the experimenter's expectancy, among his other attributes, because judgmental consistency is broken down. It is, however, problematic whether there is a linear relationship between increasing order of ambiguity and magnitude of EEE. The present research provides little

---

[5] I would like to thank Dr. M. S. Aftanas for suggesting this.

answer to this question. On the basis of these yet unanswered questions, it is suggested that research employing four levels of dimensionality i.e., two-, four-, six-, and eight-dimensional conditions, be undertaken. Critical parameters for the ambiguity model in EEE may be determined from this research thus furnishing data for the ambiguity model conceived in terms of graduated levels of dimensionality.

In summary, the results of Study I demonstrated the usefulness in EEE research of the ambiguity model conceived in terms of dimensionality. The results of this study support the general inference that unidimensional stimulus-objects which elicit corresponding unidimensional judgments from S are more resistant to the effects of E's expectancy than are multidimensional stimulus-objects. This study, then, yielded general support for the ambiguity model in EEE conceived in terms of a unidimensional-multidimensional dichotomy.

Study II, however, was designed to test the ambiguity model in EEE conceived in terms of graduated levels of dimensionality. The results of this study largely failed to support this conceptualization in that it was intended that this research would permit the more powerful inference linearly relating an increment in dimensionality to increased magnitude of EEE. There is the suggestion, based upon the performance of Ss and their responses to a postexperimental questionnaire, that this failure may have been due to inaccurate operationalization rather than inaccurate conceptualization. Specifically, it was suggested that the unidimensional condition should more properly be considered an unrestricted, unlimited dimensionality condition in view

of Ss' performance in the expectancy paradigm and their responses to postexperimental questionnaire items. It was generally,concluded, on the basis of the evidence provided by the present research, taken as a whole, that the ambiguity model may be generally useful for the purpose of explicating those contingencies in the experimental paradigm which can either facilitate or obviate S's need to seek out directive cues external to the experimental task.

Needs for Further Research

In view of the present research, Study II in particular, it is suggested that somewhat modified replications of this study be undertaken. The aim of this research would be to derive more strictly unidimensional components of "success-failure" than those employed in Study II. It has, in fact, been suggested that many of the dimensions of success-failure investigated in the pilot research preparatory for Study II (Figure 3), may themselves, have been multidimensional and, thus, could account for the operationistic difficulties encountered in this study. In fact, it may be that one or more of the dimensions which were empirically determined may be of that class of dimensions which cannot be perceived accurately by Ss. This may have been especially applicable to the dimension of "confidence" in the unidimensional condition. It is therefore suggested that more purely unidimensional components of success-failure, perhaps the physiognomic features of the person perception stimulus-objects, need be determined and employed in a graduated sequence of dimensionality in an expectancy format. There may be the problem, however, of maintaining the face validity of the experimental situation

if these unidimensional components appear, as they may, somewhat unrelated to success-failure. This can be remedied either by an appropriate cover story or by using a stimulus-object and related dimensions other than the person perception task. In either case, however, a study of this type would add immeasurably to the conceptualization of an ambiguity schema in EEE research.

More research needs to be done in the investigation of the appropriate parameters of dimensionality in its relationship with the magnitude of EEE. It may be that a stimulus-object with two, or even more physical or specified dimensions is encoded by the subject as functionally unidimensional and, thus, the specification of a multidimensional condition may be inaccurate under certain conditions. This may, in fact, have been the case with the 3-dimensional condition in Study II and, so, studies investigating such parameters would contribute to an understanding of EEE and judgmental dimensionality.

In agreement with Levy (1969), it is suggested that research is needed which would develop a situational taxonomy and, therefore, broaden the theoretical base underpinning EEE. Such a taxonomy of tasks is long overdue and it is precisely this which is needed for consideration of the hotly-debated research area dealing with the generality and pervasiveness of EEE in general research. Thus, EEE research with a broad and systematic sampling of tasks and experiments would, in the long run, provide a framework for theoretical formulation apropos of EEE and its transmission; and this would provide a base against which continuing theoretical formulations could be assessed. It is, in

fact, a rudimentary form of such a situation which permitted the formulation of the ambiguity model in the present research.

Generally, then, more research is needed for the purpose of investigating the mediation of the experimenter's expectancy. The nature of that which is being transmitted in the expectancy paradigm is not yet fully understood. Investigations into the experimenter-subject communication and feedback loop, some of which have already been undertaken (Fode, 1960; Friedman, 1964; Katz, 1964; Zoble & Lehman, 1969), will thus shed more light on both the communication process and the substance of the covert, inadvertent communication. Conjunctive with this much needed research are efforts to synthesize this into a general theoretical framework such as those which have already been undertaken (Rosenthal, 1966; 1969a).

CHAPTER VI

SUMMARY

This investigation was designed to test the ambiguity model in experimenter expectancy effects. This model postulates that the subject attends to the experimenter in particular and to extra-task cues in general for direction when faced with an ambiguous judgment situation. Ambiguity results from complexity of judgment where the subject is not able to consistently encode a number of dimensions over a series of stimuli. This model was investigated, first, by providing subjects with extremes in dimensional conditions, i.e., unidimensional and multidimensional stimuli. Secondly, ambiguity was tested as a graduated phenomenon suggesting that an increased magnitude of expectancy effects was linearly related to graduated increments in dimensionality.

Through pilot research, 10 chromatic stimuli were selected on the basis of the best approximate match of their means and standard deviations with those of 10 achromatic stimuli in order to construct task stimuli which were matched in all respects save for extremes in dimensionality. This assumed dimensional disparity was empirically established by means of paired-comparisons data which yielded indices of intrasubject intransitivity and rank-orderings which yielded indices of intersubject intransitivity. Then, in a first study, these two sets of stimuli were employed in a traditional expectancy paradigm with positive and negative expectancies induced in both stimulus conditions.

The obtained results supported the ambiguity model since expectancy effects were found only under the ambiguous, multidimensional, judgment conditions imposed by the chromatic stimuli.

For purposes of the second study, it was necessary to construct an experimental task appropriate for the test of the graduated ambiguity model. Through pilot research, the most commonly employed dimensions of success-failure, in judging person perception stimuli, were empirically determined. Then, by presenting subjects with one-, three-, or five-dimensional definitions of success-failure, graduated conditions of dimensionality were established. These levels of dimensionality were instructionally presented in a traditional expectancy paradigm with positive and negative expectancies induced across all three conditions. Generally, the results did not support the ambiguity model which related an increment in dimensionality to increased magnitude of expectancy effects. However, it was suggested that this failure was less a matter of conceptualization than operationalization. Postexperimental questionnaire data indicated that subjects perceived the unidimensional definition as more difficult to use than did subjects in the other conditions. Thus, the unidimensional condition might more properly be considered the highest level of dimensionality, i.e., unrestricted dimensionality. Moreover, the quantitative manipulation of dimensions appeared to be less appropriate in this study than in Study I since the person perception task, despite the specification of dimensional criteria, does not readily fall into a task-ability category. Complexity, in terms of the number of dimensions encoded in

a single judgment, is thus less appropriate for the person perception task because of the confounding aspects of this self-quality, attitudinal type of task. In all, then, this may have yielded an inadequate test of the graduated ambiguity model.

Generally, the ambiguity model was considered heuristic in explicating those conditions which underlie the subject seeking out extra-task, directive cues when faced with an ambiguous judgment situation. This, then, facilitates the transmission of the experimenter's expectancy. More specifically, it was suggested that the ambiguity model be considered in terms of an all-or-nothing encodability of dimensions into consistent judgments over a series of stimuli since a graduated, ambiguity model might be an inappropriate conceptualization. However, further investigation into these parameters is necessary to elucidate this conjecture.

BIBLIOGRAPHY

Adair, J. G. Demand characteristics or conformity? Suspiciousness of deception and experimenter bias in conformity research. Unpublished manuscript, University of Manitoba, 1968.

American Optical Pseudoisochromatic Plates for Testing Color Perception. Hardy-Rand-Rittler, 1940. (Copyright)

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

Barber, T. X., Calverley, D. S., Forgione, A., McPeake, J. D., Chaves, J. F., and Bowen, B. Five attempts to replicate the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969, 33, 1-16.

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

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

Barnard, P. G. Interaction effects among certain experimenter and subject characteristics on a projective test. Journal of Consulting and Clinical Psychology, 1968, 32, 514-521.

Bernstein, A. S. Race and examiner as significant influences on basal skin impedance. Journal of Personality and Social Psychology, 1965, 1, 346-349.

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

Burnham, J. R. Experimenter bias and lesion labelling. Unpublished manuscript, Purdue University, 1966.

Cahalan, D., Tamulonis, V., and Verner, H. W. Interviewer bias involved in certain types of opinion survey questionnaires. International Journal of Opinion and Attitude Research, 1947, 1, 63-77.

Cahen, L. S. An experimental manipulation of the "Halo Effect": A study of teacher bias. Unpublished manuscript, Stanford University, 1965.

Cason, H., and Cason, E. B. Association tendencies and learning ability.

Journal of Experimental Psychology, 1925, 8, 167-189.

Coombs, C. H. A theory of data. Psychological Review, 1960, 67, 143-159.

Coombs, C. H. A theory of data. New York: John Wiley & Sons, 1964.

Cooper, J., Eisenberg, L., Robert, J., and Dohrenwend, B. S. The effect of experimenter expectancy and preparatory effort on belief in the probable occurrence of future events. Journal of Social Psychology, 1967, 71, 221-226.

Cordaro, L., and Ison, J. R. Observer bias in classical conditioning of the planarian. Psychological Reports, 1963, 13, 787-789.

Ehrlich, J. S., and Riesman, D. Age and authority in the interview. Public Opinion Quarterly, 1961, 25, 39-56.

Ekman, G., and Sjöberg, L. Scaling. Annual Review of Psychology, 1965, 16, 451-474.

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

Fode, K. L. The effect of experimenters' and subjects' anxiety and social desirability on experimenter outcome-bias. Unpublished doctoral dissertation, University of North Dakota, 1965.

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

Friedman, N. The psychological experiment as a social interaction. Unpublished doctoral dissertation, Harvard University, 1964.

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

Getter, H., Mulry, R. C., Holland, C., and Walker, P. Experimenter bias and the WAIS. Unpublished data, University of Connecticut, 1967.

Harvey, S. M. A preliminary investigation of the interview. British Journal of Psychology, 1938, 28, 263-287.

Hays, W. L. Statistics for psychologists. New York: Holt, 1963.

Horst, L. Research in the effect of the experimenter's expectancies--a laboratory model of social influence. Unpublished manuscript, Harvard University, 1966.

Hurwitz, S., and Jenkins, V. The effects of experimenter expectancy on performance of simple learning tasks. Unpublished manuscript, Harvard University, 1966.

Hyman, H. H., Cobb, W. J., Feldman, J. J., Hart, C. W., and Stember, C. H., Interviewing in social research. Chicago: University of Chicago Press, 1954.

Jenkins, V. The unspoken word: A study in nonverbal communication. Unpublished AB thesis, Harvard University, 1966.

Johnson, R. W. Subject performance as affected by experimenter expectancy, sex of experimenter, and verbal reinforcement. Unpublished master's thesis, University of New Brunswick, 1967.

Johnson, R. W., and Adair, J. G. The effects of systematic recording error vs. experimenter bias on latency of word association. Journal of Experimental Research in Personality, 1970, 4, 270-275.

Johnson, R. W., and Adair, J. G. E-bias vs. systematic recording error under automated and nonautomated stimulus presentation. Canadian Psychologist, 1971, 12, 190-191.

Jourard, S. M. The influence of experimenters' disclosure on subjects' behavior in psychological experiments. In C. Spielberger (Ed.), Current topics in clinical and community psychology. New York: Academic Press, 1969.

Katz, R. Body language: A study in unintentional communication. Unpublished doctoral dissertation, Harvard University, 1964.

Kendall, M. G. Rank correlation methods. New York: Hafner Publishing Co. (2nd Ed.), 1955.

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

Kennedy, J. J., Cook, P. A., and Brewer, R. R. An examination of the effects of three selected experimenter variables in verbal conditioning research. Unpublished manuscript, University of Tennessee, 1968.

Kennedy, J. J., Edwards, B. C., and Winstead, J. C. The effects of experimenter outcome expectancy in a verbal conditioning situation: A failure to detect the "Rosenthal Effect". Unpublished manuscript, University of Tennessee, 1968.

Larrabee, L. L., and Kleinsasser, L. D. The effect of experimenter bias on WISC performance. Unpublished paper, Psychological Associates, St. Louis, 1967.

Levy, L. H. Reflections on replications and the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969, 33, 15-17.

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

Lord, Edith. Experimentally induced variations in Rorschach performance. Psychological Monographs, 1950, 64, No. 10.

Luft, J. Interaction and projection. Journal of Projective Techniques. 1953, 17, 489-492.

Marcia, J. Hypothesis-making, need for social approval, and their effects on unconscious experimenter bias. Unpublished master's thesis, Ohio State University, 1961.

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

Marwit, S. J. An investigation of the communication of tester-bias by means of modelling. Unpublished doctoral dissertation, State University of New York at Buffalo, 1968.

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

Masling, J. Role-related behavior of the subject and psychologist and its effects upon psychological data. In D. L. Levine (Ed.), Nebraska Symposium on Motivation, Lincoln, Nebraska: University of Nebraska Press, 1966, 67-103.

McFall, R. M. "Unintentional communication": The effect of congruence and incongruence between subject and experimenter constructions. Unpublished doctoral dissertation, Ohio State University, 1965.

McGinley, H., McGinley, P., and Shames, M. Failure to find experimenter expectancy effects in I.Q. estimations. Psychological Reports, 1970, 27, 831-834.

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

Milgram, S. Behavioral study of obedience. Journal of Abnormal and Social Psychology, 1963, 67, 371-378.

Mosteller, F. Correcting for interviewer bias. In H. Cantril (Ed.),

Gauging public opinion. Princeton, N.J.: Princeton University Press, 1944. Pp. 286-288.

Miller, W., and Timaeus, E. Conformity behavior and experimenter bias. Unpublished manuscript, University of Cologne, 1967.

Mulry, R. C. The effect of the experimenter's perception of his own performance on subject performance in a pursuit rotor task. Unpublished master's thesis, University of North Dakota, 1962.

Nozick, D. Experimenter bias as a function of ambiguity of stimulus. Unpublished master's thesis, University of Manitoba, 1968.

Nunnaly, J. C. Psychometric theory. New York: McGraw-Hill, 1967.

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

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

Pfungst, O. Clever hans (the horse of Mr. von Osten): A contribution to experimental, animal and human psychology. (Translated by C. L. Rahn) New York: Holt, Rinehart, & Winston, 1911.

Pierce, A. H. The subconscious again. Journal of Philosophical and Psychological Scientific Method, 1908, 5, 264-271.

Prince, A. I. Relative prestige and the verbal conditioning of children. American Psychologist, 1962, 17, 378. (Abstract).

Raffetto, A. M. Experimenter effects on subjects' reported hallucinatory experiences under visual and auditory deprivation. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May, 1968.

Reece, M. M., and Whitman, R. N. Expressive movements, warmth, and verbal reinforcements. Journal of Abnormal and Social Psychology, 1962, 64, 234-236.

Rice, S. A. Contagious bias in the interview: A methodological note. American Journal of Sociology, 1929, 35, 420-423.

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

Robinson, D., and Rohde, S. Two experiments with an anti-Semitism poll. Journal of Abnormal and Social Psychology, 1946, 41, 136-144.

Rosenberg, M. J. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. Journal of Personality and Social Psychology, 1965, 1, 28-42.

Rosenthal, R. Experimenter attributes as determinants of subjects' responses. Journal of Projective Techniques and Personality Assessment, 1963, 27, 467-471.

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

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

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

Rosenthal, R. Experimenter expectancy, experimenter experience, and Pascal's Wager. Psychological Reports, 1967, 20, 619-622 (b).

Rosenthal, R. Experimenter expectancy, one tale of Pascal, and the distribution of three tails. Psychological Reports, 1967, 21, 517-520 (c).

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

Rosenthal, R. Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal and R. L. Rosnow (Eds.), Artifact in behavioral research. New York: Academic Press, 1969. Pp. 181-277 (a).

Rosenthal, R. Task variations in studies of experimenter expectancy effects. Perceptual and Motor Skills, 1969, 29, 9-10 (b).

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

Rosenthal, R., and Fode, K. L. The problem of experimenter outcome-bias. In D. P. Ray (Ed.), Series research in social psychology. Symposia study series No. 8, Washington, D.C.: National Institute of Social and Behavioral Science, 1961.

Rosenthal, R., and Fode, K. L. The effect of experimenter bias on the performance of the albino rat. Behavioral Science, 1963, 8, 183-189 (a).

Rosenthal, R., and Fode, K. L. Three experiments in experimenter bias. Psychological Reports, 1963, 12, 491-511 (b).

Rosenthal, R., Friedman, C. J., Johnson, C. A., Fode, K. L., Schill, T. R., White, C. R., and Vikan-Kline, L. L. Variables effecting experimenter bias in a group situation. Genetic Psychological Monographs, 1964, 70, 271-296.

Rosenthal, R., Friedman, N., and Kurland, D. Instruction-reading behavior of the experimenter as an unintended determinant of experimental results. Journal of Experimental Research in Personality, 1966, 1, 221-226.

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

Rosenthal, R., Kohn, P., Greenfield, P. M., and Carota, N. Experimenters' hypothesis-confirmation and mood as determinants of experimental results. Perceptual and Motor Skills, 1965, 20, 1237-1252.

Rosenthal, R., Kohn, P., Greenfield, P. M., and Carota, N. Data desirability, experimenter expectancy, and the results of psychological research. Journal of Personality and Social Psychology, 1966, 3, 20-27.

Rosenthal, R., and Lawson, R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. Journal of Psychiatric Research, 1964, 2, 61-72.

Rosenthal, R., Persinger, G. W., Mulry, R. C., Vikan-Kline, L., and Grothe, M. Changes in experimental hypotheses as determinants of experimental results. Journal of Projective Techniques and Personality Assessment, 1964, 28, 465-469 (a).

Rosenthal, R., Persinger, G. W., Mulry, R. C., Vikan-Kline, L., and Grothe, M. Emphasis on experimental procedure, sex of subject, and the biasing effects of experimenter hypotheses. Journal of Projective Techniques and Personality Assessment, 1964, 28, 470-473 (b).

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

Rosenthal, R., Persinger, G. W., Vikan-Kline, L. L., and Fode, K. L. The effect of experimenter outcome-bias and subject set on awareness in

verbal conditioning experiments. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 275-283 (b).

Rosenthal, R., Persinger, G. W., Vikan-Kline, L. L., and Mulry, R. C. The role of the research assistant in the mediation of experimenter bias. Journal of Personality, 1963, 31, 313-335.

Rosenzweig, S. The experimental situation as a psychological problem. Psychological Review, 1933, 40, 337-354.

Sarason, I. G. Individual differences, situational variables, and personality research. Journal of Abnormal and Social Psychology, 1962, 65, 376-380.

Sarason, I. G. The human reinforcer in verbal behavior research. In L. Krasner & L. P. Ullman (Eds.), Research in behavior modification: New developments and implications. New York: Holt, Rinehart, & Winston, 1965. Pp. 231-243.

Sarason, I. G., and Minard, J. Interrelationships among subject, experimenter, and situational variables. Journal of Abnormal and Social Psychology, 1963, 67, 87-91.

Sattler, J. M. Racial "experimenter effects" in experimentation, testing, interviewing, and psychotherapy. Psychological Bulletin, 1970, 73, 137-160.

Shames, M. L., and Adair, J. G. Experimenter bias as a function of the type and structure of the task. Canadian Psychologist, 1967, 8, 176 (Abstract).

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

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

Smith, E. L., and Dixon, T. R. Verbal conditioning as a function of race of the experimenter and prejudice of the subject. Journal of Experimental Social Psychology, 1968, 4, 285-301.

Spires, A. M. Subject-experimenter interaction in verbal conditioning. Unpublished doctoral dissertation, New York University, 1960.

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

Stevenson, H. W., and Allen, Sara. Adult performance as a function of sex of the experimenter and sex of the subject. Journal of Abnormal and Social Psychology, 1964, 68, 214-216.

Stevenson, H. W., Keen, R., and Knights, R. N. Parents and strangers as reinforcing agents for children's performance. Journal of Abnormal and Social Psychology, 1963, 67, 183-186.

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

Summers, G. F., and Hammonds, A. D. Effect of racial characteristics of investigator on self-enumerated responses to a Negro prejudice scale. Social Forces, 1966, 44, 515-518.

Symons, R. T. Specific experimenter-subject personality variables pertinent to the influencing process in verbal conditioning situation. Unpublished doctoral dissertation, University of Washington, 1964.

Taffel, C. Anxiety and the conditioning of verbal behavior. Journal of Abnormal and Social Psychology, 1955, 51, 496-501.

Timaeus, E., and Lueck, H. E. Experimenter expectancy and social facilitation: I. Aggression under coaction. Psychological Reports, 1968, 23, 456-458 (a).

Timaeus, E., and Lueck, H. E. Experimenter expectancy and social facilitation: II. Stroop-test performance under the condition of audience. Perceptual and Motor Skills, 1968, 27, 492-494 (b).

Torgerson, W. S. A theoretical and empirical investigation of multidimensional scaling. Educational Testing Services Research Bulletin, RB-51-14, Princeton, N.J., 1951.

Torgerson, W. S. Theory and methods of scaling . New York: John Wiley & Sons, Inc., 1958.

Turner, G. C., and Coleman, J. C. Examiner influence on Thematic Apperception Test responses. Journal of Projective Techniques, 1962, 26, 478-486.

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

Weiss, L. R. Experimenter bias as a function of stimulus ambiguity. Unpublished manuscript, State University of New York at Buffalo, 1967.

Wessler, R. L. The experimenter effect on a task-ability experiment. Dissertation Abstracts, 1966, 27, Number 6, 2173-B.

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

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

Zoble, E. J., and Lehman, R. S. Interaction of subject and experimenter expectancy effects in a tone length discrimination task. Behavioral Science, 14, 1969.

APPENDIX A

Rating scale employed in the first, "colour-richness", condition of the pilot study and in Study I.

EXTREMELY COLOUR POOR	VERY COLOUR POOR	MODERATELY COLOUR POOR	MODERATELY COLOUR RICH	VERY COLOUR RICH	EXTREMELY COLOUR RICH

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

RATING SCALE

Rating scale employed in the second, "colour-value", and third, "colour-value" with defining criterion, conditions of the pilot study.

EXTREMELY LOW COLOUR VALUE	VERY LOW COLOUR VALUE	MODERATELY LOW COLOUR VALUE	MODERATELY HIGH COLOUR VALUE	VERY HIGH COLOUR VALUE	EXTREMELY HIGH COLOUR VALUE
-10 -9 -8 -7 -6 -5 -4	-3 -2 -1		+1 +2 +3 +4 +5 +6 +7 +8 +9 +10		

RATING SCALE

APPENDIX B

INSTRUCTIONS I.P.

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

"We are in the process of developing a battery of human engineering tests, one of which entails the judgment of colour. This test is designed to test your ability to discriminate colours and tap your aesthetic preferences for different colours. I will show you a series of cards. For each one, I want you to assess the "richness" or "poorness" of colour.

"To help you make more exact judgments you are to use this rating scale. As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the particular colour to be "extremely poor", or, more precisely, "extremely colour poor". A rating of +10 means that you judge the colour to be "extremely rich", or, more exactly, "extremely colour rich". A rating of -1 means that you judge the colour to be "moderately colour poor" while a rating of +1 means that you assess the colour as being "moderately colour rich".

"You are to rate each colour as accurately as you can. Just tell me the rating you assign to each colour on the cards I will show you. All ready? Here is the first card."

INSTRUCTIONS II.P.

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

"We are in the process of developing a battery of human engineering tests, one of which entails the judgment of colour. This test is designed to assess your ability to discriminate colour and tap your aesthetic preference for different colours. I will show you a series of cards and for each one I want you to assess the colour "value" of the colour stimulus mounted on the card.

"To help you make more exact judgments you are to use this rating scale in front of you. As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the particular colour to be of "extremely low colour value". A rating of +10 means that you judge the colour to be of "extremely high colour value". Similarly, a rating of -1 means that you judge the colour to be of "moderately low colour value" while a rating of +1, conversely, means that you assess the colour to be of "moderately high colour value".

"You are to rate each colour as accurately as you can. Just tell me the rating you assign to each colour stimulus on the cards which I will show you. All ready? Here is the first card."

INSTRUCTIONS III.P.

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

"We are in the process of developing a battery of human engineering tests, one of which entails the judgment of colour. This test is designed to assess your ability to discriminate colour and may, incidentally, tap your aesthetic preference for different colours. I will show you a series of cards and for each one I want you to assess the colour "value" of the colour stimulus mounted on the card.

"It will aid you in your task if the term value is given some definition. In the world of art and, more specifically, painting, value is conceived in terms of the "relation of one colour to another with respect to light and shade". Physicists refer to this property as "luminance" but it might, perhaps, be more helpful for your purposes to consider value in terms of colours "which differ among one another with respect to lightness or brightness. Thus, high value would be equatable with high intensity or brightness of a colour.

"To help you make more exact judgments you are to use this rating scale in front of you. As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the particular colour to be of "extremely low colour value". A rating of +10 means that you judge the colour to be of "extremely high colour value". Similarly, a rating of -1 means that you judge the colour to be of "moderately low colour value" while a rating of +1, conversely, means that you assess the colour to be of "moderately high colour value".

"You are to rate each colour as accurately as you can. Just tell me the rating you assign to each colour stimulus on the cards which I will show you. All ready? Here is the first card."

APPENDIX C

INDIVIDUAL SCORING SHEET

NAME _____

AGE _____ SEX _____

The presentation order of the pairs is:

(1) card #1 with card #2
(2) #3 and #4
(3) #5 and #6
(4) #7 and #8
(5) #9 and #10
(6) #1 and # 3
(7) #2 and #4
(8) #5 and #7
(9) #6 and #8
(10) #1 and #4
(11) #2 and #5
(12) #3 and #6
(13) #4 and #7
(14) #5 and #8
(15) #6 and #9
(16) #7 and #10
(17) #1 and #5
(18) #2 and #6
(19) #3 and #7
(20) #4 and #8
(21) #5 and #9
(22) #6 and #10
(23) #2 and #3
(24) #8 and #9
(25) #1 and #6
(26) #2 and #7
(27) #3 and #5
(28) #8 and #10
(29) #7 and #9
(30) #3 and #8
(31) #5 and #10
(32) #4 and #9
(33) #1 and #7
(34) #2 and #8
(35) #3 and #9
(36) #4 and #10
(37) #1 and #8
(38) #2 and #9
(39) #3 and #10
(40) #4 and #6
(41) #1 and #9
(42) #2 and #10
(43) #6 and #7
(44) #4 and #5
(45) #1 and #10

APPENDIX D

OBTAINED RANK-ORDERS OF THE CHROMATIC AND ACHROMATIC SETS OF STIMULI

Achromatic Stimuli	Chromatic Stimuli
8-1-6-10-2-4-7-9-5-3	6-1-2-9-3-10-4-7-8-5
8-1-6-10-2-4-7-9-5-3	8-7-6-10-1-4-3-2-5-9
8-1-10-2-6-5-4-7-9-3	6-10-7-1-8-4-2-3-9-5
8-10-2-4-1-7-6-9-5-3	6-7-10-1-8-3-4-5-2-9
8-1-6-10-2-4-7-9-3-5	6-10-7-8-1-5-2-3-4-9
8-1-6-10-2-4-7-9-3-5	6-10-8-7-4-1-3-2-9-5
8-1-6-10-2-4-7-9-3-5	10-6-7-8-1-4-2-3-5-9
8-1-6-10-2-4-7-9-3-5	6-1-3-2-10-7-9-4-5-8
8-1-6-10-2-4-7-9-3-5	6-7-10-1-3-8-4-2-9-5
8-1-6-10-2-4-7-9-3-5	6-7-10-8-1-2-4-3-5-9
8-1-6-10-2-4-7-9-3-5	6-7-10-1-8-4-2-3-5-9
8-1-6-10-2-4-7-9-3-5	6-7-10-8-4-1-3-2-5-9
8-1-6-10-2-4-7-9-3-5	6-10-7-1-8-2-3-5-4-9
8-1-6-10-2-4-7-9-3-5	6-10-8-7-1-4-5-2-9-3
8-1-6-10-2-4-7-9-3-5	6-2-7-1-5-3-10-8-4-9
8-1-6-10-2-4-7-9-3-5	6-7-8-10-1-4-2-3-5-9
8-1-6-10-2-4-7-9-3-5	6-7-8-10-1-4-2-9-3-5
8-1-6-10-2-4-7-9-3-5	6-7-8-4-10-1-2-5-3-9
8-1-6-10-2-4-7-9-3-5	6-10-1-7-3-5-2-8-4-9
8-1-6-10-2-4-7-9-3-5	6-7-10-1-8-3-4-2-5-9
8-1-6-10-2-4-7-9-3-5	6-7-8-10-1-3-2-5-9-4
8-1-6-10-2-4-7-9-3-5	6-7-10-8-1-4-3-5-2-9
8-1-6-10-2-4-7-9-3-5	10-1-6-7-8-5-3-2-4-9
8-1-6-10-2-4-7-9-3-5	6-8-7-10-4-1-5-3-2-9
8-1-6-10-2-4-7-9-3-5	10-6-1-7-4-8-2-9-3-5
8-1-6-10-2-4-7-9-3-5	6-10-7-1-3-2-4-8-9-5
8-1-6-10-2-4-7-9-3-5	6-8-7-10-1-4-3-2-9-5
8-1-6-10-2-4-7-9-3-5	10-6-1-7-8-3-5-4-2-9
8-1-6-10-2-4-7-9-3-5	6-7-10-8-1-4-2-9-3-5
8-1-6-10-2-4-7-9-3-5	6-10-1-7-3-8-5-2-4-9
8-1-6-10-2-4-7-9-3-5	6-7-8-4-10-1-3-2-5-9
8-1-6-10-2-4-7-9-3-5	8-7-4-6-10-1-3-5-2-9
8-1-6-10-2-4-7-9-3-5	6-8-7-10-1-3-5-4-2-9
8-1-6-10-2-4-7-9-3-5	7-6-2-4-9-10-1-8-3-5
8-1-6-10-2-4-7-9-3-5	6-7-10-8-5-1-4-3-2-9
8-1-6-10-2-4-7-9-3-5	6-8-4-7-2-9-5-3-1-10
8-1-6-10-2-4-7-9-3-5	6-10-7-1-8-4-5-3-2-9
8-1-6-10-2-4-7-9-3-5	6-8-7-4-10-1-3-2-9-5
8-1-6-10-2-4-7-9-3-5	6-7-10-8-4-2-9-1-3-5
8-1-6-10-2-4-7-9-3-5	6-10-1-8-4-7-3-2-9-5
8-1-6-10-2-4-7-9-3-5	6-7-8-4-10-1-5-3-2-9
8-1-6-10-2-4-7-9-3-5	6-10-7-1-8-2-3-4-5-9
8-1-6-10-2-4-7-9-3-5	6-7-8-10-4-1-9-2-3-5
8-1-6-10-2-4-7-9-3-5	6-7-8-4-10-1-5-3-2-9
8-1-6-10-2-4-7-9-3-5	6-10-7-1-8-5-3-4-2-9
8-1-6-10-2-4-7-9-3-5	6-7-10-1-8-2-3-4-9-5
8-1-6-10-2-4-7-9-3-5	6-7-8-10-1-3-5-2-4-9
8-1-6-10-2-4-7-9-3-5	6-7-8-1-10-4-3-5-2-9
8-1-6-10-2-4-7-9-3-5	6-7-10-8-4-1-2-3-9-5
8-1-6-10-2-4-7-9-3-5	6-1-10-7-2-3-5-9-4-8

APPENDIX E

INSTRUCTIONS

The task you have to perform today is a relatively simple one. All that is required is that you rate certain sets of stimulus cards on a scale of "colour-richness". You will be given two sets of ten stimulus cards each -- one set at a time -- which you will be required to rank-order, from your left to your right, in terms of colour-richness. Thus, the first stimulus card on your left will be the richest in colour, the one next to that the next richest in colour, and so on. You will then have a scale of colour-richness, ranging from the richest on your left-hand side through to the least rich stimulus card on your right-hand side. This will, then, be a ten-step scale of colour-richness. You will be required to do this for both sets of ten stimulus cards.

DO NOT ASK ANY QUESTIONS OF THE EXPERIMENTER since he is obliged to say nothing beyond the instructions you have before you. Simply ask the experimenter for the first set of stimulus cards upon completion of reading these instructions.

****

QUESTIONNAIRE

1. State as accurately as you can the criterion, or criteria you used for "colour-richness" in ordering the stimuli employed in the first scale_____

2. Why did you employ this (these) criterion (criteria)?

3. State as accurately as you can the criterion, or criteria you used for "colour-richness" in ordering the stimuli employed in the second scale_____

4. Why did you employ this (these) criterion (criteria) ?

5. Which set of stimuli were more difficult to order, the first or the second?_____

6. Why?_____

7. Apart from the stimuli employed in this study, what does the concept of "colour-richness" generally mean to you?

8. Could you repeat the ordering for scale 1 precisely if you had to perform this task, say, one month from now? Yes____ No____. For scale 2? Yes____ No____.

Name_____ Age_____

APPENDIX F

DATA RECORDING SHEET

Name_____

_____

Age_____ Sex_____

Faculty_____ Yr._____

Stimulus Card Rating

#1 _____

#2 _____

#3 _____

#4 _____

#5 _____

#6 _____

#7 _____

#8 _____

#9 _____

#10 _____

Comments (if necessary):

_____

_____

_____

_____

_____

_____

_____

APPENDIX G

Ss' MEAN RATINGS IN STUDY I

		Chromatic Stimulus Set					Achromatic Stimulus Set				
		$\underline{E}_1$	$\underline{E}_2$	$\underline{E}_3$	$\underline{E}_4$	$\underline{E}_5$	$\underline{E}_1$	$\underline{E}_2$	$\underline{E}_3$	$\underline{E}_4$	$\underline{E}_5$
+5 Bias	$\underline{S}_1$	+2.1	+3.3	+3.1	+3.1	+1.4	-0.2	-5.9	+3.2	-7.1	+1.7
	$\underline{S}_2$	+3.3	+1.4	+1.6	+0.5	+1.9	-1.3	-6.0	+2.6	-5.0	-0.8
	$\underline{S}_3$	+3.5	+1.4	+3.9	+0.2	+2.0	-2.3	+3.3	+3.9	-2.2	+1.9
	$\underline{S}_4$	+2.6	-0.3	+3.5	+0.6	+1.3	-0.5	+6.1	-1.4	+3.0	-5.0
	$\underline{S}_5$	+2.8	+1.4	+1.7	+0.5	+1.0	-9.2	-7.4	-4.1	-6.0	-0.7
Means		+2.9	+1.4	+2.8	+1.0	+1.5	-2.7	-2.0	+.84	-3.5	-.58
-5 Bias	$\underline{S}_1$	-0.1	+0.7	+0.4	-0.3	-4.4	-4.9	+2.7	-7.7	-8.7	+1.1
	$\underline{S}_2$	-1.8	+1.3	-0.8	+2.0	+3.0	+1.9	+2.4	+0.8	+0.1	-0.8
	$\underline{S}_3$	-1.6	-1.4	-3.5	+1.7	-3.2	-7.8	-6.2	+3.3	-1.9	+0.4
	$\underline{S}_4$	-2.5	-0.7	-2.6	+0.3	+0.7	-2.9	-7.4	+7.5	+3.3	+0.8
	$\underline{S}_5$	-1.0	-1.2	-0.0	+0.7	+3.0	-1.6	+1.3	-4.2	-1.9	+1.6
Means		-1.4	+.02	-1.3	+.88	-.18	-3.1	-1.4	-.06	-1.8	+.62

Grand Means

	Chromatic Stimulus Set	Achromatic Stimulus Set
+5 Bias	+1.912	-1.576
-5 Bias	-0.452	-1.152

APPENDIX H

QUESTIONNAIRE

The practice of modern psychology, more than ever before, is dependent upon quantification -- more precisely, the measurement of psychological (as well as physical) entities. To this end, certain measurement concepts have become basic to psychology. One such fundamental concept is dimensionality, the basis upon which judgments are made. Typically, the term dimension is defined as "an attribute or characteristic by means of which an object, event or score can be located on a quantitative scale of measurement".

Dimensionality can more easily be understood, however, as the basis upon which one makes a particular judgment -- this basis being more precisely defined in terms of the criterion (or criteria, since often there is more than one) which underlies the judgment. The first step in making a particular judgment is the identification (blatantly conscious or not) of the criterion, or criteria.

E.g. A subject is asked to judge the artistic merit of a number of paintings, i.e. he must rate the paintings on a continuum of artistic merit.

(1) He may base his judgment of artistic merit on only one criterion -- color harmony.

(2) He may base his judgment on two criteria -- color harmony and use of perspective.

(3) He may use a much more complex model by using many criteria upon which he bases his judgment of artistic merit.

Criteria such as these are herein defined as dimensions, the basis upon which judgments are made. Thus, (1) is using a single dimension, or unidimensional approach in making a judgment of artistic merit. (2) is using a multidimensional approach as is (3).

You have just made 10 judgments of relative "success" and "failure" of photographs of faces. Your task, in this questionnaire, will be to delineate the criteria or, more properly, the dimensions upon which you based your judgments of "success" and "failure". Try, as much as possible, to specify those dimensions you used when you were making your judgments -- not those that occur to you as you respond to this questionnaire.

Your frank and accurate responses to the items following will be of immense benefit to the experimenter in analysing the obtained results and in planning future research. You may, of course, feel free to include any additional information not presently in the questionnaire which you think might prove helpful to the experimenter.

Read the questions carefully before writing down your response. Most importantly, ANSWER THE QUESTIONS IN THE ORDER WHICH THEY APPEAR. Do not, for instance, freely respond so that question #5 is answered before question # 4 or #3, etc. DO NOT TURN OVER ANY SHEET AND LOOK AT THE NEXT QUESTION TILL YOU HAVE COMPLETED THAT SHEET -- IN ORDER.

Your task, thus, is quite simple. You may take any reasonable length of time to complete it. Thank you.

QUESTIONNAIRE 1

1. In your performance of the photo-judging task, what dimension (dimensions) of "success" and "failure" did you employ in making your judgments?

_____

_____

_____

2. Did you use only a single dimension of "success" and "failure"?

Yes_____ No_____

3. If you responded "No" to question #2, write down the dimensions you used in making your "success-failure" judgments in the space provided below. Write these down in terms of their order of importance as they influenced your judgments.

IMPORTANCE	DIMENSION
1st	_____
2nd	_____
3rd	_____
4th	_____
5th	_____

4. If you used multiple dimensions in your judgments of "success-failure", in what combination did you use them?

_____

_____

Did you apply this same combinational use of dimensions to all of the photos you judged, or were there variations?

_____

_____

DO NOT TURN THIS SHEET OVER

UNTIL YOU HAVE COMPLETED ALL OF THE ITEMS ON IT

QUESTIONNAIRE 2

1. Although you may not have used all of them, or any of them for that matter, would you please attempt to rank-order the following dimensions of "success-failure" in terms of their order of importance in making your judgments of "success-failure". Thus, if you consider a given dimension to be most important in making a "success-failure" judgment, you would place a "1" next to it in the blank space provided. The next most important dimension would then have a "2" placed next to it in the space provided, and so on for all the remaining dimensions.
These are the dimensions to be ordered:

   The intelligent-appearing aspect of the persons pictured _____

   The confident-appearing aspect of the persons pictured _____

   The pleasant-looking aspect of the persons pictured _____

   The happy-appearing aspect of the persons pictured _____

   The healthy-appearing aspect of the persons pictured _____

   The well-dressed aspect of the persons pictured _____

   The prosperous-appearing aspect of the persons pictured _____

   The fulfilled-appearing aspect of the persons pictured _____

2. If you feel that there are certain dimensions which might be of some importance in making judgments of "success-failure", and which have not been included in this listing, please specify these in the space provided below. Such a specification could prove to be of immense value to the experimenter in planning future research.

   _____

   _____

   _____

   _____

   _____

Name_____ Year in Course_____

APPENDIX I

Empathy Test Rating Scale employed in Study II for the purpose of judging the "success" and "failure" of the persons pictured in the person perception task.

EXTREME FAILURE			MODERATE FAILURE				MILD FAILURE			MILD SUCCESS			MODERATE SUCCESS				EXTREME SUCCESS	
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9	+10

EMPATHY TEST RATING SCALE

APPENDIX J

GENERAL EXPERIMENTAL PROCEDURE

In front of you, you will find the instructions that you are to read to your subjects, sheets of paper for recording each subject's rating of each of the stimulus cards, and a set of 10 numbered stimulus cards.

After recording the data required for each subject as specified by the data recording sheet, i.e. Name, Age, Sex, etc. and reading the instructions to the subject, you are ready to begin.

Take stimulus card #1 and hold it in front of the subject for approximately 5 seconds (you can judge this time very approximately by a mental count). After she tells you her rating of the stimulus card, in terms of a numeral (not a verbal label), you will then record this rating in the appropriate space on the data recording sheet which corresponds to the number on the stimulus card. Thus, the rating for stimulus card #1 will be entered in the blank space provided on the data recording sheet opposite #1 and so on, for the 10 stimulus cards. The order of presentation is given by these numbers on the back of the stimulus cards, from #1 to #10.

You will follow this procedure for all of the 10 stimulus cards and for all 10 of your subjects that you will be running. When you have collected all 10 ratings for each subject, simply calculate the average of the 10 ratings for each subject and notate it on the bottom of the data recording sheet.

PLEASE MAKE A DEFINITE ATTEMPT TO LIMIT THE SUBJECT'S VIEWING OF EACH STIMULUS CARD TO APPROXIMATELY 5 SECONDS, AND TRY NOT TO LET THEM SEE THE DATA RECORDING SHEET DURING THE PROGRESS OF THE EXPERIMENT.

APPENDIX K

INSTRUCTIONS TO SUBJECTS

UNIDIMENSIONAL CONDITION

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

We are in the process of developing a test of empathy. This test is being designed to show how well a person is able to put herself into someone else's place. For this purpose, I will show you a series of 10 photographs of faces. For each one I want you to judge whether the person pictured has been experiencing success or failure. To aid you in your empathic judgments of success and failure, previous research in this area has established that the most commonly used dimension or definition of success and failure is the appearance of <u>confidence</u>. Thus, the more confident a person appears, the more is he judged as successful. I want you to use precisely this definition of success and failure -- no other -- in making your judgments of the photographs. This is crucial to the study.

To help you make more precise judgments, using <u>confidence</u> as your criterion, you are to use this rating scale in front of you. As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the person to have experienced EXTREME FAILURE. A rating of +10, on the other hand, means that you judge the person to have experienced EXTREME SUCCESS. Similarly, a rating of -1 means that you judge the person photographed to have experienced MILD FAILURE, while a rating of +1, conversely, means that you judge the person to have experienced MILD SUCCESS.

You are to rate each photograph as accurately as you can, given the definition and rating scale provided. Just tell me the rating you assign to each photo in terms of a number. Try to make your ratings quite quickly since you will be presented with the photos for a very short period of time only. All ready? Here is the first photo.

(No further explanation may be given to the subject although all or part of the instructions may be repeated.)

INSTRUCTIONS TO SUBJECTS

THREE-DIMENSIONAL CONDITION

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

We are in the process of developing a test of empathy. This test is being designed to show how well a person is able to put herself into someone else's place. For this purpose I will show you a series of 10 photographs of faces. For each one I want you to judge whether the person pictured has been experiencing success or failure. To aid in your empathic judgments of success and failure, previous research in this area has established that the most commonly used dimensions or definitions of success and failure are (1) the appearance of confidence, (2) the appearance of fulfillment, and (3) the appearance of intelligence -- all three taken together. Thus, the more confident, fulfilled, and intelligent, a person appears, the more is he judged as successful. I want you to use precisely these three definitions of success and failure, TOGETHER IN COMBINATION, and no other definitions, in making your judgments of the photographs. This is crucial to the study.

To help you make more precise judgments, using confidence, fulfillment and intelligence as your criterion, you are to use this rating scale. As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the person to have experienced EXTREME FAILURE. A rating of +10, on the other hand, means that you judge the person to have experienced EXTREME SUCCESS. Similarly, a rating of -1 means that you judge the person photographed to have experienced MILD FAILURE, while a rating of +1, conversely, means that you judge the person to have experienced MILD SUCCESS.

You are to rate each photo as accurately as you can, given the definitions and rating scale provided. Just tell me the rating you assign to each photo in terms of a number. Try to make your ratings quickly since you will be presented with the photos for a very short period of time only. All ready? Here is the first photo.

(No further explanation may be given to the subject although all or part of the instructions may be repeated.)

(NOTE: Place the sheet labelled DEFINITIONS OF SUCCESS AND FAILURE in front of the subject, along with the EMPATHY TEST RATING SCALE.)

DEFINITION OF SUCCESS AND FAILURE

1) The appearance of CONFIDENCE
2) The appearance of FULFILLMENT
3) The appearance of INTELLIGENCE

these three taken together, in combination.

INSTRUCTIONS TO SUBJECTS

FIVE-DIMENSIONAL CONDITION

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

We are in the process of developing a test of empathy. This test is being designed to show how well a person is able to put herself into someone else's place. For this purpose, I will show you a series of 10 photographs of faces. For each one I want you to judge whether the person pictured has been experiencing success or failure. To aid you in your empathic judgments of success and failure, previous research in this area has established that the most commonly used dimensions or definitions of success and failure are (1) the appearance of confidence, (2) the appearance of fulfillment, (3) the appearance of intelligence, (4) the appearance of happiness, and (5) the appearance of prosperity -- all five of these taken together. Thus, the more confident, fulfilled, intelligent, happy, and prosperous a person appears, the more is he judged as successful. I want you to use precisely these five definitions of success and failure, TOGETHER IN COMBINATION, and no other definitions, in making your judgments of the photographs. This is crucial to the study.

To help you make more precise judgments, using confidence, fulfillment, intelligence, happiness, and prosperity, as your criterion, you are to use this rating scale as well. As you can see, the scale runs from -10 to +10. A rating of -10 means that you judge the person to have experienced EXTREME FAILURE. A rating of +10, on the other hand, means that you judge the person to have experienced EXTREME SUCCESS. Similarly, a rating of -1 means that you judge the person photographed to have experienced MILD FAILURE, while a rating of +1, conversely, means that you judge the person to have experienced MILD SUCCESS.

You are to rate each photo as accurately as you can, given the definitions and rating scale provided. Just tell me the rating you assign to each photo in terms of a number. Try to make your ratings quickly since you will be presented with the photos for a very short period of time only. All ready? Here is the first photo.

(No further explanation may be given to the subject although all or part of the instructions may be repeated.)

(NOTE: Place the sheet labelled DEFINITIONS OF SUCCESS AND FAILURE in front of the subject, along with the EMPATHY TEST RATING SCALE.)

DEFINITION OF SUCCESS AND FAILURE

1) The appearance of CONFIDENCE
2) The appearance of FULFILLMENT
3) The appearance of INTELLIGENCE
4) The appearance of HAPPINESS
5) The appearance of PROSPERITY

— these five taken together, in combination.

APPENDIX L

Ss' MEAN RATINGS IN STUDY II

		1-Dimension			3-Dimension			5-Dimension		
		E_1	E_2	E_3	E_1	E_2	E_3	E_1	E_2	E_3
	S_1	-0.9	+3.7	+2.9	+6.7	+2.5	+3.2	+6.3	+2.7	+4.5
	S_2	+2.4	+0.4	+3.3	+3.9	-0.1	+5.8	+5.2	+3.9	+2.9
	S_3	+4.6	+8.1	+1.7	+4.4	+0.8	-0.7	+0.5	+0.4	+3.0
	S_4	+5.3	+5.2	+3.2	+1.2	+1.6	+2.6	+3.1	+5.1	+3.4
	S_5	+2.8	+5.2	+1.9	+3.7	+3.7	+2.3	+3.2	+0.9	+2.4
+5	S_6	+1.9	+6.9	+0.7	+3.9	+1.0	+0.5	+6.4	+1.2	+3.4
	S_7	+2.3	+4.1	+3.8	+2.8	+3.6	+1.4	+4.0	+1.2	+1.7
	S_8	+2.3	+4.2	+4.5	+4.8	+1.0	+1.4	+2.1	+2.7	+1.7
	S_9	+2.6	+4.2	+3.2	+4.5	+2.9	+1.0	+4.4	+3.0	+1.7
	S_{10}	+3.0	+3.8	+4.4	+7.1	+3.6	+4.6	+8.9	+5.0	+0.1
	S_1	+1.4	+3.3	+3.8	+3.7	+2.9	+1.4	+1.1	+3.3	+3.9
	S_2	+2.2	+2.4	+3.6	+3.0	+6.1	+1.7	+5.0	-1.1	+0.3
	S_3	+0.5	+2.0	+1.4	+2.7	+2.3	+3.2	+0.8	+5.5	+1.3
	S_4	+2.7	+6.2	+2.0	+0.5	+1.3	+3.9	+1.2	+2.6	+3.0
	S_5	+1.5	+2.3	+4.4	+3.3	+3.4	+1.3	+1.1	+4.5	+4.8
-5	S_6	+2.7	+1.2	+2.3	+2.9	+3.4	+1.1	+1.5	-2.3	+1.8
	S_7	+2.8	+0.8	+4.1	+0.1	+3.0	+3.9	+0.6	+1.3	+2.1
	S_8	+2.8	+0.7	+1.5	+5.2	+3.0	+3.0	+2.4	+1.7	+ 3.5
	S_9	+0.1	+1.3	+1.0	+4.3	+3.0	+0.7	+3.7	+2.5	+1.7
	S_{10}	+3.4	-0.9	+3.1	+3.3	+2.3	+2.3	+1.2	+4.4	+2.0

APPENDIX M

QUESTIONNAIRE

1) Did you find this judgment task difficult? Yes _____ No _____

Why? _____

_____

2) Did you use the definition of success and failure provided precisely as you were instructed to do? Yes _____ No _____

Why? _____

_____

3) Did you find it difficult to translate success and failure into the terms provided by the definition given? Yes _____ No _____

Why? _____

_____

4) Did you find it difficult to translate success and failure into numerical terms -- as provided by the EMPATHY TEST RATING SCALE? Yes _____ No _____

Why? _____

_____

5) Did you find it easy to integrate all of the information provided for the purpose of making an empathy judgment? Yes _____ No _____

Why? _____

_____

6) What was the purpose of the experiment? _____

_____

7) Do you think you contributed to the purpose of the experiment? Yes _____ No _____

If Yes, how? _____

If No, why? _____

Name _____