THE ROLE OF THE EXPERIMENTER IN A CONCEPT LEARNING TASK

A Thesis

Presented to

the Faculty of the Department of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Martha H. Tyson

June, 1967

THE ROLE OF THE EXPERIMENTER IN A CONCEPT LEARNING TASK

An Abstract of a Thesis

Presented to

the Faculty of the Department of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Martha H. Tyson

June, 1967

ABSTRACT

Since the Renaissance scientific experimentation has been plagued by the experimenter's effect upon the observation, collection, and reporting of data. Only recently, however, has this variable been recognized as one that can be investigated. Psychological experiments with animal and human subjects in clinical, experimental, social, survey, and educational research have shown that experimenters can and do influence their data. Specific factors which have been investigated include experimenter's status, prior experience, previous interaction with subject, modeling effects, early data returns effects, sex, and expectancy. This experiment was designed to investigate the effect of experimenter's expectancy. The experimenter was told he was dealing with two different groups--a bright one and a dull one--and that these two groups would perform very differently in a concept learning task. A 2x2x2x4 mixed hierarchical analysis of variance design was used, with eight experimenters testing 64 subjects. No significant effects were found. A discussion follows, and the differences between this study and others which found significant differences in their results are explored. In future experiments it is suggested that the experimenter be given an opportunity to score, summarize, and interpret his data, and that the subject be allowed to make subjective judgment of his behavior

in relation to his perceived idea of whether or not he is pleasing the experimenter, and, if not, an opportunity should be provided for him to be able to alter his own behavior in keeping with the set or expectancy of the experimenter.

Table of Contents

Table of Contents (Continued)

List of Tables

Chapter I

The Nature of the Problem

Since the Renaissance, with the beginning of the application of the scientific method to experiments, the effect of the experimenter himself on his data has been significant in the observation, collection, and interpretation of his results. Boring (1950) describes an incident which illustrates the error of the experimenter as observer. Maskelyne, the astronomer royal at the Greenwich Observatory in England, had to discharge his assistant, Kinnebrook, because he was consistently "too slow" in his observations of the movements of the stars. Bessel, also an astronomer, twenty years later concluded that Kinnebrook's "error" must have been beyond his control, since differences in observation were the rule and not the exception and that these differences of "personal equations" varied over time (Boring, 1950). Rosenthal (1966) describes many other cases of observer, scorer, and interpreter errorsmade by the experimenter himself.

Definition

Various names have been given to this experimenter effect. The experimenter himself is called "a neglected stimulus-object" (McGuigan, 1963), and a "non-person"(Goffman, 1956). The process of interaction

between subject and experimenter has been described as being "learning without awareness" (Postman & Jarrett, 1954), "unconscious communication" (Rosenthal, 1966), and "selective perception" influenced by projection of needs, or "apperception"(Bellak, 1959). No matter what semantic differences exist, the experimenter effect is recognized as the results of the impact of the experimenter upon the subject, of the subject upon the experimenter, and in the interaction between the two.

Description

Lane (1953) writes in a paper on the octopus that scientists may "equate what they <u>think</u> they see and sometimes what they <u>want</u> to see with what actually happens." M. L. Johnson (1953), in a paper entitled "Seeing's Believing," said that, "Our assumptions define and limit what we see, i.e., we tend to see things in such a way that they will fit in with our assumptions, even if this involves distortion or omission. We therefore may invert our title and say 'Believing's Seeing'"(p. 79). Bean (1948) in research on reports of nutritional examinations by physicians found that many experienced doctors disagreed in the diagnosis of deficiency, even when objective standards were used. He pleads that, "Our aim must not be to deny error, but to learn from it, avoiding the stability it gets from repetition" (p. 454). Bertrand Russell (1927) has a pertinent and

sharp observation to make about the whole process:

> One may say broadly that all the animals that have been carefully observed have behaved so as to confirm the philosophy in which the observer believed before his observations began. Nay, more, they have all displayed the national characteristics of the observer. Animals studied by Americans rush about frantically, with an incredible display of hustle and pep, and at last achieve the desired result by chance. Animals observed by Germans sit still and think, and at last evolve the solution out of their inner consciousness (pp. 29-30).

Experimenter effects tend to be consistently too high or too low, almost never being randomly distributed. Roe (1961) and Pearson (1902), among others, comment that this bias bears some relation to the characteristics of the observer and to the observation situation. Yule (1927) and Fisher (1936) take opposing views about the possibility of eliminating this source of variance. Yule, reporting on the tendency to read a scale in quarters rather than in tenths, believed that training could eliminate the error, but Fisher was quite pessimistic about the outcome.

Rosenthal (1966) in a book entitled, Experimenter Effects in Behavioral Research, gives a detailed review of the literature, describing a number of experiments he and his associates have done in an effort to identify the ways the researcher influences his data. Rosenthal thinks "the study of the behavioral-scientist-experimenter is crucial, for there are important implications for how we conduct and how we assess our research" (p. viii). He doubts that anyone will be surprised

about the possibility of unintentional influence of data, but it is important to realize that this process can be observed in the laboratory, and that an attempt can be made to investigate its dynamics fully and systematically. Rosenthal centers his research around the question of whether it is possible to account for increasing proportions of the total variance in experiments by a consideration of experimenter effects, and thus reduce these sources of error (see p. viii).

Kintz, Delprato, Mattee, Persons, and Chappee (1965) note that most crucial experiments, particularly in psychological learning theory, have produced generally inconclusive results, with the exception of a high correlation between the theory of an experimenter and the support of his theoretical position by his research (p. 230). They believe this is caused by the experimenter variable, and that it contaminates most psychological investigations of today. They observe that, to the extent that we hope for dependable knowledge in the behavioral sciences generally, and to the extent that we rely on the methods of empirical research, we must have dependable knowledge about the researcher and the research situation.

This present study of the experimenter effect on the performance of subjects in a concept-learning task investigates whether, in fact, the expectancy of the experimenter about the intelligence of the subject he is testing, i.e., whether he is bright or dull, changes the performance of the subject in the direction of expectancy.

Chapter II

Review of Previous Research Findings

This chapter will present some of the abundant evidence that the experimenter can and does influence the results of his research in his observation, in his reporting, and in his interpretation of his data.

The Beginnings

Boring (1950) describes the beginning of the notation of observer error in the incident of the discharge by Maskelyne of his assistant, Kinnebrook, because he was consistently "too slow" in his observations of the movements of the stars.

The actual study of experiment effect as an independent variable began in Germany in 1904. Professor Stumpf and his associates investigated the "amazing" horse of Mr. von Osten. They made meticulous and systematic measurements, noting in particular how the horse was able to pick up unconscious cues from the questioners (Pfungst, 1911). The horse, Hans, by tapping his foot, was able to add, subtract, multiply and divide. He could read, spell, and solve problems of musical harmony. His owner was a mathematics instructor, and the investigators found that he had trained Hans to pick up the smallest cues from him. A forward inclination of his head would start Hans tapping, and a straightening-up would cause him to stop. Even the raising of an eyebrow

or the dilation of a nostril was enough to start the tapping.

Both Ebbinghaus and Pavlov speculated about the effect of unconscious influences. Ebbinghaus (1913) mentioned the importance of the "secret" influence of theories and opinions held before an investigation started or which developed during the research. He felt that this influence was a source of error "which may occur, and...when it does, it is a source of great danger." As the investigation is carried further, these theories and "these suppositions...constitute a complicating factor which probably has a definite influence upon the subsequent results" (p. 184). Pavlov (cited in Gruenberg, 1929, p. 327) explained that in checking on experiments with successive generations of mice who had an apparent increase in learning ability he found that the apparent improvement in the ability to learn on the part of successive generations of mice was really due to an improvement in the ability to teach on the part of the experimenter.

Other Disciplines

Examples of experimenter effect from other disciplines are abundant. In the physical sciences Newton, Blondlot, and Michelson and Morley were victims of this source of error. Newton, according to Boring (1962), did not see and report the absorption lines in the prismatic solar spectrum because of his theoretically-biased expectations. Blondlot, reported by Rostand (1960), discovered the famous

or rather the infamous "N-rays "which appeared to reflect light more intensely than other rays. Only a few experimenters were unable to detect these rays, which were later evaluated as emanating from at least a colossal observer error, if not a fraud. In 1887 Michelson and Morley conducted their famous experiment on the speed of light, which is said to be most important in the development by Einstein of the theory of relativity. Polanyi (1958) gives the details of the experiment, which, in his opinion, were subject to observational error.

In the biological sciences, one of the most famous cases has to do with Mendel's classic monograph, Experiments in Plant-Hybridization. Although it was presented in 1865, it was 1900 before three different investigators, all working independently, found error in this experiment. Fisher (1936) believed that this gap was caused by the fact that, "Each generation, perhaps, found in Mendel's paper only what it expected to find,... each generation, therefore, ignored what did not confirm its own expectations" (p. 137). In his statistical analysis of Mendel's data, Fisher showed that Mendel could not reasonably have obtained the data he reported; it could have only occurred by a bias in either Mendel, his assistant, or both. Another important, systematic, and replicable error for many years occurred in the counting of blood cells until Berkson, Magath, and Hurn (1940) reported a way of counting these cells more accurately with an electrical apparatus.

Rosenthal (1966) states that the investigators were led to conclude inescapably that laboratory technicians had for years reported blood cell counts that could have agreed with one another so well only 15 to 34 per cent of the time (p. 5). Darwin, Lister, Pasteur, and Semmelweiss are also mentioned as being among those whose research was contaminated.

Alfred Binet, who later helped devise the first test to predict school success, discharged one of his assistants for making errors in cephalometric measurements in an anthropological study. These errors, like those possibly committed by Mendel's assistant, were not necessarily errors of observation (Wolf, 1961).

With the advent of psychiatry upon the scene Harry Stack Sullivan (1936) and Wirth (1936), among others, were very much aware of the implications of the fact that the observer or experimenter affected his results in the fields of psychiatry and clinical psychology. Roy Schafer (1954) gives a detailed picture of the effects of interaction between the examiner and examinee from both points of view; his observations concerning the effects of each upon the other have had a significant effect upon clinical reports and upon research in the field of testing.

Animal Research

Brogden (1962) writes that it "has been long recognized that the behavior of the experimenter may be an important source of stimulation

to an animal subject in the conduct of experiments, and thus a source of experimental error" (p. 239). He reports data involving speed of conditioning of a first and a second group of rabbits, in which there was a significant experimenter practice effect and a significant difference between examiners.

Cordaro and Ison (1963) and Rosenthal and Halas (1962) report experiments using planaria (flatworms placed low on the phylogenetic scale) as examples of experimenter effects in invertebrate behavior. In the former experiment, seven observers were led to expect a very high incidence of turning and contracting for half of the worms. For the remaining worms the same observers were led to expect a very low incidence of turning and contracting. The worms observed were, however, essentially identical. Observers reported twice as many head turns and three times as many body contractions when the set was for high rates of response as contrasted with the set for low rates. This experiment was then repeated, employing a new set of ten observers. This time half the experimenters were to observe only "high-response-producing" worms, and the remaining were to observe only "low-response-producing" worms. Again there was no real difference in the worms. The observers found nearly five times as many head turns and twenty times as many contractions under the high level of responding set. The Rosenthal and Halas experiment investigated

"natural" differences among workers interested in planaria. It was found that differences among observers in the reported number of turnings and contractions were nearly all statistically significant when high levels of responding were expected. These experimenters were not given any false expectancies; they were engaged in actual research on planaria, and they were more experienced than those used in the Cordaro and Ison project. Even for these experienced experimenters it was concluded that there are individual differences in the extent to which behavior modifications in planaria are observed, and that the particular differences found are affected by the specific type of behavior being observed. Unlike the Brogden studies, no practice effects were found, and the experimenter effect did not disappear with further experience of the experimenter. Rosenthal (1966) points out that neither in the case of the rabbits nor in the case of the planaria could it be specified just what the experimenters did differently that could have led to such different records of animal learning.

Human Learning

Clinical Research

The rise and spread of clinical psychology has led to research in several areas demonstrating the universality of the experimenter effect: in the use and interpretation of projective techniques, in psychological reports, and in the area of psychotherapy. Clinical psychology has led

the way to investigation of the interaction of the personalities of the experimenter-subject and of the effect of the situation in which they find themselves.

Studies using the Rorschach ink blots have been numerous and important in the field of projective techniques, but only a few of the most significant can be mentioned here. They all have in common the fact that the interviewer or clinician projects his own needs and emotions into the testing situation. Hunter (1937) compared the responses on the Rorschach test of whites and negroes, finding less production when negro interviewed negro, and more when white interviewed negro. Schactel (1945) preceded Schafer in his emphasis on the subjective definitions of the test situation, noting that the testee may feel competitive, bored, listless, resistant, or depressed, with each mood having its effect on the Rorschach responses. Lord (1950) examined responses on the Rorschach in three different types of situations, with 36 subjects taking the Rorschach three times--once each from three different female examiners. Twenty-seven out of the 48 differences within subjects were attributed to examiner differences.

The personality of the tester has been assessed in different ways, and the results have been correlated with the results obtained from the giving of the Rorschach test. Sanders and Cleveland (1953) found that overtly anxious experimenters, as indicated by their own

Rorschach responses, tended to elicit more subject flexibility and responsiveness, while overtly hostile experimenters drew more passive and stereotyped responses and less of the hostile responses. Questionnaires filled out by the subjects after taking the Rorschach from the examiners indicated that experimenters who were most liked were those who had been rated low on anxiety and hostility. Masling (1959) found that personal warmth or coldness caused different results in projective testing. Gibby, Miller, and Walker (1952) examined the protocols of inexperienced trainees, and then the protocols of the same trainees after completion of their training, and they found the protocols now more in keeping with their supervisor's personality and more nearly the same. Rosenthal, Persinger, and Fode (1962) found that the experimenters' personality and personal bias do interact. This was one of a long series of studies carried out by Robert Rosenthal and his associates.

Other projective techniques that have been studied are home-made ink blots (Wickes, 1956), the Thematic Apperception test (Murstein and Easter, 1966), the Kohs Block Design test (McGuigan, 1960), the Draw-A-Person test (Holtzman, 1952), and photographs of people taken at random from magazines (Friedman, Kurland, and Rosenthal, 1965). This last experiment was filmed in an attempt to pin down how this non-verbal communication takes place. Experimenters had been

led to expect ratings of success from some of the subjects and failures from others. Again, the subjects were randomly assigned to groups. Experimenters whose behavior reflected greater interpersonal involvement or warmth obtained ratings of the photographs as those of more successful people, while experimenters whose behavior reflected a greater task orientation, greater competence, and a more professional manner obtained ratings significantly more in accord with their expectancy, regardless of the particular nature of that expectancy.

Clinical psychology is inextricably involved in report writing; in this area, studies by Robinson and Cohen (1954), Hyman, Feldman, Hart, and Stember (1954), and Star (1950) are representative. Robinson and Cohen investigated individual biases of three internes in psychology who had, in a rotating service over a year's period, studied thirty patients each. Pronounced and reliable differences among all three were found; the authors concluded that systematic individual biases exist in the reporting of patients' personalities, and they may be related to the individual personality of the reporter in a systematic fashion. They suggest these findings raise a serious question about the use of psychological reports for the evaluation or prediction of behavior--an everyday occurence. Hyman, et al., (1954) compared counselor's written reports of interviews with an electrical transcription, and found large and significant omissions of content in the written records, alterations in

the time sequence of remarks, and lack of precision in the notes, leading to ambiguity. They refer to the work of nine different clinicians who administered about 500 Rorschach tests to soldiers in World War II. All examiners had received the same rigorous course and had the same standardized instructions to give to their subjects. Significant differences were obtained in the number of responses given. These authors also mention one of the California Growth studies in which three clinicians, working in close cooperation with a given group of children over a period of seven years, rated the presence of certain needs. Although there was considerable agreement in the ratings of single needs, there were marked differences in the degree to which each clinician found sets of needs co-existing in the subjects (p. 13). Star (1950), in a report of the magnitude of differential diagnoses of abnormal behavior, reported that during World War II one induction center rejected 100 times more recruits than did another.

The golden, growing specialty of psychotherapy has generated many studies about the therapist and the therapeutic process. When Greenspoon (1955) demonstrated the potentially reinforcing effect of the interviewer's behavior in survey research, it occurred to some clinicians that non-directive therapy was not as non-directive as had been thought. Graham (1960) reports a study in which it was found that in therapy the patients tend to become more like their therapists.

Therapist expectancy was found to be the critical factor in studies by Goldstein (1962), Sanders and Cleveland (1953), and Verplanck (1955). Goldstein showed that clients who are rehabilitated by a particular technique may be more products of perceived therapist expectancies than of therapeutic techniques; Sanders and Cleveland suggested the possibility that the therapist may be a contributing factor to the patient's failure to recover as the result of perceived negative therapist expectancy. Carson and Heine (1962) and Edward Lichtenstein (1966) found contrasting results in studies of a relationship between patient-therapist personality similarity and success of psychotherapy. The first study hypothesized and found empirical support for a curvilinear relationship, but the second found none. One theory about therapeutic success which is understandably not very popular is that patients called "rehabilitated" may only have adjusted to the wishes of the therapist, and not necessarily to the emotional problems that brought them there in the first place (Verplanck, 1955).

Experimental Research

Since 1955 there has been a marked increase in the number of reported studies on the experimenter variable in conditioning studies, especially in the area of verbal learning. Two interesting experiments similar to Greenspoon's (1955) were done by Verplanck (1955) and Azrin, Holz, Ulred, and Goldiamond (1961). Verplanck's experiment concerned

the control of the content of conversation by reinforcement of statements of opinion; he concluded that, following certain reinforcement procedures, the content of conversation could be changed. Azrin, et al., obtained similar results, but later found that some of the data were fabricated by the student experimenters. Azrin's group then suggested that Verplanck's student data collectors had deceived him, but of course that was only supposition. In an interesting study on "learning without awareness" Postman and Jarret (1952) used 30 different experimenters; these experimenters required subjects to respond to each of 240 stimulus words with another word which "came to mind." One-half of the subjects were told to guess, and the other half were told the "correct" principle of answering, which was to give common associations such as those found in speaking and writing. Differences among experimenters were found to be highly significant sources of data variance. Kanfer (1958) reinforced verb responses with a flashing light under three difference reinforcement schedules. The experimenters were simply required to distinguish between verb and non-verb in reinforcing subjects, and yet even with this relatively elementary task it was found that ability to perform was highly subject to individual differences.

Three teams of researchers, investigating the human reinforcer in verbal behavior research, found significant differences between their experimenters: Friedman, Kurland, and Rosenthal (1965), using sound

motion pictures of examiners conducting a person perception experiment, found that experimenters showed significant behavioral variations in the way in which they conducted the experiment; Binder, McConnell, and Sjoholm (1957), studying verbal conditioning without awareness, used two experimenters of different sex and markedly different height, weight, age, appearance, and personality, who reinforced subjects' use of a hostile verb by saying "Good" whenever such a verb was used in a freely constructed sentence. They found that the rates of learning for the subjects of the two experimenters differed significantly, with the female experimenter's group learning much more quickly; Sarason and Minard (1963), manipulating experimentally the degree of personal contact between subject and experimenter and the value of experimenter's prestige, reinforced subjects for emission of first person pronouns, and found that both individual differences and experimental variables significantly influenced subjects' performances, which findings they interpret as indicating the value of approaching experimental situations from the point of view of interpersonal transactions.

There have been several studies on the subject's perception of the experimenter: Spires (1960, cited in McGuigan, 1963), Rosenthal, Fode, Friedman, and Vikan-Kline (1960), and Rosenthal and Persinger (1962). All these studies found a significant effect. The latter study is especially interesting since, although the experimenter

was imaginary, a median correlation of .81 was obtained between subjects' ratings of experimenters on a number of variables. Rosenthal (1966) attributes this high correlation value to the pre-determined idea by the subject of what a "typical" experimenter is like--scientific, intelligent, etc.

In the study of the effect in question many different procedures have been used. the galvanic skin response (Rankin & Campbell, 1955); digit grouping (Severin & Rigby, 1963), digit span (Young, 1959); performance on arithmetic tasks (Murstein & Easter, 1965); and the prediction of subject compliance (Gore, 1962). Performance with all procedures has been shown to be subject to widespread examiner variance.

Social and Survey Research

In the area of social psychology and survey research many investigators have assessed the effect of the interviewer's own opinion, attitude, or ideology on the reponses obtained from the respondents. Evidence for this phenomenon has been discussed and summarized by Hyman, et al., (1954) and Maccoby and Maccoby (1954). Where modeling effects were found, they tended to be in the direction of those of the interviewer, but in a minority of cases the effect of the interviewer's own opinion or ideology has been negative, so that the subject responded in a direction significantly opposite to that favored by the interviewer himself (Rosenthal (1963b).

The work previously mentioned by Greenspoon (1955) and Verplanck (1955) on "verbal" reinforcement contributed to the awareness of the problem of interviewer influence. Wickes (1956) in a similar study had the examiners' comment, "Good," "Fine!," and "All right," while at the same time commenting non-verbally by nodding the head and leaning forward in the chair. His results support his contention that examiners should be alert to the fact that even under presumably "standardized" conditions it is possible for their behavior to be reflected in test results.

Studies by Clark (1927) and Rice (1929) illustrate positive modeling effects. Clark had two interviewers inquire of 193 subjects how much of their time was devoted to various daily activities. One of the interviewers was more athletically inclined than the other, his subjects reported a greater amount of time spent in athletic activities than did the subjects contacted by the other interviewer. The direct effect of interviewers' attitudes towards drinking and politics was found in reported reasons by derelicts for their "downfall" (Rice, 1929).

Hyman, et al., (1954) report a study of the effect of sex differences of interviewers. Respondents were given a brief description of a proposed motion picture plot and asked whether they would like to see such a movie. Both male and female interviewers obtained responses significantly contingent upon their sex, and, perhaps, upon the respondents'

inferences of what movies the interviewers (perhaps because of their sex) would themselves enjoy. The plot of "Lawrence of Arabia" was used. When asked by members of their own sex, male subjects were 50 per cent more often favorable to the film than were the females; but, when the interviewer was of the opposite sex, male subjects responded favorably only 14 per cent of the time more often than female subjects. How much of this was due to the fact that Lawrence of Arabia is widely known to be a homosexual is not known.

In cross-cultural research the assessment of cultures is also subject to widely divergent interpretations (Hyman, et al., 1954). Oscar Lewis and Robert Redfield described the same Mexican village, Tepoztlan, quite differently: Redfield saw a highly cooperative, integrated and happy society, Lewis saw a poorly integrated, uncooperative society whose members were anything but happy. In describing the Arapesh Margaret Mead and Reo Fortune used quite different terms: Mead saw a placid, domestic people characterized by a maternal temperament, Fortune saw a national attitude that "warfare was good Arapesh custom." This warfare was not conducted in order to promote the natural increase of clans, but when one man coveted another man's wife his desire was implemented by divorce and war. He did note, however, that under German and Australian administration warfare had been suppressed (Fortune, 1939, p. 27). These differences are quite clearly attributable to

differences in the perception of the investigator, and thus are a result of experimenter error.

In an effort to control this source of variance, Dohrenwend and Richardson (1956) present a scheme for classification of interviewers' behavior. Also, Franzer and Williams (1956) suggest that when variance within subjects, as compared to the variance of total subject is proportionate, the questioning is dependable but that, when the ratio is greater than 1.25 or less than .75, the questioning is not objective, and the difference in responses lies in the differences of the interviewers.

Educational Research

The general field of education is closely involved with testing, including IQ tests, placement tests, readiness tests, etc., and thus is also plagued with the problem of experimenter effect.

Even though there have been rigorous attempts at standardization of test items and test procedures, examiners still influence the test taker in other ways. In an early study by Marine (1929), it was found that familiarity with the experimenter had a significant effect on performance on the Stanford-Binet. Sacks (1952) studied experimentally established social relationships with the children, and he also found a significant effect.

In a clear demonstration of the experimenter effect or bias in education, 256 prospective school teachers (Cahen, 1965; cited in

Rosenthal, 1966, p. 22) were each asked to score several test booklets ostensibly filled out by children being tested for academic readiness. Each of the 30 test items was to be scored on a four-point scale; a scoring manual provided examples of answers of varying quality. On each of the answer booklets some background information for that child was given, including an alleged IQ score. The purpose of this background information was to create an expectation in the scorer that the child whose booklet they scored was (a) above average, (b) average, or (c) below average in intelligence. The scoring of the tests supported Cahen's hypothesis that children thought to be brighter receive higher scores for the same performance than children believed to be less able.

Kintz, et al., (1965) suggest that the accurate interpretation of a person's IQ score requires knowledge of kinds of interactions and expectancies on the part of both teacher and subject--e.g., whether the teacher expected a high or a low score, whether the subject was tense or at ease in the testing situation, and how the experimenter was perceived (docile, friendly, etc). Sometimes teachers and experimenters rate their subjects on cooperation instead of achievement (Braunstein, Braunstein, and Blumenfeld, 1965). This is particularly noted in education in the lower grades.

Various suggestions have been made concerning the control of the teacher-child interaction (Medley & Klein, 1947; Weitz, 1956; and

Mayhew, 1957). Medley and Klein, in a study inferring classroom behavior from pupil response, suggest ways to control the "halo effect"; Weitz, in an interesting short article, pleads for replication of critical experiments in the classroom as a means of control; Mayhew (1957) suggests control through the personality of the teacher. Binder, McConnell, and Sjoholm (1957), McGuigan (1963), Rosenthal (1963), and Kintz, et al., (1966) all suggest that the administrator effect and other contamination problem may eventually be resolved by the use of machines in the administration of tests. It may be, however, that using computers or mechanical devices only delays the experimenter effect until a later time.

Specific Factors Influencing the Experimenter Effect

Many studies of specific factors which can be identified as components of this variable have been done, but only more effort and time will complete the task. Research on only a few of these specific factors is described here.

Experience of Examiners

Investigators with widely variant amounts of experience are busily conducting studies every day, but very often this is not taken into account in reporting the results of the studies. Cantril (1944) and Brogden (1962) come to different conclusions about the amount of bias experimenters gain or lose with more experience. Cantril's study of refusals

as a source of bias in public opinion research found that interviewers with a great deal of experience showed as much bias as those who were less experienced, while Brogden's study of rabbits previously mentioned found a significant practice effect and a significant difference between examiners.

This particular area needs to have more research on the differential effects of naive vs. experienced experimenters.

Status

Surveys conducted by the Federal Bureau of Investigation are likely to earn a degree of cooperation quite different from that earned by the manufacturer of soap or detergent, a fact which is well known among specialists in survey research (Hyman, et al., 1954).

McConnell (1955), Verplanck (1955), Orne (1962), and several of the Rosenthal studies (1966) demonstrated the effect of the perceived status of the experimenter upon the results of the study. McConnell demonstrated that prestige of the examiner influenced the type of response received. More specifically, Verplanck showed that examiners with the most prestige were more successful in bringing about unconscious reinforcement in persons of lower prestige. Orne's study emphasized the social psychology of the psychological experiment with particular reference to the expectations of the subjects in a test or interview situation. The examiner was perceived as knowing what

he is doing, and thus provided special motivation for the subject to agree with the examiner.

In contrast, Blaufarb's study (1960) on the relation of experimenter status and achievement imagery to success in conditioning verbal behavior found that the status of the experimenter was not a significant factor in the results.[a]

Vikan-Kline (1962) used six male faculty members and six male graduate students as experimenters. They were asked to attempt to influence half of their subjects to rate photographs as those of successful people, and the other half in the opposite direction. Results showed that the faculty experimenters were more successful in influencing their subjects, but only among subjects contacted later in the experiment. In the early series of subjects the faculty experimenters were, if anything, less successful influencers than the graduate student experimenters. Rosenthal (1966) attributes this to the "early data returns" effect. These experimental results are also confounded with modeling effect--to be considered next.

Modeling Effect

A subject will frequently try to model himself to be like the experimenter, or will try to be what he thinks the experimenter wants him to be.

---

[a]This negative finding is the only one found in the literature available, but it seems likely that others do exist.

Birney (1958), like Vikan-Kline, used faculty members as investigators with the expectation that they would probably produce a significant modeling effect. He found that his two faculty experimenters obtained responses from subjects reflecting a higher need for achievement than did student experimenters.

Blankenship (1940) and Cantril (1944), among others, found that interviewers elicit from their respondents responses which agree with their own beliefs (see also Rice, 1929, Rosenthal, 1963, and other studies discussed under survey research). Blankenship used a constant time element for each interviewer and comparable subjects. In general the attitudes of the interviewers correlated with the results they secured. In a study of psychotherapists, Graham (1960) found that those who showed a higher movement response in their own Rorschach protocols had a significantly higher movement response in their subjects' protocols.

Early Data Returns Effect

The "early data returns" effect is defined by Rosenthal (1966) as the contamination that is apt to be produced when the experimenter receives feedback of obtained results during earlier phases of his experiment. It may be that this is due to a mood change in the experimenter (Rosenthal, Persinger, Vikan-Kline & Fode, 1963) brought about by "good results," which might lead him to be perceived by the

subject as more likeable and personable. In this particular study three groups of four experimenters each had three groups of subjects rate the apparent success of people in photographs. In each of the two experimental groups, two subjects were confederates, while in the control groups all were naive. Each confederate was instructed to give "good" or "bad" data which would be contrary to the experimenters' expectations. Results for the experimental groups were significantly different from the control groups; a tendency for these "early returns" to affect later responses was reflected in later stages of data gathering.

Although there is undoubtedly evidence for this effect in political elections, this general area needs further investigation.

Previous Interaction

Sacks (1952) and Marine (1929) studied the effect of previous contact on intelligence test scores. In the former study, done with three-year-old children, significant effects were observed, these effects were also related to the warmth of that contact. The latter study was carried out with older children, here, previous interaction had no effect on gain in IQ scores.

Kanfer and Karas (1959), in a study using college students as subjects, investigated the effects of prior contact on success of conditioning first-person pronouns. Of the four groups of subjects, three had had prior contact with the experimenter and one had not. All three with prior

contact conditioned more rapidly than did the other group.

The nature of the task seems to be a confounding factor in the effect of previous interaction. In a simple, repetitive motor task--dropping marbles in holes--strangers seem to be more effective reinforcers than experimenters previously known by the subjects (Stevenson, Keen, & Knights, 1963). In a study by Berkowitz (1964) with 39 chronic schizophrenics and 39 medically hospitalized normals, the effect of previous contact, both "warm" and "cold," was studied; the subjects were given a task consisting of simple finger-lift reactions in response to a buzzer. He found that prior contact in the simple task seemed to decrease the anxiety of the patient, and his performance was slower. Rosenthal's(1966) interpretation was that when the performance required is difficult, prior contact (especially when "warm") improves performance, when the task is simple, performance may deteriorate, although subjects feel more relaxed about it. When the task is of medium difficulty, no reliable or clear prediction is possible.

Sex

The effects of sex of the experimenter is clearly shown in a verbal conditioning study (Binder, McConnell, & Sjoholm, 1957). Two very different experimenters were used: a young, pretty, feminine girl and a mature, large masculine male. The fact that she tested six females and sixteen males, and he tested five females and ten males, may have

been an important factor in the results obtained: the attractive female examiner's subjects conditioned significantly more quickly than did those of the "ex-marine" male's. Another clear-cut demonstration of the effect of sex on results was provided by Stevenson and Allen (1964). In a simple sorting task, eight male and eight female experimenters tested eight male and eight female subjects. The mean number of responses was recorded at 30-second intervals. With both sexes of experimenters female subjects gave more responses than males, but all performed better under an experimenter of the opposite sex.

Rosenthal (1966) found that, in such things as accuracy of instruction reading, time required for stimulus preparation, degree of leaning towards subjects, and ratings of experimenters, sex of the experimenter was a determinant which confounded experimental results, usually to a statistically significant degree.

Sex of the experimenter, therefore, clearly affects results and should be taken into account in any experiment.

Expectancy

Rosenthal (1966) devotes over one-third of his book to studies of experimenter expectancy effects. He reviews the literature in detail on such sub-topics as animal and human studies, subject set, excessive rewards, structural and behavioral variables, and communication

of expectancy.

This is the component of experimenter effect which causes the greatest concern to those interested in getting objective results in experiments. Relatively little research has been carried out in this specific area until recently, but Rosenthal and his group at Harvard are now engaged in a series of experiments specifically designed to pin down the method by which this expectancy is communicated.

Two significant studies, one by Rosenthal and Fode (1963) and the other by Rosenthal and Lawson (1964), have been widely discussed and have generated further research. In the former study two groups of randomly assigned rats were provided to two groups of six experimenters. One group of the experimenters was instructed that its group of rats was "maze-bright," while the second group was instructed that its rats were "maze-dull." And they were! In a simple T-maze, the maze-bright performed significantly better than the maze-dull. In the latter study, the investigators divided 38 experimenters into 14 research teams, each of which had one rat randomly assigned to it. Six of the teams were told that their rats were bred for dullness, and the other eight were told their rats were bred for brightness. Seven experiments, using such tasks as operant acquisition, stimulus discrimination, and chaining of responses, were conducted. In seven out of the eight

comparisons (overall $p < .02$), differences in performance again favored the experimenters who believed their subjects to be bred for brightness.

Chapter III

The Purpose

There is no apparent reason to think that the experiments on the maze-bright and maze-dull rats were not reported correctly, and there is little doubt that this experiment did, in fact, obtain significantly different results from two different, yet same, groups of rats. There is, however, doubt about whether this study could be replicated, and especially with the use of human subjects instead of rats.

This investigation of the role of the experimenter was done with human subjects, and the concept-learning (Bruner, 1956) task selected for humans was intended to parallel roughly the maze-running task for rats. Since sex might be a complicating factor, it was controlled statistically for both subjects and experimenters. An effort will be made to control age, education, socio-economic level, and intelligence in the selection of the subjects.

It is hypothesized that subjects who are expected to be "bright" by their experimenters differ significantly from those expected to be "dull," and that this difference is in the direction of the expectancy of the experimenters.

Chapter IV

The Method

Subjects

The subjects were 64 students from three senior high schools in Houston, Texas suburbs. They were chosen from a population of juniors who had scored between 90 and 110 on the Otis Gamma test, which had been previously administered by their schools and recorded on their permanent record. The schools are located in an area with an upper middle-class population. It should be pointed out that most families in the area had moved to Houston for business reasons, so that they represent many different backgrounds.

The subjects were contacted by mail and asked to volunteer to take part in this experiment. About one-half of the subjects used did volunteer from the public schools.[b] The other half of the subjects were chosen by the principal of a parochial high school as fitting the required qualifications. These students all participated in the experiment, and of course they could not be called volunteers.

Experimenters

The experimenters were four male and four female undergraduate

---

[b]This number represents less than one out of ten contacted.

psychology students, and all were considered "naive" as far as psychological experiments are considered.

Each experimenter tested eight subjects, two "bright" and two "dull" of each sex. The subjects were assigned randomly to the experimenters. Printed instructions were given about test procedure along with a demonstration of the experimenter's expected behavior. They were told the experimental subjects were from two extreme groups of the high schools; the highest accelerated classes, made up of pupils with highly efficient problem-solving abilities, and from the dullest and lowest schedules, composed of students who were deficient in problem-solving abilities. All experimenters were reminded that Houston high schools group students according to their tested abilities and achievements. They were told that the object of the experiment was to investigate the students' approaches to problem-solving in order to find possible reasons why the differences were so large between the two groups being studied. They were cautioned not to let the subjects know the purpose of the experiment nor to which group they belonged; if the subjects asked how they were chosen (and it turned out they nearly always did), the answer was to be that they were representatives of their class. The atmosphere of the testing situation itself was to be informal and permissive, and the experimenters themselves were to be as relaxed and warm as possible. Before the testing began there were

to be a few minutes of general conversation with the subject in the hope of increasing motivation and establishing good rapport.

Procedure

The procedure was essentially that of Bruner's (1962) which is described in his book, A Study of Thinking. Index-size cards were prepared which varied in four attributes: color--red, green, or black middle figure, number of border--0, 1, 2, or 3, shape of the middle figure--cross, square, or circles, and the number of middle figures--1, 2, or 3. The cards were laid out in order and were shown to the subject. After the preliminary remarks the experimenters approached the actual test period with a remark similar to this:

> (Lucy or Tom), you have been chosen to represent your class at _____ High School, and so try to do the best you can. We are going to play a game which is similar to "Twenty Questions," or "What's My Line?"

The experimenters demonstrated at least once what each subject was expected to do. The subject was told he would have to guess what concept the examiner was thinking of, i.e., "all red cards," or "all cards with one cross and three borders," or "something like that." The experimenter then pointed to a card that was a positive example or instance of what the examiner had in mind. The subject was to point to cards, one at a time, from those spread out before him. After each guess or

choice, the examiner let him know whether or not it was a positive instance of the concept (whether it was a card which was included in the concept or not). Whenever the subject felt he thought he knew what it was the examiner was thinking about, he was to guess. All the examiners were given the same three concepts to use for the three times the subject had to guess: (a) all red circles; (b) all cards with 2 squares and 2 borders, and (c) all cards with just one object. However, the examiners were free to pick the first example of each concept, so long as it was recorded. The examiners recorded the amount of time taken to guess each of the three concepts, and, as far as possible, the exact words of the subjects. The experimenters were told to concentrate on the thought processes of the two different groups in order to identify the differences in approaches to problem-solving--in this case, guessing the concept.

A pilot study using this procedure was carried out. The results suggested that the groups might differ significantly in (a) the total number of choices; (b) the number of repeated choices; (c) the number of ways the first choice of the subjects after the positive example given differed from the example itself; (d) the number of incorrect hypotheses guessed; and (e) the length of time taken by subjects.

Experimental Design

The experimental design is a 2x2x2x4 mixed hierarchical analysis

of variance design, A(e/A)BD((s/ABDe) (Winer, 1962). An inspection of the layout for this design, shown in Table 1, shows that the sex variable is completely crossed, and experimenters and subjects are nested under sex.

Table 1

The Layout For

The Mixed Hierarchical Analysis of Variance Design Used

A(e/A)BD (s/ABDe)

			B_1 (Bright)		B_2 (Dull)	
			D_1 (Male)	D_2 (Female)	D_1 (Male)	D_2 (Female)
A_1 (Male)	e_1	s_1				
		s_2				
	e_2	s_1				
		s_2				
	e_3	s_1				
		s_2				
	e_4	s_1				
		s_2				
A_2 (Female)	e_1	s_1				
		s_2				
	e_2	s_1				
		s_2				
	e_3	s_1				
		s_2				
	e_4	s_1				
		s_2				

Note.--Fixed factors are capitalized (B, A, D), and random factors are designated by small letter (e, s). (After Winer, 1962).

Chapter V

Results

Data were analyzed for five different variables: (a) length of time taken by subjects; (b) number of choices made by subjects; (c) number of incorrect hypotheses guessed by subjects; (d) number of repeated choices made by subjects, and (e) number of ways the first choice of the subjects after the positive example given differed from the example itself. The statistical tests of the design are shown in Table 2, and the numerical values for the sources of variances are shown in Tables 3 through 7.

There were no statistically significant differences for any of the variables analyzed. The only F-test which approached significance was that between male and female experimenters on the total length of time taken. This value, 1.301, would have to be 1.53 (6, 32 degrees of freedom) in order to be significant at the .20 level (Lindquist, 1956). The results were all apparently due only to chance.

Inspection of the raw data shows that a great deal of variance was present. In the total number of choices, for instance, one subject gave 111 answers, while another gave none--he just started guessing hypotheses; the number of incorrect hypotheses varied from 11 to 0; the

number of repeated choices varied from 28 to 0, and the amount of time varied from 69" to 3". All of these ranges produced relatively large error terms.

Table 2

Statistical Tests of Design A(e/A)BD(s/ABDe)
(N = 64)

A_1, A_2	Sex of experimenter
B_1, B_2	"Bright" or "Dull"
D_1, D_2	Sex of subject
e_1, e_2, e_3, e_4	Experimenter
s_1, s_2	Subject

Source	Degrees of Freedom	Error Term
A	1	e/A
e/A	6	s/ABDe
B	1	Be/A
AB	1	Be/A
Be/A	6	s/ABDe
D	1	De/A
AD	1	De/A
De/A	6	s/ABDe
BD	1	BDe/A
ABD	1	BDe/A
BDe/A	6	s/ABDe
s/ABDe	32	No error term

Note.--See Winer, 1962.

Table 3

Length of Time Taken by Subjects

(N =64)

Source	Value	Degrees of Freedom	Error Term	F-Test
A	.562	1	2682.438	Not Significant
e/A	2682.438	6	2062.000	1.301 (Not Significant)
B	22.562	1	436.937	Not Significant
AB	16.001	1	436.937	Not Significant
Be/A	436.937	6	2062.000	Not Significant
D	121.000	1	1467.937	Not Significant
AD	264.063	1	1467.937	Not Significant
De/A	1467.937	6	2062.000	Not Significant
BD	60.063	1	157.188	Not Significant
ABD	2.249	1	157.188	Not Significant
BDe/A	157.188	6	2062.000	Not Significant
s/ABDe	2062.000	32	- - - -	

Table 4

Number of Choices Made by Subjects

(N = 64)

Source	Value	Degrees of Freedom	Error Term	F Test
A	144.000	1	3768.438	Not Significant
e/A	3768.438	6	6188.063	Not Significant
B	49.000	1	1695.937	Not Significant
AB	148.563	1	1695.937	Not Significant
Be/A	1695.937	1	6188.063	Not Significant
D	121.000	1	1739.437	Not Significant
AD	76.563	1	1739.437	Not Significant
De/A	1739.437	6	6188.063	Not Significant
BD	175.563	1	3252.375	Not Significant
ABD	430.499	1	3252.375	Not Significant
BDe/A	3252.375	6	6188.063	Not Significant
s/ABDe	6188.063	32	- - - -	

Table 5

Number of Incorrect Hypotheses Guessed

(N = 64)

Source	Value	Degrees of Freedom	Error Term	F Test
A	21.390	1	91.469	Not Significant
e/A	91.469	6	162.500	Not Significant
B	.765	1	48.344	Not Significant
AB	9.766	1	48.344	Not Significant
Be/A	48.344	6	162.500	Not Significant
D	11.390	1	115.219	Not Significant
AD	26.266	1	115.219	Not Significant
De/A	115.219	6	162.500	Not Significant
BD	8.265	1	127.047	Not Significant
ABD	77.859	1	127.047	Not Significant
BDe/A	127.047	6	162.500	Not Significant
s/ABDe	162.500	32	- - - -	

Table 6

Number of Repeated Choices Made by Subjects

(N = 64)

Source	Value	Degrees of Freedom	Error Term	F Test
A	21.390	1	91.469	Not Significant
e/A	91.469	6	162.500	Not Significant
B	.765	1	48.344	Not Significant
AB	9.766	1	48.344	Not Significant
Be/A	48.344	6	162.500	Not Significant
D	11.390	1	115.219	Not Significant
AD	26.266	1	115.219	Not Significant
De/A	115.219	6	162.500	Not Significant
BD	8.265	1	127.047	Not Significant
ABD	77.859	1	127.047	Not Significant
BDe/A	127.047	6	162.500	Not Significant
s/ABDe	162.500	32	- - - -	

Table 7

Number of Ways the First Choice of the Subjects After the Positive Example Given Differed From the Example Itself

(N=64)

Source	Value	Degrees of Freedom	Error Term	F-Test
A	.140	1	8.094	Not Significant
e/A	8.094	6	18.500	Not Significant
B	.140	1	13.984	Not Significant
AB	2.641	1	13.984	Not Significant
Be/A	13.984	6	18.500	Not Significant
D	.765	1	5.219	Not Significant
AD	.766	1	5.219	Not Significant
De/A	5.219	6	18.500	Not Significant
BD	2.641	1	.328	Not Significant
ABD	.016	1	.328	Not Significant
BDe/A	.328	6	18.500	Not Significant
s/ABDe	18.500	32	- - - -	

Chapter VI

Discussion

This discussion takes up the basic question of why experimenter effects show up in other studies and not in this one, and what variables are involved in this difference in results.

In many studies reported there were errors made in the observation, recording, scoring, and interpretation of data. The studies with planaria (Cordaro and Ison, 1963; Rosenthal and Halas, 1962) involve errors in observation; significantly more head turns and body contractions were reported when the experimenter expected that he would get that kind of results from his observations. Rosenthal (1966) concluded that there are situational differences in the extent to which behavior modifications in planaria are observed, and that these particular differences depend upon the specific type of behavior being observed.

In the clinical interaction between patient and psychologist, events occurring are often not observed, or at least not reported, by the psychologist, and events not occurring are sometimes erroneously reported. A case in point is the study by Hyman, et al., (1954) in which counselor's written reports of interviews are compared with electrical transcriptions of the interviews. They found extensive and significant

omission of content in the written record, alteration of the time sequence, and lack of precision in the notes taken. Marked contrast in the descriptions by Margaret Mead and Reo Fortune (Hyman, et al., 1954, Fortune, 1939) of the Arapesh must be attributable to selective choosing of which customs and behavior of the tribe were typical.

Rosenthal (1966) cites study after study in which errors of record and computation were involved and, in his own experiments, he found a regularity of the occurrence of such errors. One example of bias in scoring is found in the scoring on the WISC, the benefit of the doubt usually going in the direction of the set of the experimenter. In a study in which false IQ scores were given in background information Cahen (1965; cited in Rosenthal, 1966, p. 22) found that children thought to be brighter received higher scores than those thought to be less able.

Errors of interpretation are perhaps best illustrated by the significant biases seen in psychological evaluations, including Rorschach protocols. Robinson and Cohen (1954) found significant biases in the psychological reports of 30 patients by three different examiners. Star (1950) reported the magnitude of differential diagnoses of abnormal behavior, noting that during World War II one induction center rejected 100 times more recruits than did another. Lord (1950), Masling (1959) and Gibby, Miller, and Walker (1952), among others, found differences

in Rorschach protocols which contributed to variance in the examiners.

In this study the experimenters did not score, compute, nor interpret their findings. They merely recorded the total time used, the verbatim answers of the subjects, and then submitted their data to the writer. The measures used--time, number of incorrect hypotheses, number of recurring errors, number of ways the first choice after the given example differed from the example itself, and the total number of choices--were all based upon simple counting. Thus there was no opportunity for the examiners to bias the results by scoring, computing, or interpreting their findings, since this was all done by someone else.

In any of the situations described which deal with the experimenter effect, the expectancy of the experimenter about his subjects (in this study, whether the children were bright or dull) has to be applicable to the measures of performance; it has to be communicated somehow to the subject, either verbally or non-verbally; the subject has to be able to interpret this information correctly; and then the subject must want to and be free to alter his behavior in the direction of the expectancy. In the studies of perception of photographs of people who were "successful" or "failures" (Rosenthal, 1966), for example, the subject's awareness of the experimenter's set was applicable to his rating of the photograph,

this set was communicated to the subject either by the appearance of the experimenter, by his perceived status (successful or not), by his verbal utterances and behavioral mannerisms: all of these aspects were shown to be quite variable in the motion pictures taken of the testing situation.

In studies of verbal conditioning, such as those by Greenspoon (1955), Verplanck (1955), and Azrin, et al., (1961), the subject was aware, perhaps subconsciously, of the fact that some of his behavior was pleasing the examiner and some was not; he altered his answers to fit in with his perceived idea of what the experimenter wanted--what he was reinforcing. Kanfer (1958) reinforced verb responses with a flashing light, and found significant individual differences in the alteration of behavior. Binder, McConnell, and Sjoholm (1957) found significantly differences rates of learning obtained by male and female experimenters, when they reinforced a hostile verb by saying, "Good." Sarason and Minard (1963) experimentally manipulated the degree of personal contact between subject and experimenter, as well as the perceived prestige, they found significant differences in performance when subjects were reinforced for emission of first person pronouns. These studies all provide the subject with the freedom to alter his performance in accordance with his perception of what the experimenter

wants. In the present study there was no opportunity for a shift in approach to the concept learning problem, nor could the subject perform more efficiently, even if he could have perceived the expectancy of the experimenter.

Modeling effects are aptly illustrated by the answers obtained by Rice (1929) and Clark (1927) which were like those the examiner himself would have given. Rice's subjects reported to their "teetotaler" examiners that alcohol was the "cause of their downfall." Clark's athletic examiner obtained reports of greater amounts of time spent by his subjects in athletic events than was reported by subjects interviewed by a less athletic-looking person. Modeling effects were not apparent in this present study, because the subjects knew what the experimenters wanted them to do, but were unable to alter their behavior to correspond with the experimenters' expectation of their performance.

A closer look at the way the wishes of the experimenter are communicated to the subjects is needed. In verbal conditioning studies it is the reenforcement which communicates the expectancy, in studies involving non-verbal communication it is some behavioral gesture that communicates this to the subjects. In studies involving animals, e.g., rabbits (Brogden, 1962) and rats (Rosenthal and Fode, 1963; Rosenthal and Lawson, 1964), there may have been some difference in the way the animals

were treated--either in the amount of time spent with the animals or in the way they were handled. If they were handled roughly, the animals might have become anxious, and their performance might have deteriorated, if they were handled gently and often, the animals might have been at their best when the experimental trials were run. Time spent with the animals might itself serve to allay anxiety and improve performance. In the present study there was no difference in the amount of time spent by the examiners with the subjects, and there was no personal contact between them. Rosenthal and his group at Harvard are now engaged in a series of experiments trying to pin down the cues which are picked up by the subject--verbal or non-verbal. Among these are studies in which all auditory, or visual, stimuli are blocked. His results are not yet in print.

In future research on this subject an opportunity should be provided for the experimenter to score, record, compute, and interpret his data; mannerisms or verbal communication should be used to communicate the expectancy of the examiner to the subject; the task chosen should be one in which the subject can alter his behavior in an effort to please the examiner, if in fact he does realize what the examiner expects him to do.

As for the control of experimenter effects in future research, it is suggested that an awareness of the importance of these effects may become an important factor in the design of experiments. There is also

a need to replicate these experiments using, of course, different experimenters. These will hopefully lead to more objective results and better designed experiments in psychological research.

Chapter VII

Summary

Since the Renaissance scientific experimentation has been plagued by the experimenter's effect upon the observation, collection, and reporting of data. Only recently, however, has this variable been recognized as one that can be investigated. Psychological experiments with animal and human subjects in clinical, experimental, social, survey, and educational research have shown that experimenters can and do influence their data. Specific factors which have been investigated include experimenter's status, prior experience, previous interaction with subject, modeling effects, early data returns effects, sex, and expectancy. This experiment was designed to investigate the effect of experimenter's expectancy. The experimenter was told he was dealing with two different groups--a bright one and a dull one--and that these two groups would perform very differently in a concept learning task. A 2x2x2x4 mixed hierarchical analysis of variance design was used, with eight experimenters testing 64 subjects. No significant effects were found. A discussion follows, and the differences between this study and others which found significant differences in their results are explored. In future experiments it is suggested that the experimenter be given an opportunity to score, summarize, and

interpret his data, and that the subject be able to make a subjective judgment of his behavior in relation to his perceived idea of whether he is pleasing the experimenter, and, if not, an opportunity should be provided for him to be able to alter his own behavior in keeping with the set or expectancy of the experimenter.

References

Azrin, N. H., Holz, W., Ulrich, R., & Goldiamond, I. The control of the content of conversation through reinforcement. J. exp. anal. Behav., 1961, 4, 458-468.

Bean, W. B. An analysis of subjectivity in the clinical examination in nutrition. J. appl. Physiol., 1948, 1, 458-468.

Bellak, L. On the problem of the concept of projection: a theory of apperceptive distortion. In L. E. Abt & L. Bellak (Eds.), Projective psychology. New York: Grove, 1959. Pp. 7-32.

Berkowitz, H. Effects of prior experimenter-subject relationships on reinforced reaction time of schizophrenics and normals. J. abnorm. soc. Psychol., 1964, 69, 522-530.

Berkson, J. Magath, T. B., & Hurn, Margaret. The error of estimate of the blood cell count as made with the hemocytometer. Amer. J. Physiol., 1940, 128, 309-323.

Binder, A., McConnell, D., & Sjoholm, Nancy A. Verbal conditioning as a function of experimenter characteristics. J. abnorm. soc. Psychol., 1957, 55, 309-314.

Birney, R. C. The achievement motive and task performance: A replication. J. abnorm. soc. Psychol., 1958, 56, 133-135.

Blankenship, A. B. The effect of the interviewer upon the response in a public opinion poll. J. consult. Psychol., 1940, 4, 134-136.

Blaufarb, H. The relation of experimenter status and achievement imagery and the conditioning of verbal behavior. Dissert Abstr., 1961, 29, 2793 (Abstract)

Boring, E. G. A history of experimental psychology. (2nd ed.) New York: Appleton-Century-Crofts, 1950.

Boring, E. G. Newton and the spectral lines. Science, 1962, 136, 600-601.

Braunstein, D. N., Braunstein, Harriet M., & Blumenfeld, W. S. Performance in training and an achievement effort rating. Psychol. Rep., 1965, 16, 1077-1080.

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

Bruner, J., Goodnow, J. J., & Austin, G. A. A study of thinking. New York: Wiley, 1956.

Cahen, L. S. An experimental manipulation of the "Halo Effect"; A study of teacher bias. Unpublished manuscript, Stanford Univer., 1965. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966. P. 22.

Cantril, H., & Research Associates. Gauging public opinion. Princeton: Princeton Univer. Press, 1944.

Carson, R. C., & Heine, R. W. Similarity and success in therapeutic dyads. J. consult. Psychol., 1962, 26, 28-43.

Clark, E. L. The value of student interviewers. J. Pers. Res., 1927, 5, 204-207. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966. P. 113.

Cordaro, L., & Ison, J. R. Observer bias in classical conditioning of the planarian. Psychol. Rep., 1963, 13, 787-789.

Dohrenwend, Barbara S., & Richardson, S. A. Analysis of the interviewer's behavior. H. Org., 1956, 15, 29-32.

Ebbinghaus, H. An early approach to the study of memory. In R. L. Wrenn (Ed.), Basic contributions to psychology: readings. Belmont, Calif.: Wadsworth, 1966. Pp. 181-186. Abridged from H. Ebbinghaus, Memory, a contribution to experimental psychology. Trans. Henry Ruger. New York: Bureau of Publications, Teachers College, Columbia Univer., 1913. Pp. 22-33.

Fisher, R. A. Has Mendel's work been rediscovered? Ann. Sci., 1936, 1, 115-137.

Fortune, R. Arapesh warfare. Amer. Anthrop., 1939, 14, 22-41.

Frankel, R., & Williams, R. A method for measuring error due to variance among interviewers. Publ. Opin. quart., 1956, 20, 507-592.

Friedman, N., Kurland, D., & Rosenthal, R. Experimenter behavior

as an unintended determinant of experimental results. J. Proj. Tech. pers. Assess., 1965, 29, 479-490.

Gibby, R. G., Miller, D. R., & Walker, E. L. Examiner variance in the Rorschach protocols of neuropsychiatric patients. Amer. Psychologist, 1952, 7, 338-339. (Abstract)

Goldstein, A. P. Therapist-patient expectancies in psychotherapy. New York: Pergamon Press, 1962. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York Appleton-Century-Crofts, 1966. P. 153.

Gore, Pearl M. Individual differences in the prediction of subject compliance to experimenter bias. Dissertation Abstr., 1963, 24, 390-391. (Abstract)

Graham, S. R. The influence of therapist character structure upon Rorschach changes in the course of psychotherapy. Amer. Psychologist, 1960, 15, 415. (Abstract)

Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. Amer. J. Psychol., 1955, 68, 409-416.

Gruenberg, B. C. The story of evolution. Princeton: Van Nostrand, 1929.

Hammer, E. F., & Piotrowski, Z. Hostility as a factor in the clinician's personality as it affects his interpretation of projective drawings (H-T-P). J. proj. Tech., 1965, 29, 210-216.

Harris, H. E., Piccolino, E. B., Roback, H. B., and Sommer, D. K. The effects of alcohol on counterconditioning of an avoidance response. Quart. J. Alcoholic Stud., 1964, 25, 490-497.

Holtzman, W. H. The examiner as a variable in the Draw-a-Person test. J. consult. Psychol., 1952, 16, 145-148.

Hunter, Mary. Responses of comparable white and negro adults to the Rorschach test. J. Psychol., 1937, 3, 173-182.

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

Johnson, M. L. Seeing's believing. New Biology, 1953, 15, 60-80. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966. P. 6.

Kanfer, F. H. Verbal conditioning. reinforcement schedules and experimenter influence. Psychol. Rep., 1958, 4, 443-452.

Kanfer, F. H., & Karas, Shirley C. Prior experimenter-subject interaction and verbal conditioning. Psychol. Rep., 1959, 5, 345-353.

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

Lane, F. W. Kingdom of the octopus. New York. Sheridan House, 1960. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966. P. 6.

Lichtenstein, E. Personality similarity and therapeutic success: a failure to replicate. J. consult. Psychol., 1966, 30, 282.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston Houghton Mifflin, 1953.

Lord, Edith. Experimentally induced variations in Rorschach performance. Psychol. Monogr., 1950, 64, No. 10 (Whole No. 316).

Maccoby, Eleanor E., & Maccoby, N. The interview: A tool of social science. In G. Lindzey (Ed.), Handbook of social psychology. Vol. I. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 449-487.

Marine, Edith L. The effect of familiarity with the examiner upon Stanford-Binet test performance. Teach. Coll. contr. Educ., 1929, 381, 42. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York Appleton-Century-Crofts, 1966. P. 88.

Masling, J. The effects of warm and cold interaction on the administration and scoring of an intelligence test. J. consult. Psychol., 1959, 23, 336-341.

Masling, J. The influence of situational and interpersonal variables in projective testing. Psychol. Bull., 1960, 57, 65-85.

Mayhew, L. B. Personality and teaching. J. Communication, 1957, 7, 83-89. (Psychol. Abstr., 1957.6005)

McFall, R. F. "Unintentional communication": The effect of congruence and incongruence between subject and experimenter constructions. Unpublished doctoral dissertation, Ohio State Univer., 1965. Cited by Rosenthal, Experimenter Effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

McGuigan, F. J. Variation of whole-part methods of learning. J. educ. Psychol., 1960, 51, 213-216.

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

Medley, D. M., & Klein, A. A. Studies of teacher behavior: inferring classroom behavior from pupil response. New York: Board of Higher Education of City of New York, Division of Teacher Education, Office of Research and Evaluation, 1956. (Psychol. Abstr., 1957: 1791)

Miller, D. R. Prediction of behavior by means of the Rorschach test. J. abnorm. soc. Psychol., 1953, 48, 367-375.

Murstein, B. I., & Easter, L. V. The role of achievement, anxiety, stimulus, and expectancy, on achievement motivation in arithmetic and Thematic tests. J. proj. Tech. pers. Assess., 1965, 29, 491-499.

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

Pearson, K. On the mathematical theory of error of judgment with special reference to the personal equation. Phil. Trans. Roy. Soc. London, 1902, 198, 235-299.

Pfungst, O. Clever Hans (the horse of Mr. Von Osten). (Orig. Publ., 1905, trans. by C. L. Rahn) New York: Holt, Rinehart & Winston, 1965.

Polanyi, M. Personal knowledge. Chicago: Univer. of Chicago Press, 1958.

Postman, L., & Jarrett, R. F. Learning without awarenes . Amer. J. Psychol., 1952, 65, 244-255.

Rankin, R. & Campbell, D. Galvanic skin response to Negro and white experimenters. J. abnorm. soc. Psychol., 1955, 51, 30-33.

Rice, S. A. Contagious bias in the interview: A methodological note. Amer. J. Sociol., 1929, 35, 420-423.

Robinson, J. T., & Cohen, L. D. Individual bias in psychological reports. J. clin. Psychol., 1954, 10, 333-336.

Roe, Ann. The psychology of the scientist. Science, 1961, 134, 456-459.

Rosenthal, R. Experimenter attributes as determinants of subjects' responses. J. proj. Tech. pers. Assess, 1963, 27, 324-331. (a)

Rosenthal, R. Experimenter modeling effects as determinants of subject's responses. J. proj. Tech. pers. Assess., 1963, 27, 467-471. (b)

Rosenthal, R. Clever Hans: A case study of scientific method. Introduction to Pfungst, O., Clever Hans. New York: Holt, Rinehart & Winston, 1965. Pp. ix-xlii.

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

Rosenthal, R., & Fode, K. L. The effect of experimenter bias on the performance of the albino rat. Behav. Sci., 1963, 8, 183-189.

Rosenthal, R., Fode, K. L., Friedman, C. J., & Vikan-Kline, Linda. Subjects' perception of their experimenter under conditions of experimenter bias. Percept. mot. Skills, 1960, 11, 325-331.

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

Rosenthal, R., & Lawson, R. A longitudinal study of the effects of experimenter bias on the operant learning of laboratory rats. J. Psychiat. Res., 1964, 2, 61-72.

Rosenthal, R., Persinger, G. W., & Fode, K. L. Experimenter bias,

anxiety, and social desirability. Percept. mot. Skills, 1962, 15, 73-74.

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

Rostand, J. Error and deception in science. New York. Basic Books, 1960.

Russell, B. Philosophy. New York Norton, 1927. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York Appleton-Century-Crofts, 1966. P. 179.

Sacks, Elinor L. Intelligence scores as a function of experimentally established social relationships between child and examiner. J. abnorm. soc. Psychol., 1952, 47, 354-358.

Sanders, R., & Cleveland, S. E. The relationship between certain examiner personality variables and subject's Rorschach scores. J. proj. Tech., 1953, 17, 34-50.

Sarason, I. G., & Minard, J. Interrelationships among subject, experimenter, and situational variables. J. abnorm. soc. Psychol., 1963, 67, 87-91.

Schactel, E. G. Subject definitions of the Rorschach test situation and their effect on test performance. Psychiat., 1945, 5, 419-448.

Schafer, R. Psychoanalytic interpretation in Rorschach testing: theory and application. New York: Grune & Stratton, 1954.

Severin, F. T., & Rigby, Marilyn K. Influence of digit grouping on memory for telephone numbers. J. appl. Psychol., 1963, 47, 117-119.

Spires, A. M. Subject-experimenter interaction in verbal conditioning. Unpublished doctoral dissertation, New York Univer., 1960. Cited by Rosenthal, R., Experiment effects in behavioral research. New York Appleton-Century-Crofts, 1966. P. 82.

Star, Shirley A. The screening of psychoneurotics: comparison of psychiatric diagnoses and test scores at all induction stations. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarfeld,

Star, Shirley, A. & Clausen, J. A. Measurement and prediction. Princeton. Princeton Univer. Press, 1950. Pp. 548-567.

Stevenson, H. W. & Allen, Sara. Adult performance as a function of sex of experimenter and sex of subject. J. abnorm. soc. Psychol., 1964, 68, 214-216.

Stevenson, H. W., Keen, Rachel, & Knights, R. M. Parents and strangers as reinforcing agents for children's performance. J. abnorm. soc. Psychol., 1963, 67, 183-186.

Sullivan, H. S. A note on the implications of psychiatry, the study of interpersonal relations, for investigations in the social sciences. Amer. J. Sociol., 1936-37, 42, 848-861.

Verplanck, W. S. The control of the content of conversation: Reinforcement of statements of opinion. J. abnorm. soc. Psychol., 1955, 51, 668-676.

Vikan-Kline, Linda L. The effect of an experimenter's perceived status on the mediation of experimenter bias. Unpublished master's thesis, Univer. of North Dakota, 1962. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966.

Weitz, J. Let's do it again. Psychol. Rep., 1956, 2, 391.

Wickes, T. H. Examiner difference in a test situation. J. consult. Psychol., 1956, 20, 23-24.

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

Wirth, L. Preface to K. Mannheim, Ideology and utopia. New York: Harcourt, Brace & World, 1936. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966. P. 7.

Wolf, Theta H. An individual who made a difference. Amer. Psychologist, 1961, 16, 245-248.

Young, R. K. Digit span as a function of the personality of the experimenter. Amer. Psychologist, 1959, 14, 375 (Abstract)

Yule, G. U. On reading a scale. J. Roy. statist. Soc., London, 1927, 90, 570-587.