DOCUMENT RESUME

ED 101 009 TM 004 136

AUTHOR Dolly, John P.; Tillman, Murray H.
TITLE An Overview of Research Artifacts Supposedly Causing Data Bias.
PUB DATE [Apr 74]
NOTE 52p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, Illinois, April, 1974)

EDRS PRICE MF-$0.75 HC-$3.15 PLUS POSTAGE
DESCRIPTORS *Bias; *Design Needs; Ethics; Experiments; Literature Reviews; Participant Characteristics; *Research Design; Researchers; Research Methodology; *Research Problems

ABSTRACT

An attempt has been made to show how the problem of data bias had been approached, conceptualized and studied. A review of some of the major studies in the area of research artifacts is presented. Following each major topic, suggestions for future research are discussed. The difficulty in complying with the current APA guidelines on ethics is also examined. Some of the research on experimenter expectancy, experimenter bias, demand characteristics and evaluation apprehension are presented and discussed as causes of data bias. (Author)

An Overview of Research Artifacts Supposedly Causing Data Bias

By

John P. Dolly
University of South Carolina

and

Murray H. Tillman
University of Georgia

A paper presented at the annual meeting of the
American Educational Research Association
Chicago, 1974

TABLE OF CONTENTS

INTRODUCTION

Recent research has questioned the degree and type of bias entering into much of the research being conducted, especially in the areas of personality and social psychology (Gephart & Antonopolis, 1969; Rosenthal, 1966; Rosenthal & Rosnow, 1969; Weber & Cook, 1972). At the same time, Argyris (1968) has pointed out how current research, especially in organizational settings, has neglected the problems created by the distortion of data by subjects (Ss). Argyris (1968) suggested that Ss in a research study, as employees in an industrial setting have a tendency to adopt roles, either positive or negative, which have a marked effect on any research data being gathered concerning them. Weber and Cook (1972), Ausubel (1968) and Bugelski (1971) have all expressed much concern about the validity of inferences about human behavior that can be drawn from laboratory experiments. There is agreement among these researchers that the generalizability of laboratory research to situations outside this artificial setting are questionable. Bugelski (1971) has pointed out how psychologists have invented tasks unfamiliar to research Ss that are usually short, meaningless and isolated. He questions this whole approach to research, especially the contribution of such research to the classroom.

Gephart and Antonopolis (1969) point out that there are at least five different kinds of bias which may effect any research study: 1) experimenter bias effect; 2) Hawthorne effect; 3) demand characteristics;

4) halo effect; and 5) placebo effect. Not only is there research to show that Ss want to please an experimenter (E) when they perceive being in an inferior position (Orne, 1962) or when they feel threatened (Argyris, 1968) but there is also some evidence that under certain conditions negativistic Ss may purposely decide to distort the data. Cook, Bean, Calder, Frey, Krovetz and Reisman (1970) found that the negativistic S may try to distort the data by corroborating some hypothesis other than the E's or by giving useless responses to the E. This has been called the "recalcitrant subject" by Fillenbaum and Frey (1970) and the effects resulting from this orientation has been referred to as the "screw-you effect" (Masling, 1966) and the "boomerang effect" (Silverman, 1965).

Another consideration which has received much comment recently has been a new emphasis on ethical considerations when using human subjects (APA, 1972). One position frequently advocated has been that Es should obtain the permission of the Ss involved and when possible inform the S of the purpose of the research prior to conducting it. Such actions may have some far reaching effects on research data obtained. In most of the research on bias effects Ss have never been told the purpose or intent of the research. In many of these studies the Ss was unaware he was participating in a research study. Such deceptions in future research may be ruled unethical. This would place serious limitations on the E conducting research in the area of bias effects. For example, Weber and Cook (1972) point out in six studies where Ss were informed of the research hypothesis, such knowledge typically caused bias. In a more recent study by Resnick and Schwartz (1973), a group of Ss treated in accordance with the proposed APA ethical guidelines, gave results that were significantly different from

a group of Ss who were not briefed about the nature of the study as the guidelines require.

As Weber and Cook (1972) indicate, the majority of the research being done on research bias has been conducted in social and personality psychology. Little if any research has been done outside these areas although conclusions drawn from some of this research has been generalized to the classroom (Grieger, 1970; Rosenthal & Rosnow, 1969). Even in this research most of it is limited to attitudes and attitude change. Although the bias problem in research has been studied recently, only the articles by Argyris (1968), Weber and Cook (1972) and the text by Rosenthal and Rosnow (1969) give a comprehensive view of research done.

There are some crucial issues raised by the research on bias effects which have not been adequately considered in classroom or instructional research. In a very brief review of some of the articles dealing with classroom and instructional problems in the Journal of Educational Psychology from February, 1971, to February, 1973, there is almost a total disregard for how the research was conducted, the type of Ss used, the materials used and the setting of the research (Table 1). Yet most authors draw conclusions and generalizations to the classroom. Below are the 23 studies reviewed. Only five of the researchers conducted their research in the classroom using relevant educational materials. The studies were under the direction of the classroom teacher and apparently in all five of these studies the Ss were unaware of the research being conducted (Cobb, 1972; Domino, 1971; Hammer, 1972; Holmes, 1971; Rhodes, 1971).

TABLE 1

Review of Articles Attempting to Generalize Findings to Classroom and Instructional Settings from the Journal of Educational Psychology February 71 to February 73

Authors	Subjects	Experimenter Perceived As	Materials Used & Setting
Anderson et al. (1972)	ed. psych. volunteers	experimenter	non-relevant in an artificial setting
Cobb (1972)	classroom students	classroom teacher	relevant in a naturalistic setting
Deno et al. (1971)	individual volunteers	experimenter	non-relevant in an artificial setting
Domino (1971)	classroom students	classroom teacher	relevant in a naturalistic setting
Friedman et al. (1972)	intro psych students	experimenter	non-relevant in an artificial setting
Goldman (1972)	psych students	not specified	relevant material setting not specified
Hammer (1972)	classroom students	classroom teacher	relevant in a naturalistic setting
Hicks et al. (1971)	first & second grade children	experimenter	non-relevant in an artificial setting
Holmes (1972)	classroom students	classroom teacher	classroom material and tests in a naturalistic setting
Jenkins et al. (1971)	individual volunteers	experimenter	relevant materials in an artificial setting

Table 1
(Continued)

Authors	Subjects	Experimenter Perceived As	Materials Used & Setting
Kulhavy et al. (1972)	individual volunteers	experimenter	academic may have been relevant to some artificial setting
Levin et al. (1971)	sixth grade students	experimenter	paired associate task in an artificial setting
Mayer et al. (1972)	psych students	experimenter	math concepts in an artificial setting
Quirk et al. (1971)	classroom students	classroom teacher	classroom material in a special program
Rhodes (1971)	classroom students	classroom teacher	classroom material with standardized test, pre-test may have biased the naturalistic setting
Rosenfeld (1972)	classroom students	not clearly specified	relevant math material naturalistic setting
Rothkopf et al. (1972)	industrial trainees	not specified apparently as an instructor	relevant materials in a naturalistic setting
Rychlak et al. (1971)	individual volunteers	experimenter	non-relevant material in an artifical setting
Scandura et al. (1971)	individual students	experimenter	non-relevant material in an artificial setting
Stevenson et al. (1971)	day care center volunteers	not specified	paired associate task in an artificial setting

Table 1
(Continued)

Authors	Subjects	Experimenter Perceived As	Materials Used & Setting
Tallmadge et al. (1971)	classroom students	experimenter	relevant materials in a naturalistic setting
Todd et al. (1971)	psych. students	experimenter	English short stories in an artificial setting
Venezky et al. (1973)	grade school students	not specified	relevant material in an artificial setting

These five studies were conducted in a setting where generalizations could be made to similar situations. A study using industrial trainees (Rothkopf & Kaplan, 1972) used relevant industrial training materials in a training situation where Ss were unaware they were in a research study. In this study although the setting was industrial, the use of relevant materials in a training situation seems to allow for greater generalization of the results obtained to similar industrial settings. In the remaining 17 studies there is at least one major deficit and in most of these studies there is more than one. All these studies leave serious doubts as to the generalizability of the results they reported. In one study where teachers conducted the research using relevant materials the teachers spent so much time teaching the task to the experimental group that the results can hardly be meaningful (Quirk, Steen, & Lipe, 1971). Rosenfeld (1972) used sixty grade school children in a reinforcement experiment. He used relevant materials in a classroom setting but neglected to specify whether or not the researcher was perceived as a teacher or E. In 12 of the studies the individual conducting the research was presented to the Ss as an E, the Ss were tested individually, many of them volunteers or selected out of S pools. The materials used ranged all the way from relevant classroom materials (Goldman, 1972; Jenkins & Deno, 1971; Rosenfeld, 1972) to paired-associate trigrams (Rychlak & Tobin, 1971). In the other eight studies the situation was an artificial experimental one, the materials irrelevant and the Ss all tested individually. However, all these studies attempt to generalize their results to the classroom (Anderson, Kulhavy & Andre, 1972; Deno, Jenkins & Marsey, 1971; Friedman & Greitzer, 1972; Hicks & Packwood, 1971; Kulhavy & Parsons, 1972; Levin & Horvitz, 1971; Mayer & Greeno, 1972; Scandura & Vorhies, 1971).

In the remaining studies there were also procedural problems. In one study, the material used was supposedly meaningful English short stories but the Ss were psychology students and the E was seen as a researcher (Todd & Kessler, 1971). A study by Tallmadge and Shearer (1971) used a classroom in which the Ss apparently did not know they were in a research study, and the material presented was classroom related and rote. This research was conducted by an E who administered psychological tests to the group. Venezky and Johnson (1973) used grade school children as Ss in an experiment on learning four letter-sound generalizations. They neglected to specify how the researcher was perceived and although they used relevant materials they dealt with the students on a one-to-one basis. In the final article generalizing to classroom situations, the authors did not state how the researcher was perceived, used lower socio-economic children from day care centers where children were approached individually and asked to participate. Each child was then tested individually. The students who were unfamiliar with the names of the animals in the paired associate task had difficulty completing this task, as was pointed out by the author (Stevenson, Williams, and Coleman, 1971).

Another significant question to consider is the effect on future research in the area of classroom learning. If it is true that an E or the experimental situation creates an atmosphere of artificiality and negativism in the subjects, then perhaps research in the future should be left up to the Ss' teachers to carry out in naturalistic settings with little or no artificial manipulation of the Ss involved. If the APA ethical guidelines should be adopted then researchers in the future may have to depend a great deal on readily available data such as school records and classroom performances.

Authors' Assumptions

The authors of this paper assume that there are various artifacts that will bias the data in an experiment. Historically, researchers have looked at this problem in settings other than the classroom. Most of the studies quoted have been conducted by social and personality psychologists trying to define various constructs. The rest of the research has been conducted in industrial or military settings. In these studies the researchers often use the constructs being studied, to explain their results or to account for the data obtained. Constructs such as E bias or demand characteristics are used without adequate operational definitions in explaining much of the data reported. At the present time there is a large body of information referring to various bias effects entering into research settings. Most of this research has concerned itself with the E and the social setting in which the interaction between E and S takes place. There have been very few studies dealing solely with the S and S bias. Although researchers, such as Orne (1969) and Rosenberg (1969) address themselves to the role of the S, their research is not as thorough or conclusive compared to the studies on E bias.

One of the problems in looking at the research is the lack of agreement on what causes the bias. Some researchers such as Rosenthal and Rosnow (1969) seem to imply that most of the artifacts biasing research data are extraneous variables that should be controlled or eliminated. Others such as Orne (1962) seem to think that some internal psychological motive on the part of the S is responsible. There are still others such as Rosenberg (1969) who seem to indicate the setting of the research is the major cause of data bias.

This paper discusses 1) the problem of generalizing results from the lab to the classroom, 2) the three major categories of bias, experimenter bias, demand characteristics and evaluation apprehension. These major categories of bias are followed by a discussion of various factors mediating bias, S as the mediating variable, and research settings and E characteristics. A brief section on design problems and a section on ethics is presented followed by a summary statement.

The Problem of Generalization

Bugelski (1971) pointed out the obvious differences in research done in the laboratory and that done in the classroom. Bugelski says that the typical psychologist

> "tries to devise a task that is completely unfamiliar to the learner . . . Not only are the tasks short, meaningless, and isolated, but also the degree of learning is not usually high (p. 23)."

In contrasting the above with human subjects, he went on to say:

> "For humans, the tasks have usually been divorced from meanings--the subjects memorize series of nonsense syllables, numbers or bits of poetry. The most popular task of all in the laboratory has been the learning of a group of pairs of syllables like KAL-MES, where the first syllable is supposed to represent a "stimulus" and the second a "response" (p. 23)."

Bugelski (1971) went on to point out how this use of completely nonrelevant material is one of the most obvious deficiencies in current learning research. As Bloom (1968) pointed out, education is supposed to be meaningful and have a purpose. If this is the case, then there is a large body of knowledge being collected in the literature that meets all the criteria of good research except for its total lack of a relevant task.

There appears to be a large body of material dealing with Ss in artificial situations learning meaningless material upon which researchers try to find solutions to instructional problems.

Ausubel (1968) also felt that much of the research conducted especially in the area of educational psychology has little value or use for generalizing to classroom data of a meaningful nature:

> " . . . methods of learning that children employ in rotely mastering lists of nonsense syllables in the laboratory do not necessarily correspond to methods of learning children use in classrooms to acquire a meaningful grasp of subject matter (p. 17)."

Earlier we made specific reference to a major research problem, the lack of naturalistic settings for research. This problem compounds the difficulty of trying to isolate variables contributing to various forms of research bias. In much of the research to be quoted, the problem of the setting in which the research has taken place has not been properly considered. Only one researcher, Argyris (1968), who looked at research bias in industrial settings, recognized the importance of the research context. Although the industrial setting appears to be a perfect naturalistic setting in which to operate, the Es conducting the research, quoted by Argyris (1968), managed to destroy whatever naturalistic setting they had by getting supervisors and other management level personnel to direct employees into a research setting without adequate briefings or information on the purpose of the data collection.

On the other hand, in attempting to isolate variables which distort research findings it may be necessary to first isolate them in a highly structured artificial situation. Once these variables are identified then

they can be observed in a variety of settings to determine their biasing effects. This seems to be the approach taken by most of the researchers to be mentioned. They are still in the process of trying to isolate the various artifacts causing research bias. Some researchers place the emphasis on the E assuming that if data bias occurs it is due to the influence of the E in the study. The E somehow conveys his research expectations to the Ss who usually comply with these expectations. Other researchers emphasize the role of the S. They point out that the S often has preconceived notions about psychological research. The S tends to respond in the research in accordance with these preconceptions and the data provided tends to be biased.

Experimentar Bias

There have been several researchers who have concerned themselves with various forms of data bias introduced in experimental research. The most important individual appears to be Robert Rosenthal. He has generated most of the relevant research and has caused the greatest amount of controversy concerning various research artifacts. According to Rosenthal (1966), a distinction can be made between observer errors and observer bias. When the observations are non-randomly distributed around a true value, they are referred to as biased observations. When interpretations do not vary randomly, they are referred to as experimenter bias. Rosenthal in much of his research concerns himself with the E as the cause of data bias (1966). In the more recent book by Rosenthal and Rosnow (1969) the E as the cause of bias is still emphasized but other factors mediating bias effects such as the S, the setting, the design of the study, etc., are also looked at as possible unwanted research artifacts.

In Rosenthal's 1966 review of E effects, the most frequently employed task was the person perception task. Photos are rated for the degree of success or failure the person pictured has been experiencing. No naturalistic settings have been provided, usually, the S knew he was in an experiment. In 1969 Rosenthal looked at 64 studies on the E bias effect in which the person perception task had been performed. A total of 33 studies showed inconclusive results; 23 studies showed positive results and eight of the studies showed negative results. Below is a brief description of the "typical" experimental situation using the person perception task. In these studies not only did the major researcher look at the Ss but he also manipulated the Es involved to observe the effect of this manipulation on the data generated by the Ss.

Typically in the person perception task the Ss are given a series of photographs of faces supposedly neutral in appearance, and are asked to rate the photos as experiencing "success" or "failure." The ratings are placed on a scale from (-10) extreme failure to (+10) extreme success. The researcher actually running the study usually tells the Es involved that their Ss should average about -5 or +5 ratings. This is reinforced by telling them (Es) that the results are "well established," the reason for the research is to duplicate these experimental findings or that the Es will be paid a certain sum of money depending on the results they (Es) obtain. Examples of this research are presented below.

The following studies all used this paradigm in obtaining data. Ekman and Friesen (1960) employed two military Es to administer a photo judging task to army recruits. Sometimes the Es were presented to the Ss as officers, sometimes as enlisted men. Sometimes Es reinforced Ss for

liking the persons pictured in the photos and sometimes for disliking them. The overall results suggested that the officer E was more effective at increasing Ss' rate of disliking photographs, whereas the enlisted man E was more effective at increasing Ss' rate of liking photographs. Similar results were found by Rosenthal (1966). He found that the more dominant E drew more failure ratings from the Ss. He also found that lower status Es tended to obtain more favorable reactions to the photos. A problem that can be noted here is the interpretation given to data by Rosenthal and others. They tend to describe their data using terms such as liking, dominant, lower status, etc. The researchers seem to make the inference that giving a photo a rating of +5 or higher indicates some inferred behavioral state referred to as liking. Although these types of inferences are common throughout the research quoted they are not always tied to overt behavior nor are these constructs often operationally defined.

The following two studies used the person perception paradigm and both found data that appears to support the E bias effect.

Using the person perception task, Rosenthal and Fode (1963a) used 10 students in a course on experimental psychology as Es. Each of these student Es administered the task to approximately 20 undergraduate Ss. Each E was given the same instructions and was told not to deviate from them. Some of the Es were then led to expect high ratings from the Ss while other Es were led to expect low ratings. The results supported the expectations the Es were given.

In a study by Laszlo and Rosenthal (1967) three law students were given the role of E. Each of the Es was led to expect low ratings from some of the Ss and high ratings from the remainder of the Ss. The results

indicated a mean rating of -.50 when Es expected positive ratings and a mean rating of -.99 when E expected negative ratings. The difference between the two is significant and supposedly the difference is attributed to E bias.

There have also been studies using similar procedures, that have not shown the E effect. Using 18 student Es White (1962) gave the person perception task to 108 student Ss. The Es were placed in one of six groups. Each group was given information that the Ss would give one of six ratings (-6, -3, -.5, +.5, +3, or +6) on the task. The findings indicated that the Es who were given information that the ratings should be high did not differ significantly from the Es given the low rating expectancy. Es who were given a low rating expectancy obtained results opposite to their expectancy. The overall results were not as predicted for all the groups, they were in a slightly negative direction.

Rosenthal and Fode (1963a) using the same person perception task attempted to see the effect a monetary reward would have on expectancy data. Using 12 Es they told them that their Ss should give high positive ratings to the photos. The researchers also told half the Es they would get five dollars per hour if the results were as expected. The other half of the Es were given the same information and told they would receive two dollars per hour if the results turned out as predicted. The Es were also divided into two other groups with half receiving 50 cents per hour and the other half being unpaid for simply performing the task. The results indicated no significant differences among all the groups. There was no more bias found in any one of the groups than the others.

Some researchers have been very critical of Rosenthal and his methodology. Weber and Cook (1972) suggest, for example, that much of his

research involved a single experimental procedure (the person perception task) that permitted no generalization to other procedures. They also point out that much of the research is only marginally significant. Rosenthal (1966) in reporting his data often goes back to perform post hoc analysis to obtain significant results. When this fails if he repeats the study several times and obtains similar data he reports this as a significant trend. Rosenthal (1966) then refers to these significant trends as if he had obtained significance (.05 or .01) on the original data. Barber, Forgione, Chaves, Calverley, McPeake and Bowen (1969) attempted to show the E bias effect in five investigations. They duplicated the experimental procedures used by Rosenthal and Fode (1963a) and incorporated more stringent controls. In all five cases they failed to demonstrate the effect. Not only did they fail to find bias but the data did not even show a trend in support to an E bias effect. These investigators also take issue with the methodology employed and the statistical manipulation of the data. They point out how in Rosenthal and Fode's (1963a) study, they allowed the use of Ss familiar to the E on obvious form of E bias. They too take issue with the use of post hoc analysis to obtain significant results. Greiger (1970) was also unable to replicate many of the findings presented by Rosenthal and his associates.

Barber and Silver (1968) have cast some doubt on the expectancy phenomena being as pervasive and general as claimed by Rosenthal and others. Their attack has been on the faulty data analysis techniques being employed by many of these studies. They reviewed 31 E bias studies using the person perception task. They concluded that of the 31 studies supposedly demonstrating bias, 19 did not do so adequately. They claimed these 19 showed either negative results or poor data analysis upon which clear cut

decisions were impossible. They were very critical of the tendency of the authors to perform post hoc analysis after failing to reject the null hypotheses at an .05 level of significance. They did find that 12 of the studies seem to demonstrate the effect, indicating under certain circumstances it is a cause of bias, but it is not as general a problem as presented by many researchers.

Besides the studies using the person perception task there have been some research studies that used animals to demonstrate the E bias effect. The animal studies appear to have the advantage of not introducing the possibility of S bias that seems to exist when human Ss are used. Two of these studies are presented below to give the reader an idea of their design and findings.

Rosenthal and Fode (1963b) used written instructions to convey the experimental procedure to student Es in a bogus animal research project. The Es were instructed that their research was a replication of studies done with Maze-Bright and Maze-Dull rats. They were told that the Maze-Bright rats, through special breeding, do better on maze running tasks than ordinary rats. The task was to teach the rat to turn in the darker arm of a T maze. Each rat was run ten times a day for five days. For each trial the student E recorded whether the trial was correct or incorrect as well as the time it took to complete each trial. The results indicated that the rats run by Es believing them bright were significantly greater on the first, fourth and fifth days. When the data from all five days were combined, a one-tailed t was significant with a p of .01. On a post-experimental questionnaire those Es who were told they had Maze-Bright rats reported that their animals were more likeable, brighter and more

pleasant. These Es described their own behavior toward the rats as friendly and enthusiastic. In this study, Rosenthal and Fode said they found no instance of intentional error to account for the differences. They observed, however, five instances where Es prodded Ss to run the maze. The possibility exists that the Es engaged in other behaviors to get the desired results, which were not observed.

Rosenthal and Lawson (1964) in a similar study informed 39 student Es that they were to repeat some research which showed that Skinner Box-Bright rats do better than Skinner Box-Dull rats. The Es were then assigned the animals randomly half labelled "bright" and half labelled "dull." The rats in this study were to learn a series of operant behaviors such as running to the magazine and eating when the feeder was clicked; pressing a bar; withholding the response when extinguished; responding to a click as a secondary reinforcer; bar-pressing only when presented a light flash; responding to similar stimuli that resembled the original light; and learning a chain of responses. The results indicated that those Es who had the "bright" rats obtained superior performances than did the Es with the "dull" rats. Post-experimental findings also indicated the Es with the "bright" rats liked the animals and rated them as more friendly and pleasant than those assigned the "dull" animals. The authors also pointed out that the Es assigned to the bright rats, handled the rats more frequently and this may have had a reinforcing effect on the Ss.

Although in both these studies the authors refer to those conducting the research as Es and the rats as Ss this may be a point of confusion in making comparisons with other studies. We believe that the student Es in these studies are really the S of the study. In fact, the way the data is

analyzed, this is the case. The rats are not really Ss but simply part of the apparatus used to check on the behaviors of the student Es. The rats or the Ss in these studies serve the same purpose as a personality questionnaire in a data gathering sense. They are simply a tool used to measure the behaviors of the student Es conducting the research.

Demand Characteristics

Much of the research quoted does assume that if the S gives biased data he does so only because he has been given an expectancy by the E. These studies rarely look at the S as an independent factor capable of acting without the expectancy of the E in distorting data. Orne (1962, 1969) looked at the problem of data bias from a different perspective. Orne believes that the S's thoughts about an experiment may affect his behavior in carrying out an experimental task. Orne (1969) says that Ss are never neutral in an experiment. The Ss tend to submit to the demands of the situation and the E. Orne (1969) quotes some of his early research indicating how this affects the data gathered. Orne seems to be implying that the S's perception of what is going on is more important to the data you get than the actual experimental situation. The problem with this hypothesis is that the S may not perceive the situation in the same way as the E depending on the past experience of the S. Orne (1962, 1969) has given Ss meaningless tasks to perform to see how long an S would stay at it. His results seem to indicate Ss will do almost anything to the point of exhaustion just to complete the task. In one study Ss were required to carry out dangerous tasks such as picking up a poisonous snake or removing a penny from fuming nitric acit with their bare hands. Ss complied, surmising that, the E had somehow taken appropriate precautions for their safety.

This perception by the S of his role in research and the expectation the E has for him has been labelled the "demand characteristics" by Orne (1959). These include the rumors about an experiment, its setting, implicit and explicit instructions, and the person of the E. Orne (1969) also makes a distinction between the effects of data bias introduced by the S and those introduced by the E. "Experimenter bias effects are rooted in the motives of the experimenter, but demand characteristic effects depend on the perception of the subject (p. 147)." Orne does not deny that E bias is important. Instead he places the emphasis on the S's perception of the situation rather than the cues provided by the E. In his research Orne (1962) indicates that when demand characteristics influence the S, they tend to bias the data in a positive direction. The S wants to be a "good" S and help the E prove his hypothesis. In his research, Orne (1962) often used small groups, five Ss or less and contacted them individually or in small groups. In such a relationship the S may not have any other option but to do his "best" as the S himself perceives the situation. Orne does not operationally define demand characteristics as a single factor but instead has tried to show with his research how the behavior of S alone is not accounted for by the expectancy of the E and therefore must be attributed to the setting, the S himself or some other variable.

The following psychophysiological study by Gustafson and Orne (1965) attempts to demonstrate how demand characteristics might work in a research study. The study attempted to replicate one by Ellson, Davis, Saltzman and Burke (1952). Their research dealt with the knowledge of results on a galvanic skin response (GSR). Using a "lie detection" device they told some Ss their lies were detected, other Ss were told the opposite. On the

second trial they found that the S who believed they had been found out became harder to detect the second time while those Ss who thought they deceived the device on Trial 1 became easier to detect on Trial 2. Gustafson and Orne (1965) provided two groups of Ss different information about the effectiveness of the lie detection. One group was told:

> "This is a detection of deception experiment. We are trying to see how well the lie detector works. As you know, it is not possible to detect lying in the case of psychopathic personalities or habitual liars. We want you to try your very best to fool the lie detector during this experiment. Good luck (p. 413)."

The other group was told:

> "This is a lie detection study and while it is extremely difficult to fool the lie detector, highly intelligent, emotionally stable, and mature individuals are able to do so. (p. 413)."

It was expected that the latter group would want to deceive the E to comply with the demand characteristics while the first group would want to be detected so as not to be labelled liars of psychopaths. After the initial instructions both groups were treated alike. Each S drew a numbered card from a deck and were told to keep it secret. On the first trial the differences on GSR responses for the "secret numbers" were not significantly different. After the first trial was over the E told half the Ss in each group that they were detected and told them their secret number to prove it. The other half of the Ss were told they successfully fooled the lie detector.

On the second trial new numbers were given. The results indicate that two groups of Ss gave large GSR responses to the critical number. The Ss who wanted to be detected but had not been on the first trial and those who had tried to deceive on the first trial but were unsuccessful all gave

large GSR responses. Those Ss whose "hopes" had been confirmed now became harder to detect regardless of which group they belonged to. Ss who had originally tried to be detected and had been detected on Trial 1 now behaved physiologically like those Ss who tried to deceive and did deceive. The authors stress that the effect found is a powerful one. That on the basis of demand characteristics Ss were able to control a physiological response supposedly beyond volitional control.

Orne's work has not been as critically appraised as Rosenthal's. The reason may not necessarily be that the concept of demand characteristics has more validity than E expectancy. Instead it may be due to the lack of specificity of the concept of demand characteristics that makes it invulnerable to empirical validation. Yet Orne has pointed out that the assumption of a passive S responding to the E may not be justified unless the other variables entering into the interaction between E and S are somehow controlled.

Evaluation Apprehension

Another form of bias that has been put forth as a possible explanation for the role the S assumes is called evaluation apprehension (Rosenberg, 1965). According to Rosenberg (1969):

> "It is proposed that the typical human subject approaches the typical psychological experiment with a preliminary expectation that the psychologist may undertake to evaluate his (the subject's) emotional adequacy, his mental health or lack of it. Members of the general public, including students in introductory psychology courses, have usually learned (despite our occasional efforts to persuade them otherwise) to attribute special abilities along these lines to those whose work is perceived as involving psychological interests and skills.

> Even when the subject is convinced that his adjustment is not being directly studied he is likely to think that the experimenter is nevertheless bound to be sensitive to any behavior that bespeaks poor adjustment or immaturity.
>
> "In experiments the subject's initial suspicion that he may be exposing himself to evaluation will usually be confirmed or disconfirmed (as he perceives it) in the early stages of his encounter with the experimenter. Whenever it is confirmed, or to the extent that it is, the typical subject will be likely to experience 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. Personality variables will have some bearing upon the extent to which this pattern of apprehension develops. But equally important are various aspects of the experimental design such as the experimenter's explanatory 'pitch,' the types of measures used, and the experimental manipulations themselves (p. 281)."

In the last sentence above, Rosenberg (1969) seems to be implying that E bias contributes to evaluation apprehension. He also seems to be saying some of the same things as Orne (1969) but under a different label. Both researchers stress the perception of the S as being an important determinator of how he will react to the experiment. They place the emphasis on some internal construct, not clearly defined, operating to produce "good" data. The results of evaluation apprehension as explained by Rosenberg (1969) tend to produce positive data generated by the S in order to satisfy the possible evaluation of the E. Rosenberg (1969) reviewed research on dissonance theory and attitudinal research and is critical of the interpretation of some of the research in this area. Unfortunately, he doesn't explain his reasons in terms of experimental paradigms but rather in vague reinterpretations of conclusions presented so as to support his notion of evaluation apprehension. His research quoted is theoretical in nature and

does offer alternate explanations, but it lacks the experimental validation to support many of his generalizations.

In the same article Rosenberg (1969) describes a study where evaluation apprehension is supposedly operating. The task consisted of having undergraduate student Ss tap upon a key with their right and left index fingers for six separate ten-second intervals, half with one finger and half with the other. The number of taps were registered on a meter. The normal expectation is for a large discrepancy between the performance of the two fingers. The index finger of the dominant hand producing more taps.

As each S entered the lab he was asked "Did you take the general abilities test and personality inventory during freshman week? (p. 296." This supposedly helped create some prompting toward evaluation apprehension. The S then was administered three brief abilities tests on verbal and symbolic skills to supposedly arouse his interest in his own performance. The E then gave a description of the task to the S. For the control group only an explanation of the task was presented. In the experimental group Ss were told that recent research with graduate students at Yale and Michigan showed that the number of taps with the nondominant index finger was virtually equal to the number with the dominant index finger. The implication being that people with higher intelligence performed differently than so-called normal persons. The results indicated a significant difference ($p > .005$) between the control and experimental groups. The control group had more taps with their dominant index finger while the experimental group produced virtually equal taps with the nondominant and dominant index fingers.

Rosenberg (1969) uses the above study and several like it in his article to support his claim that evaluation apprehension is responsible for the difference in data obtained in the two groups. Actually, the difference found between the groups can be more realistically attributed to environmental manipulation rather than the construct evaluation apprehension. Looking at the study above, a different conclusion could be drawn. Ss in the study simply acted in the way the E told them to act. By such an obvious bias as telling Ss how others who are so-called "bright" did on the task, it is really not surprising that the Ss also wanted to do the task in the same manner. The study seems no different than others in which the S is given the hypothesis. As Weber and Cook (1972) have stated when this is done there tends to be research bias, ". . . a clear picture emerges of what subjects do when they know a hypothesis: They use it to determine performance, and bias is produced (p. 280)."

Ohler (1971) took 125 college student Ss and gave them a personality test and survey sheet. Half the Ss were given a choice between two apparently different experiments and the other half were told they had no choice. Ss were also allowed to read a sheet supposedly telling them the purpose of the study. Half the Ss read an explanation designed to arouse high evaluation apprehension; the other half of the Ss read an explanation designed to supposedly lower evaluation apprehension. The Ss were then presented slides and asked to indicate their like or dislike of the faces presented and how successful-unsuccessful they appeared. The results indicated that the Ss did use the written information as a cue and tried to support the E's hypothesis and that there was no difference between the low and high evaluation apprehension groups. In this study where the so-called

evaluation apprehension dimension was separated from directions stating what the E expected, nothing happened to the construct evaluation apprehension.

Blake and Heslin (1971) attempted to study evaluation apprehension by providing a group of Ss written instructions only. This it was felt would eliminate any E expectancy cues. The Ss entered a room where written instructions told them they were in a market research study (Group 1) or in a psychological health study (Group 2). The task was the person perception task. The authors found evaluation apprehension on the initial photos to be strong and significant but as the task was continued and the longer the Ss participated the effect was less and less effective.

The research quoted here does not indicate very strong support for the construct evaluation apprehension. Until Rosenberg or others more clearly define the term, research to validate its effect will be difficult to conduct.

Various Factors Mediating Bias

When Barber et al. (1969) were attempting to explain some of the differences found in their research attempts to replicate bias effects, they concluded that there seems to be an eight step transmission process involved. They credited this outline to McGuire (1967):

> "1. The student-E must attend to the expectancy communication from the principle investigator.
> 2. The student-E must comprehend the expectancy communication.
> 3. The student-E must retain the communication.
> 4. The student-E (intentionally or unintentionally) must attempt to transmit the expectancy to S.
> 5. The S ("consciously" or "unconsciously") must attend to the expectancy communication from E.

6. The S ("consciously" or "unconsciously") must comprehend E's expectancy.
7. The S ("consciously" or "unconsciously") must retain E's expectancy.
8. The S (wittingly or unwittingly) must act upon (give responses in harmony with) E's expectancy (p. 4)."

They suggested that in the various investigations that failed to replicate bias effects there is a break down in one of these eight links.

From this eight-step breakdown it can be seen that the expectancy need not be written or verbal but can be any form of communication capable of transmitting some sort of expectancy. Some of the research below has been reported by Rosenthal (1966). The research has usually been referred to as "paralinguistic cues" which cause directly or indirectly the data bias.

Adair and Epstein (1967) show that auditory cues alone may be sufficient to mediate expectancy effects. They found that Ss listening to tape-recorded instructions from Es who were given different expectancies picked up the bias and performed in accordance with it.

Zoble (1968) had two experimental groups perform at a tone-length discrimination task. The Es were all given various expectancies about the performance of the Ss. One of the experimental groups was provided only with auditory cues from their E. For this group the results showed that 53 per cent of the total expectancy effect was obtained when Ss were provided only auditory cues. The other experimental group was given only visual access to cues from their E. He found that the visual cues were more important than the auditory ones. Ss who had access to only the visual cues gave data that was consistent with their Es expectancy 75 per cent of the time, compared with Ss having access to both modes of information, auditory and visual combined.

Other studies looking at mediating variables have concentrated on the role portrayed by the E as perceived by the S. Rosenthal, Kohn, Greenfield and Carota (1966) performed an experiment on verbal reinforcement with the Es status defined in terms of his behavior during his interaction with the S. Each S was asked to rate his E on four behavior categories (Businesslike, Expressive Voice, Professional, Use of Legs). These categories were not operationally defined for the Ss so the results are based entirely on the Ss subjective evaluation of these so-called behaviors. They predicted that a more professional businesslike, less noisy, and more consistent E would be ascribed a higher status by his S. Using 19 male Es who said the word "good" whenever the Ss used first person pronouns, they found higher status Es as defined by their Ss' perception of their behavior were significantly more influential in changing their Ss' responses.

In another study Rosenthal, Kohn, Greenfield and Carota (1965) used praise to induce a good mood in Es and reproof to induce a negative mood. Twenty-six college seniors were used as Es. One-hundred and fifteen female Ss were tested using the person perception task. To eliminate as much bias as possible from the interaction between the Es and Ss, written instructions alone were used to convey the experimental conditions to the Es participating. Results indicated that the Ss tended to give more positive data (labelled photos as successful people) when Es were evaluated positively. When E was evaluated negatively, Ss tended to give negative ratings.

Barber and Calverley (1964) in an experiment in hypnosis, had the single E sometimes adopt a forceful, authoritative tone of voice and sometimes a lackadaisical one. Ss under hypnosis accepted more suggestions when offered in the authoritative tone than when offered in a bored

disinterested tone. They concluded that Es who are more likely to influence their Ss responses, behave in a more professional way.

Burney (1958) found that Ss responses obtained by two faculty Es reflected a higher need for achievement than did Ss responses obtained by a student E. Rosenthal (1966) suggests that Ss may feel a greater need to achieve when in interaction with others who have probably achieved more.

Most of the studies supporting the notion that the role of the E as perceived by S affects the data, used a rating scale usually from +5 to -5. The S is asked to rate his E on several constructs such as Relaxed, Likable, Professional, Honest, Casual, etc. (Rosenthal, 1966). As previously noted these are referred to often as behaviors but are not operationally defined in a behavioral context.

When the E has had prior contact with his S, even when the contact is brief, the S may respond differently in the experimental task. When the performance required is difficult, prior contact, especially when of a "warm" quality, seems to improve performance (Rosenthal, 1966). It has also been suggested that in a military or an academic setting, a higher status E may evoke more negative feelings from the S. In the academic setting, the higher status or more professional acting E may be seen as a psychologist capable of looking into the mind of the S (Riecken, 1962) and as a result is feared by the S.

According to Rosenthal (1966) the kind of person the E is, how he or she looks and acts, may by itself affect the S's response. Sometimes the effect is a direct and simple one, but sometimes too, the effect is found to interact with S characteristics, task characteristics or situational characteristics.

Rosenthal (1969) concludes that:

> "the sex, age and race of the investigator have all been found to affect the results of his research. What we do not know and what we need to learn is whether Ss respond differently simply due to the presence of Es varying in these biosocial attributes. We also need to determine if due to these variables the Es behave differently toward Ss varying in these characteristics (p. 186)."

S as the Mediating Variable

There has been research which has looked at both the role of the S and the interaction of the E and S as being the primary cause of data bias. Some of the studies concentrating on the S have already been presented under Demand Characteristics and Evaluation Apprehension. Below are listed some of the articles which concentrate on the role of the S in biasing data.

Kintz, Delprato, Metter, Persons and Schappe (1965) stress the importance of various organismic variables (age, sex, etc.) and various personality factors having a direct influence on the outcomes of E and S interaction. In one of these studies cited, by Woods (1961), 1,737 published experiments were examined closely for the treatment of the E and S interaction. None of these studies ran an analysis of E interaction. Although these factors are considered academically important, virtually no one looks at their influence in the actual research they conduct.

According to McGuire (1969) there is growing concern regarding the probability and effect of a growing suspiciousness regarding E's intent by S drawn from heavily used populations such as college sophomores. This group is becoming research "wise" and is not at all an adequate sample for generalizable research. Orne (1962), Schultz (1963), Silverman (1965), Resnick and Schwartz (1973) have all suggested that when attempts to

influence Ss become more obvious, Ss become less influenceable. This may be the trend of the future. As more Ss of psychological experiments become acquainted with the results of classic research in conformity there may be more and more determination to show the E that the S is not to be regarded as "one of those mindless acquiescers (p. 220)."

In his chapter on S set, Rosenthal (1966) deals with expectations but limits them in terms of E expectancy. The expectancy of the S may be more important than the expectancy of the E as Rosenberg (1969) and Orne (1969) point out. A conclusion that can be drawn from some of the research available is that Ss tend to respond in a way that they perceive to be appropriate (in light of the investigator's attributes). Ss in experiments apparently want to do the right thing and be well evaluated (Orne, 1962), Riecken (1962), Rosenberg (1965). To insure positive evaluation they react positively to the investigator. The S trying to be a good S has been the trend up until the present time.

Rosenthal (1966) notes that students' motivation for distorting or changing data is not to hoax their instructors but to hand in a "better report," one that conforms to the expected data. He also points out that students sometimes fear that a poor grade will be the result if the data does not conform to the expected event. The data to support this notion has been obtained through the use of post-experimental questionnaires. These questionnaires are usually given to the Ss right after the experiment. They usually ask the S what he thought the hypotheses were and why he responded the way he did. The problem with such questionnaires, however, is that they are also subject to such influences as E expectancy bias (Rosenthal, 1966).

Rosenberg (1969) suggests that most experimental psychologists are aware that their Ss are prone to "faking it" and particularly, to faking it "good." His research shows that there are problems even using only written instructions. For example, if the S obtains cues that "personality" is to be measured, evaluation apprehension will be created causing the S to do his best to look good by biasing the data in a positive direction. In several studies by simply indicating to one group that they were in a personality study, their scores on various tasks tended to be higher than control group Ss.

Rosenberg (1969) says that the E oriented theories, such as Rosenthal's (1966) view the S as a passive recipient in the whole process of recieving cues from the E. He describes his view toward bias introduced by the S in the following statement:

> "This would suggest that where such cues are absent or imperceptible systematic bias would be unlikely to occur. In distinction, a subject-oriented theory of the experimental transaction views the subject as seeking something from the experimental experience. In the present theoretical view that 'something' is the experimenter's judgmental validation of the subject's psychological adequacy and on this basis, the ultimate maintenance of enhancement of the subject's self-esteem (p. 344)."

Criswell (1958) claimed that S's curiosity about the research project and a need on the part of the S's to ingratiate himself with the E were factors causing data bias. This last motive supposedly explains why subjects who correctly "read" the E's unintentionally communicated expectancy generally go along with it rather than choosing to disregard it or defy it. The S in this situation also perceives of himself as making a useful contribution to the research by confirming the E's hypotheses as Orne (1962) suggests.

A systematic analysis of the social interactions in the psychological experiment was made by Riecken. According to Rosenthal (1966), Riecken (1962) notes three possible motives of the S.

> "First, the S strives to attain those rewards, such as money, course credit, and psychological insight, which he feels are his due from having accepted the invitation to participate. Second, Ss strive to 'penetrate the experimenter's inscrutiability and discover the rationale of the experiment.' The third motive is to 'represent himself in a favorable light to the E (p. 181)."

Riecken (1962) is basically summarizing what both Orne (1962, 1969) and Rosenberg (1969) have said. In the recent study by Resnick and Schwartz (1973), a group of Ss provided with information on the exact purpose of the study, provided data completely opposite from a group not given similar information. Perhaps the trend has finally changed. Today's S, instead of complying, may go out and try to give the E data contrary to the hypotheses. The Resnick and Schwartz (1973) article is presented in detail later on. Weber and Cook (1972) also indicated this possibility when they pointed out in their review of the literature on S bias that when the S is provided the hypothesis the result in six out of six cases was data bias.

Research Settings and E Characteristics

There has been a great deal of research on data bias but very little of it has ever been collected in what might be called a naturalistic setting. In most cases the Ss knew they were in a study. In the studies where the Es behavior was the variable studied, they too were aware that research was going on even if they were not sure of the purpose. The effects of the lab setting on the data is not clear but several researchers have expressed concern over the artificiality involved.

Riecken (1962) has pointed out how much there is we do not know about the effects of the physical scene in which an experimental transaction occurs. For example, Riecken wondered about the effect on his Ss of the E's white coat. Even the E may think he is more of a scientist.

Not only is the setting itself important but the appearance of the E in that setting also seems to have an influence. Some researchers now think that once a S hears the words "psychological research" the Ss assume a deception of some sort will take place.

Ss are far from being the automated data production units that psychologists believe. Riecken (1962) has suggested that this current view of the S is seriously in error. Argyris (1968) has pointed out that some of the current problems of data bias have been caused by the means in which psychological research has been conducted. Most Ss no longer believe the E; instead, they expect to be tricked in some way and act accordingly. Argyris (1968) suggests that if the S likes the E he will cooperate. On the other hand, if he does not like the E, the S may enjoy "botching the works with such great skill that the experimenter is not aware of this behavior (p. 187)." Unfortunately, Argyris never states the nature of the attraction or how he validated the construct "like."

Other evidence also suggests that Ss are sometimes interested in their E as a person rather than simply as an inscrutable scientist-psychologist. Rosenthal, Kohn, Greenfield and Carota (1966) found that on post-experimental questionnaires, 20 per cent of the Ss made some reference to one or more physical characteristics of their E which were "irrelevant" to the E's role performance. References were made the E's posture, clothing, facial blemishes, eyeglasses, dental condition, and relative attractiveness.

Weber and Cook (1972) support the notion that we must conduct more of our research outside of the laboratory. This may eliminate some of the more obvious elements providing cues to the S as to the nature of the experiment. They also state that Es should not be labeled as psychologists. In stating their view on the present state pf research, they say:

> " A wider contact with naturalistic studies, both experimental and otherwise, is required as a corrective for the systematically distorted picture of social behavior that emerges because of the roles that subjects may adopt in laboratory experiments (p. 291)."

By providing naturalistic setting as Weber and Cook (1972) propose, you eliminate many of the problems cited. The person collecting the data is no longer an E but a teacher or supervisor or colleague. The so-called "mental sets" that Ss have about research should not affect the data since most would be unaware their behavior was being observed for "research" projects.

Design Problems

Rosenthal (1966) says that any method making it less likely that experimenter effects will interact with treatment conditions would reduce our problem of assessing adequately the effects of our treatment conditions. Any data obtained by a single E may be due as much to the E as to his treatment conditions. He offers the suggestion that we use several Es who don't know the hypotheses. According to Brunswick (1956), Hammond (1954), and Rosenthal (1966), one way to increase generalizability is to employ samples of Es. Due to differences in appearance and behavior, different Es serve as different stimuli to their Ss, often changing the experimental situation as the Ss confront it. When only a single E is

employed, you don't know how much of the results are a reflection of his stimulus value, results are confounded. By eliminating E and S contact, we can help eliminate experimenter expectancy effects. Dolly (1973a) demonstrated that E bias may be eliminated by having several Es collect the data and by not allowing the Es access to the hypotheses being tested.

In research on bias effects there appears to be some limitations on the experimental design that can be employed. Most all researchers avoid giving pre-tests for fear of sensitizing their Ss. In most studies post-test designs are employed. Most researchers seem to think that Ss become suspicious of the researchers intent more often in pre-test designs than in post-test only designs. However, McGuire (1969) claims he has never found significant differences between suspicious and non-suspicious Ss regarding the effects of any of the important variables.

According to Lana (1969) the Hawthorne studies showed that the fact of measurement not only changed the magnitude of the dependent variable (rate of production), but of the very nature of the social situation. Lana (1969) argues a pre-test design only sensitizes Ss to the experimental situation and should be replaced by randomization in a post-test only design.

Rosenthal (1963) says that even when a pre-test is not used and the E does not tell the S the hypothesis this may not insure that the S is ignorant of it. This view seems consistent with the findings of Rosenberg (1969), who tried to show that any cues present in the experimental situation could provide either correct or incorrect information to the S about the hypotheses. Even if the information the S thought he learned was incorrect, it will still bias the data since the S will tend to operate on the premise of knowing the "real" reason for the research.

Campbell and Stanley (1966) said:

> "In the usual psychological experiment, if not in educational research, a most prominent source of unrepresentativeness is the patent artificiality of the experimental setting and the student's knowledge that he is participating in an experiment. For human experimental subjects, a higher order problem-solving task is generated in which the procedures and experimental treatment are reacted to not only for their simple stimulus values, but also for their role as clues in deriving the experimenter's intent (p. 20)."

There is some evidence that even though data is biased that the bias tends to balance out. In an experiment by Persinger, Knutson and Rosenthal (1968), cited by Rosenthal and Rosnow (1969), the procedures were all filmed and tape recorded without the Es or Ss knowledge. Comparing the results of the Es with independent observers it was found that .72 per cent of the Es' transcriptions were in error and that .48 per cent of the transcriptions erred in the direction of the Es' hypotheses while .24 per cent of the transcriptions erred in the opposite direction. The latter errors tended to be larger so that the mean error per E was -.0003 in a direction opposite the Es' expectancies.

According to Skinner (1953) it is accepted as a general principle in scientific method that it is necessary to interfere in some degree with any phenomenon in the act of observing it. The important thing is that behavior may be observed with a minimum of interaction between Ss and scientist, and this should be where we try to begin. Weber and Cook (1972) have pointed out that most research conducted lacks external validity since it cannot be generalized beyond laboratory settings or college student populations. They point out that much of the research on "subject effects" places further limitations on the data indicating

many studies may not even be internally valid. By this they mean the findings of the research may not be caused by the experimental treatment.

In any psychological experiment there appears to be little that can be done to eliminate all data bias. There are, however, certain generally agreed upon procedures for reducing the effects of bias in psychological research. Rosenthal (1966) outlined the following ten strategies for controlling E expectancy effects.

"Strategies for the Control
of Experimenter Expectancy Effects

1. Increasing the number of experimenters:
   decreases learning of influence techniques
   helps to maintain blindness
   minimizes effects of early data returns
   increases generality of results
   randomizes expectancies
   permits the method of collaborative disagreement
   permits statistical correction of expectancy effects

2. Observing the behavior of experimenters:
   sometimes reduces expectancy effects
   permits correction for unprogrammed behavior
   facilitates a greater standardization of
   experimenter behavior

3. Analyzing experiments for order effects:
   permits inference about changes in experimenter behavior

4. Analyzing experiments for computational errors:
   permits inference about expectancy effects

5. Developing selection procedures:
   permits prediction of expectancy effects

6. Developing training procedures:
   permits prediction of expectancy effects

7. Developing a new profession of psychological experimenter:
   maximizes applicability of controls for expectancy effects
   reduces motivational bases for expectancy effects

8. Maintaining blind contact:
   minimizes expectancy effects

9. Minimizing experimenter-subject contact:
   minimizes expectancy effects

10. Employing expectancy control groups:
    permits assessment of expectancy effects (p. 402)."

What is of major interest is the fact that so few of these procedures have been implemented. Even Rosenthal has not implemented all the controls suggested. Dolly (1973b) used procedures one and eight above and was able to eliminate any trace of E bias effects.

Ethical Considerations

A major problem of much of the research on bias effects has been the deception required to test many of the hypotheses put forth. This is contrary to the current emphasis on improving ethical standards in research being supported by the American Psychological Association (APA). Rosenthal (1966) suggested for now the fact of deception must be accepted. The hope must be that the knowledge acquired through this necessary deception is worth the price of having deceived. He felt that if you can answer the following question in the affirmative then the deception is excusable, "Is the deception warranted by the potential importance of the results of the expectancy-controlled experiment? (p. 388)"

In research on bias effects, the fact that the S must be deceived is taken for granted. This deception does seem inconsistent with the new proposed APA (1972) ethical guidelines. McGuire (1969) pointed out that in studies on attitude change it is simply taken for granted that the E cannot admit the purpose of the study to the Ss. If the E did tell the

Ss, the Ss' behavior could not be interpreted and generalized to the behavior of naive Ss to whom our theories of persuasion are supposed to apply. Weber and Cook (1972) agree; they found that in six cases where Ss were provided information concerning the hypotheses involved, data bias occurred.

The effect of being duped is not really clear. Fillenbaum (1966) found that the performance of the "faithful" S who has been exposed to prior deceptions yields results little different from those of naive subjects. Brock and Becker (1966) find that prior participation in a deception experiment with debriefing produces surprisingly little effect on performance in a subsequent test experiment, even when it follows immediately afterwards. Only when the test experiment and the prior debriefing experiment were made similar was performance found to be substantially affected.

According to McGuire (1969) to ask about the effect of "suspiciousness of experimenter's intent (p. 13)" is to ask what is the effect of awareness of what is going on in the experiment.

> "In the lab experiment we strain our intellectual and moral resources to design some elaborate deception which will hide from the S the persuasive nature of the material--we are also making it more difficult to generalize to the naturalistic situation (p. 37)."

Yet the E often justifies the deceptions to enable him to generalize. McGuire (1969) says that the usual psychological S is fairly cooperative even when he is forced to participate. He guessed that 9 out of 10 would prefer to help rather than hinder the E. This assumption they want to help may not be warranted.

Resnick and Schwartz (1973) attempted to literally implement the ethical guidelines being promoted by the APA. Using two separate groups

of Ss one group was provided with a verbal conditioning task under the usual procedures in which no information concerning the purpose of hypotheses were provided. The second group of Ss were provided information on all aspects of the research as outlined in the draft reports on "Ethical Standards and Psychological Research" (Cook, Kimble, Hicks, McGuire, Schoggen, and Smith, 1972). The study consisted of reinforcing students for using the pronouns I or We as the first word in a sentence construction task. The so-called "ethical" group was fully briefed on all aspects of the study.

Results showed that the "non-ethical" group acted in a predictable manner showing significant positive conditioning. The "ethical" group showed significantly negative conditioning indicating that implementing the ethical guidelines has a strong effect on the way Ss react in an experimental situation. The authors stress that before we fully implement these "ethical standards" more research is needed to determine the exact effects. An especially interesting point made was a number of Ss in the "ethical" condition refused to believe the E and were certain they were being duped. This type of response set has been pointed out in detail by McGuire (1969), but his assumption of their desire to help does not seem supported by the group provided with the research information. One quote of McGuire (1969) that may seem appropriate in view of the research to date is, "It's absurd to take the subject fully into the experimenter's confidence and expect to find generalizable results."

The authors of this paper see the concern of many professional people and researchers who think restrictions on human research are necessary. These restrictions were meant to prevent Ss in research from being harmed

either directly or indirectly by the experiment itself. It seems, however, that in attempts to censure a handful of psychologists who mistreated subjects, all researchers are expected to operate under a set of restrictions that eliminates many simple controls necessary in research on behavioral artifacts. The loss of control over the Ss and over the research environment itself makes it very doubtful that the answers to the questions raised in this paper will ever be answered.

Summary

This paper has attempted to show how the problem of data bias has been approached, conceptualized and studied. Some of the research on E expectancy, E bias, demand characteristics, and evaluation apprehension were presented. In addition, some of the situational and organismic variables mediating bias were also discussed. Some of the research available on the design of bias studies and research arguing for more naturalistic settings were also cited.

As mentioned previously, many of the so-called behaviors measured by this review did not fit the description as normally viewed in psychology. Rosenthal (1966) for example, refers to such constructs as Like and Professional as behaviors. Yet he defines these constructs by use of a rating scale from +5 to -5 filled out by the Ss in many of the research studies he quoted. No where are the constructs operationally defined in terms of E's behavior. The only data available tends to be a subjective rating by the S on how they perceive the E, and this rating is restricted by the concepts presented to the S for ratings. According to Cronbach and Meehl (1967) constructs of the type used in many of the articles quoted can be

validated only by research that clearly shows a connection between the construct and behavior.

> "Scientifically speaking, to "make clear what something is" means to set forth the laws in which it occurs. We shall refer to the interlocking system of laws which constitute a theory as a nomological network. The laws in a nomological network may relate (a) observable properties or quantities to each other; or (b) theoretical constructs to observables; or (c) different theoretical constructs to one another (p. 255)."

Skinner (1953) has also been critical of the use of constructs without trying these constructs to observable behaviors. Unless at some point these constructs are based on observations of behavior there is little that can be done to experimentally explore their usefulness in psychological theory. Unfortunately, the research reviewed in this paper rarely defines the "motives" and "constructs" used to explain data bias in a behavioral context.

Assuming that these constructs are valid, however, one must then look at the large volume of data in social and personality psychology and try to determine the exact behavioral causes of the data bias. The fact that data is being biased is supported by the literature. It is only the explanations of the causes of the bias that are questioned by the authors.

REFERENCES

Adair, J. G., and Epstein, J. Verbal cues in the mediation of experimenter bias. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May, 1967. Cited by R. Rosenthal and R. L. Rosnow, Artifact in behavioral research. New York: Academic Press, 1969, p. 269.

Anderson, R. C., Kulhavy, R. W., and Andre, T. Conditions under which feedback facilitates learning from programmed lessons. Journal of Educational Psychology, 1972, 63, 3, 186-188.

APA Monitor, Ethical standards for research with human subjects. 3, 5, 1972. I-XIX American Psychological Association, Washington, D. C.

Argyris, C. Some unintended consequences of rigorous research. Psychological Bulletin, 1968, 70, 185-197.

Ausubel, D. P. Educational psychology a cognitive view. New York: Holt, Rinehart and Winston, Inc., 1968.

Barber, T. X., and Calverly, D. S. Effect of E's tone of voice on "hypnotic-like" suggestibility. Psychological Reports, 1964, 15, 139-144.

Barber, T. X., Forgione, A., Chaves, J. F., Calverly, D. S., McPeake, J. D., and Bowen, Barbara. Five attempts to replicate the experimenter bias effect. Journal of Consulting and Clinical Psychology, 1969, 33, 1-6.

Barber, T. X., and Silver, M. J. Fact, fiction and the experimenter bias effect. Psychological Bulletin, 1968, 70, 6, 1-29.

Blake, B. F., and Heslin, R. Evaluation apprehension and subject bias in experiments. Journal of Experimental Research in Personality, 1971, 5, 1, 57-63.

Brock, T. C., and Becker, L. A. Ineffectiveness of "overhead" counter-propaganda. Journal of Personality and Social Psychology, 1965, 2, 654-660.

Brunswick, E. Perception and the representative design of psychological experiments. Berkeley: University of California Press, 1956.

Bugelski, B. R. The psychology of learning applied to teaching. New York: Bobbs-Merrill, 1971.

Burney, R. C. The achievement motive and task performance: a replication. Journal of Abnormal and Social Psychology, 1958, 56, 133-135.

Cobb, J. A. Relationship of discrete classroom behaviors to fourth-grade academic achievement. Journal of Educational Psychology, 1972, 63, 74-80.

Cook, T. D., Bean, J. R., Calder, B. J., Frey, R., Krovetx, M. L., and Reisman, S. R. Demand characteristics and three conceptions of the frequently deceived subject. Journal of Personality and Social Psychology, 1970, 14, 185-194.

Cook, S. W., Kimble, G. A., Hicks, L. H., McGuire, W. J., Schoggen, P. H., and Smith, M. B. Ethical standards for psychological research: Proposed ethical principles submitted to the APA membership for criticism and modification (by the) ad hoc Committee on Ethical Standards in Psychological Research. APA Monitor, 1971, 2(7), 9-28.

Creswell, Joan H. The psychologist as perceiver. In R. Taguire and L. Petrullo (eds.), Person perception and interpersonal behavior. Stanford, California: Stanford University Press, 1958, pp. 95-109.

Cronbach, L. J. Essentials of psychological testing. New York: Harper and Brothers, 1960.

Cronbach, L. J., and Meehl, P. E. Construct validity in psychological tests. In W. A. Mehrens and R. L. Ebel (eds.), Principles of educational and psychological measurement: A book of selected readings. Chicago: Rand McNally, 1967, pp. 242-270.

Deno, S. L., Jenkins, J. R., and Marsey, Judy. Transfer variables and sequence effects in subject-matter learning. Journal of Educational Psychology, 1971, 62, 365-370.

Dolly, J. P. The effect of data bias introduced by subjects in classroom research. A paper presented at the annual meeting of the Southeastern Psychological Association, New Orleans, April, 1973(a).

Dolly, J. P. The subject bias effect in classroom research. Unpublished doctoral dissertation. The University of Georgia, 1973(b).

Domino, G. Interactive effects of achievement orientation and teaching style on academic achievement. Journal of Educational Psychology, 1971, 62, 427-431.

Ekman, P., and Friesen, W. V. Status and personality of the experimenter as a determinant of verbal conditioning. American Psychologist, 1960, 15, 430. (Abstract)

Ellson, D. G., Davis, R. C., Saltzman, I. J., and Burke, C. J. A report on research on detection of deception 1952. Cited by R. Rosenthal and R. L. Rosnow, Artifact in behavioral research. New York: Academic Press, 1969, p. 178.

Fillenbaum, S. Prior deception and subsequent experimental performance: the "faithful" subject. Journal of Personality and Social Psychology, 1966, 4, 537.

Fillenbaum, S., and Frey, R. More on the "faithful" behavior of suspicious subjects. Journal of Personality, 1970, 38, 43-51.

Friedman, M. P., and Greitzer, F. L. Organization and study time in learning from reading. Journal of Educational Psychology, 1972, 63, 6, 609-616.

Gephart, W. J., and Antonopolis, D. P. The effects of expectancy and other research-biasing factors. Phi Delta Kappan, 1969, 10, 579-583.

Goldman, R. D. Effects of a logical versus a mnemonic learning strategy on performance in two undergraduate psychology classes. Journal of Educational Psychology, 1972, 63, 4, 347-352.

Grieger, R. M. The effect of teacher expectancies on the intelligence of students and the behavior of teachers. Unpublished doctoral dissertation, Ohio State University, 1970.

Gustafson, L. A., and Orne, M. T. Effects of perceived role and role success on the detection of deception. Journal of Applied Psychology, 1965, 49, 412-217.

Hammer, B. Grade expectations, differential teacher comments, and student performance. Journal of Educational Psychology, 1972, 63, 5, 454-458.

Hammond, K. R. Representative vs. systematic design in clinical psychology. Psychological Bulletin, 1954, 51,150-159.

Hicks, R. A., and Packwood, J. Effects of structural complexity of bigrams on paired-associate learning and retention of children. Journal of Educational Psychology, 1971, 62, 290-292.

Holmes, D. Effects of grades and disconfirmed grade expectancies on students' evaluations of their instructor. Journal of Educational Psychology, 1972, 63, 130-133.

Jenkins, J. R., and Deno, S. L. Influence of knowledge and type of objectives on subject-matter learning. Journal of Educational Psychology, 1971, 62, 67-70.

Kintz, B. L., Delprato, D. J., Mettee, D. R., Persons, C. E., and Schappe, R. H. The experimenter effect. Psychological Bulletin, 1965, 63, 4, 223-232.

Kulhavy, R. W., and Parsons, J. A. Learning - criterion error perservation in text materials. Journal of Educational Psychology, 1972, 63, 81-86.

Lana, R. E. Pretest sensitization. In R. Rosenthal and R. L. Rosnow (eds.), Artifact in behavioral research. New York: Academic Press, 1969, pp. 121-141.

Laszlo, J. P., and Rosenthal, R. Subject dogmatism, experimenter status and experimenter expectancy effects. Cambridge, Mass.: Harvard University Department of Social Relations, 1967 (Mimeo). Cited by R. M. Grieger, The effects of teacher expectancies on the intelligence of students and the behavior of teachers. Unpublished doctoral dissertation, Ohio State University, 1970.

Levin, J. R., and Horwitz, J. M. The meaning of paired associates. Journal of Educational Psychology, 1971, 62, 209-214.

McClelland, D. C., Clark, R. A., Roby, T. B., and Atkinson, J. W. The projective expression of needs. IV. The effect of the need for achievement on thematic apperception. Journal of Experimental Psychology, 1949, 39, 242-255.

McGuire, W. J. Suspiciousness of experimenter's intent. In R. Rosenthal and R. L. Rosnow (eds.), Artifact in behavioral research. New York: Academic Press, 1969, pp. 13-56.

McGuire, W. J. Personality and susceptibility to social influence. In E. F. Borgatta and W. W. Lambert (eds.), Handbook of personality theory and research. Chicago: Rand McNally, 1967.

Masling, J. Role-related behavior of the subject and psychologist and its effects upon psychological data. Nebraska Symposium on Motivation, 1966, 14, 67-103. Cited by Weber, S. J. and Cook, T. D. Subject effects in laboratory research: An examination of subject roles, demand characteristics and valid inference. Psychological Bulletin, 1972, 77, p. 275.

Mayer, R. E., and Greeno, J. G. Structural differences between learning outcomes produced by different instructional methods. Journal of Educational Psychology, 1972, 63, 165-173.

Ohler, F. D. The effects of four sources of experimental bias: evaluation apprehension, cueing, volunteer status, and choice. (Doctoral dissertation, University of South Carolina) Ann Arbor, Michigan: University Microfilms, 1971, No. 71-28, 881.

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.

Orne, M. T. Demand characteristics and the concept of quasi-controls. In R. Rosenthal and R. L. Rosnow (eds.), Artifact in behavioral research. New York: Academic Press, 1969, pp. 147-177.

Persinger, G. W., Knutson, S., and Rosenthal, R. Communication of interpersonal expectations of ward personnel to neuropsychiatric patients. 1968. Cited by R. Rosenthal and R. L. Rosnow, Artifact in behavioral research. New York: Academic Press, 1969, p. 259.

Quirk, T. J., Steen, Margaret T., and Lipe, D. Development of the program for learning in accordance with Needs Teacher Observation Scale. Journal of Educational Psychology, 1971, 62, 188-200.

Resnick, J. H., and Schwartz, T. Ethical standards as an independent variable in psychological research. American Psychologist, 1973, 28, 2, 134-319.

Rhodes, F. Team teaching compared with traditional instruction in grades kindergarten through six. Journal of Educational Psychology, 1971, 62, 110-116.

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

Rosenberg, M. J. When dissonance fails: on eliminating evaluation apprehension from attitude measurement. Journal of Personality and Social Psychology, 1965, 1, 18-42.

Rosenberg, M. J. The conditions and consequences of evaluation apprehension. In R. Rothenthal and R. L. Rosnow (eds.), Artifact in behavioral research. New York: Academic Press, 1969. pp. 280-348.

Rosenfeld, G. W. Some effects of reinforcement on achievement and behavior in a regular classroom. Journal of Educational Psychology, 1972, 63, 3, 189-193.

Rosenthal, R. On the social psychology of the psychological experiment: the experimenter's hypothesis as unintended determinant of experimental results. American Scientist, 1963, 51, 268-283.

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

Rosenthal, R. Unintended effects of the clinician in interaction: a taxonomy and a review of clinician expectancy effects. Australian Journal of Psychology, 1969, 21, 1-20.

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

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

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

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., and Rosnow, R. L. (eds.), Artifact in behavioral research. New York: Academic Press, 1969.

Rothkopf, E. Z., and Kaplan, R. Exploration of the effect of density and specificity of instructional objectives on learning from text. Journal of Educational Psychology, 1972, 63, 4, 295-302.

Rychlak, J. F., and Tobin, T. J. Order effects in the affective learning styles of overachievers and underachievers. Journal of Educational Psychology, 1971, 62, 141-417.

Scandura, J. M., and Voorhies, D. Effect of irrelevant attributes and irrelevant operations on rule learning. Journal of Educational Psychology, 1971, 62, 352-356.

Schultz, D. P. Time, awareness and order of presentation in opinion change. Journal of Applied Psychology, 1963, 47, 280-283.

Silverman, I. Motives underlying the behavior of the subject in the psychological experiment. In W. E. Vinacke (Chm.), Ethical and methodological problems in social psychological experiments. Symposium presented at the meeting of the American Psychological Association, Chicago, September, 1965. Cited by Weber, S. J., and Cook, T. D. Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. Psychological Bulletin. 1972, 77, p. 275.

Skinner, B. F. Science and human behavior. New York: The Macmillan Company, 1953.

Stevenson, H. W., Williams, Anne M., and Coleman, E. Interrelations among learning and performance tasks in disadvantaged children. Journal of Educational Psychology, 1971, 62, 179-184.

Tallmadge, K. G., and Shearer, J. W. Interactive relationships among learner characteristics, types of learning, instructional methods and subject matter variables. Journal of Educational Psychology, 1971, 62, 31-38.

Todd, W. B., and Kessler III, C. C. Influence of response mode, sex, reading ability, and level of difficulty on four measures of meaningful written material. Journal of Educational Psychology, 1971, 62, 229-234.

Weber, S. J. and Cook, T. D. Subject effects in laboratory research: An examination of subject roles, demand characteristics and valid inference. Psychological Bulletin, 1972, 77, 273-295.

White, C. R. The effect of induced subject expectation on the experimenter bias situation. Unpublished doctoral dissertation, University of North Dakota, 1962. Cited by R. Rosenthal, Experimenter effects in behavioral research. New York: Appleton-Century-Crofts, 1966, p. 193.

Woods, P. J. Some characteristics of journals and authors. American Psychologist, 1961, 16, 699-701.

Zoble, E. J. Interaction of subject and experimenter expectancy effects in a tone length discrimination task. Franklin and Marshall College, 1968. Cited by R. Rosenthal and R. L. Rosnow (eds.), Artifact in behavioral research. New York: Academic Press, 1969, p. 251.