ON THE SOCIAL PSYCHOLOGY OF THE PSYCHOLOGICAL EXPERIMENT:[1,2] THE EXPERIMENTER'S HYPOTHESIS AS UNINTENDED DETERMINANT OF EXPERIMENTAL RESULTS

By ROBERT ROSENTHAL

SCIENTISTS are aware of the fact that they are imperfect instruments in the quest for lawful relationships (Wilson, 1952). Errors of observation and of interpretation have been discussed systematically from the time of the discovery of the personal equation among the astronomers Bessel, Kinnebrook, and Maskelyne, *et al.* (Boring, 1950). A lively interest in these problems is to be found today among medical researchers and particularly among those working with drugs. Various techniques, such as the "double-blind" method, have been developed in which neither patient nor physician is to be aware of the nature of the substance ingested by the patient (Beecher, 1959). One purpose of this technique, of course, is to avoid errors of observation[3] and interpretation[3] in both subject (patient) and experimenter (physician).

For many of the sciences, there seems to be little danger that the act of observation itself may change the object of study, if the object be macroscopic (Reichenbach, 1951). For the behavioral sciences, however, when humans or animals are the object of study, the act of observation may very well change the object of study. Research in the assessment of personality has shown that the personality and behavior of the assessor (observer) can change the response of the subject (Masling, 1960). The interviewer in the public opinion survey has been very systematically studied for his effect upon his respondents (Hyman, *et al.*, 1954). The experimental psychologist, working in his laboratory rather than the clinic or the field, has been less systematically investigated. Nevertheless, studies have shown that different psychological experimenters (*E*s) may obtain statistically significantly different data from comparable human subjects (*S*s) (McGuigan, 1961; Mulry, 1962; Pflugrath, 1962). Further evidence suggests that different *E*s may obtain statistically significantly different data from comparable *S*s even

[1] This paper is an expanded and revised version of one presented at the symposium: *On the Social Psychology of the Psychological Experiment.* Henry W. Riecken, Chairman, Amer. Psychol. Ass., New York; Sept., 1961. I am particularly grateful for the personal encouragement and intellectual stimulation provided by Donald T. Campbell, Harold B. Pepinsky, and Henry W. Riecken.

[2] Preparation of this paper and most of the investigations summarized were supported by research grants (G-17685 and G-24826) from the division of Social Sciences of the National Science Foundation.

[3] Papers with fairly extensive bibliographies dealing with these and other types of experimenter effects have been prepared for publication.

when *S*s are planaria (an invertebrate organism placed low on the phylogenetic scale) (Rosenthal and Halas, 1962).

Findings such as those presented, and the conceptualization of the psychological experiment as a social situation led Riecken (originally in 1958, now in press) to state clearly the need for a social psychology of the psychological experiment. Orne (1961) and White (1962) have studied the role of *S* in the *E*-*S* dyad and shown that certain features of an experimental situation may cue *S* as to what responses may be desired. The purpose of this paper is to consider the role of *E* as a partial determinant of the outcome of his experiments. More specifically we will consider *E* outcome-orientation bias; that is, the notion that *E*s obtain from their *S*s, human or animal, the data they want and/or expect to obtain (Rosenthal, 1956).

When an *E* undertakes an experiment, even if it is not very explicitly formulated, he has some hypothesis or expectancy about the outcome. The expectancy may be vague, indeed it may be a family of expectancies; but the fact of having selected one or more particular variables for study rather than other variables serves as a clue to the nature of the expectancy. To the extent that this expectancy (and motivational variables associated with it) is a determinant of experimental results, we must re-evaluate most carefully the results of those experiments of the past and those proposed for the future which may not have been controlled adequately for the operation of this phenomenon.

We will first present some of the evidence for the occurrence of this phenomenon and for its generality. We will then turn to a consideration of what is known about the sources and mediation of this phenomenon and finally consider some implications of our findings for research methodology in the behavioral sciences.

Evidence for the Occurrence of Experimenter Bias

The basic paradigm for the study of this phenomenon has been to create two or more groups of *E*s with different hypotheses, or expectations about the data they would obtain from their *S*s. In those of our studies where *E*s ran human *S*s, their experimental task has been to obtain ratings of photos from their *S*s. These photos were of faces cut from a weekly news magazine and standardized in such a way that most *S*s would normally regard them as occupying a neutral position on a rating scale of success or failure. The actual rating scale employed ran from -10 (extreme failure) to $+10$ (extreme success) with intermediate labeled points. In three different experiments (Rosenthal and Fode, 1961; Fode, 1963) there was a group of *E*s who was told that they would probably obtain mean ratings of $+5$ from their *S*s while another group of *E*s of equal size was told that they would probably obtain mean ratings of -5 from their *S*s. All *E*s read identical instructions to their

*S*s (see appendix). In the three studies, a total of 30 *E*s ran about 375 *S*s. In every one of these studies the lowest mean rating obtained by *any E* expecting high ratings was higher than the highest mean rating obtained by *any E* expecting low ratings from his *S*s. The three p levels were .004, .001, and .004.

Pavlov was aware of the fact that *E*s could influence their animal *S*s. In speaking of experiments on the inheritance of acquired characteristics, he suggested that noted increase in learning ability of successive generations of mice was really more an increase in teaching ability on the part of the experimenter (Gruenberg, 1929, p. 327).

Two studies in experimenter outcome-orientation bias have been conducted using animal *S*s (Rosenthal and Fode, 1960; Rosenthal and Lawson, 1961). In each study, half the *E*s were told that the rats they would be running had been specially bred for brightness, while the remaining *E*s were told that their *S*s had been specially bred for dullness. The actual learning problems for the rats involved both maze learning and Skinner-box situations. In both studies, *E*s believing their *S*s to be bred for brightness obtained better learning from their rats than did *E*s believing their *S*s to have been bred for dullness (ps were .01 and .02).

At the conclusion of one of these studies, we told all *E*s of the nature of the experiment which had lasted the entire quarter. Their reaction was most interesting. When *E*s who had run "dull" rats were told that their *S*s were really not dull at all, their uniform reaction was: "How very interestingly you took in those other *E*s—our rat, however, was obviously *really* dull."

Several other studies, some of which will be discussed later, have also shown the occurrence of the experimenter outcome-bias phenomenon.

Generality of the Phenomenon

The five studies summarized above yield some evidence bearing on the question of the generality of the phenomenon under discussion. Within the framework of Brunswik's (1956) notion of the representative design of experiments, a more complete statement of the sampling domains involved in our research program seems indicated.

A total of twenty studies has been conducted within our research program, of which twelve could be analyzed to determine the occurrence of experimenter outcome-orientation bias. The phenomenon occurred in varying degrees in all twelve studies, the weakest p level being .08, and the median p level being .02.

There have been altogether 250 *E*s (90% of them males) running over 1700 human *S*s (approximately 50% males, 50% females). Fifty different *E*s (85% males) ran 80 rat *S*s, about two-thirds of whom were females. Most of the *E*s and *S*s were attending the University of North

Dakota and the Ohio State University with smaller samples being drawn from two smaller universities in Ohio. Several of the studies done at the University of North Dakota were conducted during Summer Sessions.

Volunteer and non-volunteer populations of *E*s have been drawn from advanced undergraduate courses in experimental, industrial, and clinical psychology and from graduate courses in psychology and education. Volunteer and non-volunteer *S*s have been drawn from introductory psychology courses for the most part, but also from other undergraduate courses in psychology, education, and the humanities. Animal *S*s were varied as to strains and home colonies.

Spatial-temporal characteristics were also varied. Some of the studies lasted only a few days, one lasted several months. Experimental room characteristics varied from a large armory, in which *E*s ran groups of *S*s simultaneously, to small rooms where *E*s ran individual *S*s, which was the more common procedure. Some of these rooms had one-way vision mirrors and microphones in view, others did not.

In view of the wide variety of *E*, *S*, and context domains sampled, we may conclude that experimenter outcome-orientation bias is both a fairly general and a fairly robust phenomenon.[4]

Sources of Experimenter Bias

E-Expectation: As described earlier, we have in general systematically varied *E*s' expectations in most of our experiments. Expectation has been one of our major independent variables and, from the studies reported earlier, it is a clearly significant one. Since statistical significance is no guarantee of practical significance, we may ask what proportion of the variance of *E*s' obtained data is determined by his pre-experimental expectation. Four of the experiments included an expectancy statement by *E*s before they actually ran their *S*s (Rosenthal and Fode, 1960; Rosenthal, Fode, and Vikan-Kline, 1960; Rosenthal, Friedman, *et al.*, in press; Rosenthal, Persinger, Vikan-Kline, and Fode, 1961a). These statements of expectancy were quite restricted in range and clustered closely around the values *E* had already been given as an expectancy by the investigators. Correlations between *E*s' specific expectancies and their subsequently obtained data were computed separately for each treatment group of *E*s. Considering now only the five groups of *E*s who were either unpaid for their participation, or paid only nominally, and not explicity instructed to bias, we find correlations ranging from .31 to .99, with a median Rho of .43. The total N of this group of *E*s was 36 and, for them, expectancy accounted for 18% of the variance of their

[4] A recent well-done master's thesis by Ursula Ekren (1962) utilizing 8 *E*s and 32 *S*s found no outcome-orientation effects determining *S*s' performance on an intelligence test task. In this study there was some question, however, whether individual *E*s actually were aware of the differential expectancies to be induced in them.

subsequently obtained data. Had we corrected for the restriction of range of *E* expectancies caused by our giving them specific expectations, the proportion of variance accounted for would have been at least doubled. Considering now only the five groups of *E*s (N = 30) who were offered more incentive to bias, or who were more explicitly instructed to bias, we find correlations ranging from .00 to −.31 with a median Rho of −.21. While for these *E*s, this obtained correlation accounted for only 4% of the variance (uncorrected for restriction of range) it is significantly opposite to that obtained by our other set of *E*s. This finding will be interpreted in the next section. We may conclude that under the more usual conditions of a psychological experiment, *E*'s expectation determines, to a significant extent, the magnitude of the data he will obtain.

E-Motivation: We saw in the last section that *E*-expectancy interacts with motivation in determining the phenomenon of experimenter bias. Apparently, when *E* is motivated to the point where he feels he is being bribed to bias, or when he is very aware of his own motivation to obtain certain data from *S*s, he tends to show a significant reverse bias effect (Rosenthal and Fode, 1960; Rosenthal, Friedman, *et al.*, in press). In some cases, it may be *E*'s need for autonomy that leads to the reverse bias as though he wanted to say "you can't influence me." In other cases, it may be *E*'s scientific integrity which leads him to obtain data significantly opposite to that hypothesized.

Can anything be said about the relative power of intrinsic *E*-motivation *versus* *E*-expectancy in determining the degree of *E* bias? Two of our studies have some bearing on this question (Rosenthal and Fode, 1960; Rosenthal and Lawson, 1961). These studies employed a group of experimental psychology laboratory students who were all intrinsically motivated to have their animals perform well so that they could complete the experiment and go on to the next one. It was for them, phenomenologically, rather important to get good learning from their *S*s. Yet, in both these studies, those *E*s expecting dull performance from their *S*s obtained dull performance, suggesting that expectancy effects may be more powerful than motivation effects.

Early Data Returns: That the "early returns" of psychological research studies can have an effect on experimenters was noted and well discussed by Ebbinghaus (1885). After saying that investigators notice the results of their studies as they progress, he stated: "Consequently it is unavoidable that, after the observation of the numerical results, suppositions should arise as to general principles which are concealed in them and which occasionally give hints as to their presence. As the investigations are carried further, these suppositions, as well as those present at the beginning, constitute a complicating factor which probably has a definite influence upon the subsequent results." (p. 28). He went on to speak of the pleasure of finding expected data and surprise at obtaining unexpected data and continued by stating the hypothesis of the present study: where "average values" were obtained initially, subsequent data would tend to be less extreme and where "especially large or small numbers are expected it would tend to further increase or decrease the values" (p. 29).

Ebbinghaus was, of course, speaking of himself as both *E* and *S*. Nevertheless, on the basis of his thinking and of more contemporary observations, it was decided to test Ebbinghaus' hypothesis of the effect of early data returns on data subsequently obtained by *E*s.

Twelve experimentally biased *E*s, each running six *S*s on a photo-rating task, were equally and randomly divided into three treatment conditions. One group of *E*s obtained "good" or expected data from their first two *S*s (who were actually accomplices), another group of *E*s obtained "bad" or unexpected data from their first two *S*s (who were also accomplices), while the third group, utilizing only naïve *S*s, served as a control. Comparisons were made of the mean data obtained by *E*s from the last four *S*s run. These *S*s were all naïve and were randomly assigned to *E*s.

Results indicated that *E*s obtaining "good" initial data obtained better subsequent data. *E*s obtaining "bad" initial data obtained worse subsequent data. This effect appeared to be stronger when female *S*s were run. In addition, the effect showed a slight "delayed action effect," appearing to be stronger later in the series of test *S*s run (Rosenthal, Persinger, Vikan-Kline, and Fode, 1961).

How may the effect of early data returns be explained? Early returns probably affect *E*'s expectation of the nature of the data he will subsequently obtain. We have already seen the powerful role played by *E*-expectancy effects. In addition, early data returns tend to affect *E*'s mood as has been autobiographically and disarmingly documented by Griffith (1961).

The improved mood of that *E* who can look forward to having his original (or newly revised) hypothesis successfully confirmed, may lead to more effective biasing, by affecting his behavior toward his *S*s (Rosenthal, Fode, Friedman, and Vikan-Kline, 1960). The darkened mood of that *E* who hypothesized incorrectly, and who did not revise his hypothesis, would likely lead to less effective biasing, by affecting his behavior toward his *S*s. In a later section, we shall summarize what we know about the effect on *E* bias of *E* attitudes and behavior in the *E-S* interaction.

Bias Origin: We have only one experiment specifically dealing with the question of the origin of the bias or the source of the hypothesis (Marcia, 1961). In that study, half the *E*s were allowed to formulate their own hypothesis about the data they would obtain from their *S*s. The remaining *E*s were matched on relevant personality variables and assigned the same hypothesis which the other pair member had evolved for himself.

There was a tendency (not statistically significant) for the *E*s whose hypothesis had been assigned to show greater bias. Jim Marcia, who ran this study, hypothesized that an expectancy induced by a more prestigeful figure might be for *E* more credible than *E*'s own expectancy or hypothesis.

Modeling Effects: In any behavioral experiment, *S* is asked to perform some task. *E* may request *S* to answer some questions, perform verbal or motor exercises, or simply allow his autonomic nervous system to generate a set of functions on a moving tape. The extent to which a given *E*'s own performance of a certain experimental task determines his *S*s' performance of the same task is the extent to which *E* "models" *S* (in his own image). In practice, we can speak of *E* modeling effects (or bias) if a set of *E*s' performances correlate significantly with their randomly assigned *S*s' performances. Modeling effects of *E*s are independent of *E* expectancy or motivation effects (or bias) only to the extent that *E*'s performance of an experimental task is unrelated to his expectancies or wishes regarding the data he obtains from his *S*s.

Eight of our experiments were designed to assess the existence and magnitude of *E* modeling effects. All of these studies employed the photo-rating task described earlier. *E*s themselves rated the photos before running their *S*s. Modeling effects were defined by the correlation between the mean ratings of the photos given by different *E*s and the mean ratings subsequently obtained by *E*s from their randomly assigned *S*s. The number of *E*s per study (and therefore the N per correlation coefficient) ranged from 10 to 26. The number of *S*s per study ranged from 55 to 206. The number of *S*s per *E* ranged from 4 to 20. In all, 145 *E*s ran more than 800 *S*s.

The correlations obtained were remarkably inconsistent: the highest two being +.65 and +.52, and the lowest being −.32 and −.49. Taken individually, only the correlation of +.65 differed very significantly from zero (p = .001).[5] Considered as a set, however, the correlations did differ significantly (p = .01) among themselves. It seems more likely than not that, in different experiments utilizing a person-perception task, there will be significantly different magnitudes of modeling effects which for any single experiment might often be regarded as a chance fluctuation from a correlation of zero.

Somewhat puzzling was the statistically significant (p = .03) trend for the correlations to become more negative in the later-conducted experiments. The only possibly relevant systematic difference between the earlier and later-run experiments was the increasing probability that *E*s suspected themselves to be objects of study. This recognition might have put *E*s on their guard to avoid biasing their *S* with a consequent reversal

[5] Hinkle, D. N., Personal communication, 1961. The data on which this correlation was based were not made available to us for closer study.

of the direction of bias. Evidence for a reversal of bias effect has been put forth elsewhere (Rosenthal, Friedman, *et al.*, in press).

Mediation of Experimenter Bias: The question which we will ask in this section is: Granting the occurrence and some generality of the phenomenon, how does it work? Cheating cannot reasonably account for the observed effect since at least those instances of cheating of which we have become aware, tended on the whole to diminish the biasing effect, as when *E*s who believed their rat *S*s to be dull, prodded them, and, in a few cases, presented fraudulent data. *E*s' data recording and computations were checked, and while the number of net recording and computational errors was very small it was not always randomly distributed. In general, more biased *E*s tended to make more and larger recording and computational errors in the direction of their hypothesis (Rosenthal, Friedman, *et al.*, in press). All of our own calculations were of course based on the raw data rather than on *E*s' computation. If neither cheating nor honest errors could account for our findings, what might?

Verbal Conditioning: The most obvious hypothesis seemed to be some form of verbal conditioning. If an *E* expects to obtain high ratings of photographs, might he not subtly reinforce this type of response? Conversely, if *E* expects low ratings of photos, might he not reinforce subtly those responses which are low? He might be capable of this system of subtle reinforcement even without any implication of dishonesty, for it might be an unintended response on his part. Fortunately, we were able to test this hypothesis. If indeed verbal conditioning were mediating the phenomenon, we might expect to find that biasing increases as a function of the number of photos rated. Certainly, we would not expect to find any biasing on the very first photos rated by *S*s run by different groups of *E*s. There had, after all, been no reinforcement possible prior to the very first response. In a test of this hypothesis (Rosenthal, Fode, Vikan-Kline, and Persinger, 1961) we found, if anything, that biasing decreased over the course of the photo ratings. Furthermore, there was a significant biasing effect in evidence on the first photo alone, thus ruling out verbal conditioning as a necessary mediator or even as an augmentor of the phenomenon. An important implication of this finding is our need to pay special attention, in our search for the mediators of the bias phenomenon, to the brief pre-data-gathering interaction during which *E* greets *S*, seats *S*, "sets" *S*, and instructs *S*.

A subsequent study (Fode, Rosenthal, Vikan-Kline, and Persinger, 1961) was conducted to learn whether operant conditioning *could* drive the ratings of photos up or down according to the will of the *E*. Results showed clearly that this was possible, and that it worked best with certain types of *S*s. We may therefore conclude, that while verbal conditioning is neither a necessary nor a necessarily frequent antecedent of biasing, it nevertheless could be.

Although not strictly relevant to our discussion of mediating factors in the biasing phenomenon, one of our most recent findings may be of interest (Rosenthal, Persinger, Vikan-Kline, and Fode, 1961a). Recent workers have been much concerned with the role of awareness in verbal conditioning (Matarazzo, Saslow, and Pareis, 1960; Levin, 1961). In keeping with our more general paradigm of experimenter bias studies, we had 18 *E*s condition their *S*s to give high positive photo ratings. Half the *E*s were told that their *S*s had personality test scores such that they would afterwards be aware of having been conditioned. The other half of the *E*s were told that their *S*s would not be aware of having been conditioned. All *E*s used identically programmed procedures but those *E*s, expecting their *S*s to be aware, did have significantly more aware *S*s than did the group of *E*s expecting non-awareness. *S*s were not tested for awareness by their *E*s but filled out questionnaires which could be reliably ($r = .98$) scored for awareness, by members of our research group, working under blind conditions.

Modality of Cue Communication: Are visual or verbal cues, such as tone, more important for the mediation of bias? Fode (1960) studied this question by using a group of *E*s behind screens to eliminate visual cues, and a group of *E*s who remained silent throughout the experiment to eliminate verbal cues. He found that verbal cues of tone are probably sufficient to mediate the biasing but that the effect can be greatly augmented by visual cues. Restriction of visual cues accounted for about 80% of the variance of bias magnitude.

In the case of the studies using rat *S*s, the picture is less clear. According to *E*s' self ratings, those who thought their rats were bright tended to handle them more, and more gently, than did the *E*s who thought their *S*s to be dull. In addition, these latter *E*s rated themselves as seeing their *S*s as less pleasant and themselves as less enthusiastic and friendly. We propose very tentatively that *E*s' attitudes towards their animal *S*s were mediated to their animals via their handling patterns, and that their expectations thus were in the nature of self-fulfilling prophecies.

Personal Characteristics of Es and Ss: E's need for approval as measured by the MMPI and by the Marlowe-Crowne-Social Desirability Scale (M-C SD) appears to predict E bias rather well. In five samples (Fode, 1963; Marcia, 1961; Rosenthal, Persinger *et al.*, in press) *E*'s SD score was correlated with his degree of biasing. For a total of 35 *E*s of medium anxiety level (or unselected for anxiety) the obtained correlations averaged .74 ($p = .001$). For a total of 33 *E*s scoring very high or very low on the Taylor Manifest Anxiety Scale (MAS) the analogous correlation was negative but not very statistically significant. Preliminary findings suggest no relationship between *S*s' need for approval and bias-ability.

E's anxiety level (MAS) has been found related to magnitude of *E*'s bias in a remarkably inconsistent manner. One experiment found least anxious *E*s to bias most (Persinger, 1962). A second experiment found medium anxious *E*s to bias most (Fode, 1963). A third experiment found most anxious *E*s to bias most (Rosenthal, Persinger, *et al.*, in press). Since each of these three findings could have occurred by chance only rarely, we are forced to conclude that *E* anxiety is related to *E* biasing, but in an as yet unpredictable manner. A similar conclusion must be drawn from the relationship between *S*'s anxiety and susceptibility to *E* bias. Persinger (1962) found least anxious *S*s, Fode (1963) found medium anxious *S*s, and Rosenthal, Persinger, *et al.* (in press), found most anxious *S*s to be most susceptible to *E* bias. It is of interest to note that in each of the three studies cited, the level of *E* anxiety associated with most biasing is also the level of *S* anxiety associated with most susceptibility to bias. While this finding may be coincidental ($p = .17$), it suggests the possibility that *S*s are more biasable by those biasing *E*s who are most like them in certain personality characteristics.

We had hypothesized that female *S*s might be more biasable by male *E*s because of their role assignment in our culture. In most of our studies sex comparisons were made and typically revealed no sex effect. One of our studies (Rosenthal, Persinger, *et al.*, 1961) did, however, show females to be more biasable than male *S*s.

One of our studies (Rosenthal, Fode, Friedman, and Vikan-Kline, 1960) suggests that *E*s who bias more, but without having been virtually bribed to do so, are seen by their *S*s as more likeable, more personal, more interested, more honest, slower speaking, and more given to the use of hand, head, and leg gestures than are *E*s who bias their *S*s less. These ratings tended to be intercorrelated and a median Rho with degree of bias was therefore computed and found to be .56 (see also Rosenthal and Persinger, 1962). On the other hand, those *E*s who, so to speak, accepted the bribe and did not bend over backwards to avoid biasing, were seen as significantly less honest by their *S*s. Incidentally, *S*s' ratings of their *E*s' "honesty" during the *E-S* interaction predicted significantly well the direction and magnitude of *E*s' subsequent computational errors. Those *E*s who were rated as less honest made more and larger errors in favor of their hypothesis (Rosenthal, Friedman, *et al.*, in press).

Prior acquaintanceship between male *E*s and their *S*s seemed to facilitate the biasing phenomenon. In one experiment this appeared to be true for both male and female *S*s (Rosenthal, Persinger, *et al.*, in press), but in a second study (Persinger, 1962) it appeared true only for female *S*s.

The perceived status of the experimenter was found to be related to his degree of biasing. Vikan-Kline (1962) found higher status *E*s better able to bias their *S*s' responses. We have some preliminary data which suggest that *E*s rated by observers as more professional are the *E*s who bias their *S*s most.

If, on the basis of the data available to date, we were forced to describe the paradigm fostering maximal *E* bias, we would postulate an *E* with a high need for social approval, with an anxiety level neither very high nor very low but similar to the anxiety level of his *S*s. *E* would have high status, be gesturally inclined, and behave in a friendly, interested manner vis-a-vis his *S*s. *S*s might best be acquainted with *E*s and perhaps be female rather than male.

The pattern described might be understood best by considering the *E-S* dyad as a signal exchange system. The signals under discussion are, of course, unintentional. *E*s high in need for social approval may typically be more precise in their signaling behavior. The business of impression management (Goffman, 1956), or signal editing, is more important to them in their everyday life. Their motivation for biasing may also be greater because of their need to please the source of the hypothesis. The high status and friendly manner may serve to focus *S*s' attention on the signal source and increase the likelihood of *E*'s unintentional message being understood.

Methodological Implications of Studies of Experimenter Bias

It seemed reasonable to conclude from our findings that systematic *E* biasing effects might be eliminated by employing as data collectors, research assistants (*A*s) who did not know *E*s' hypothesis or expectancy. Not only did this technique seem logically implied by our data, but it would be practical as well. More and more data collection is actually carried out for *E*s by research assistants. We decided, however, to test the soundness of this methodological suggestion in an empirical manner.

We began by conducting a by now fairly standard experiment in experimenter bias. Fourteen *E*s ran a total of 76 *S*s in the photo rating task with half of the *E*s led to expect +5, and half the *E*s led to expect −5 mean ratings of the success of persons pictured in photos. At the conclusion of this experiment, each *E* was awarded a "research grant" from which he could draw a small salary and also hire two research assistants (*E*'s *A*s). *A*s were randomly assigned to *E*s who then trained and supervised their two *A*s. Each *A* then ran 5 or 6 randomly assigned *S*s of his own (*A*'s *S*s). Unlike the original instructions to *E*s, instructions to *A*s did not inform them as to what perceptions to expect from their *S*s. *E*s, however, were subtly led, by their printed instructions, to expect their *A*s to obtain data of the same sort they had themselves obtained from their own *S*s.

*E*s biased their *S*s, and *E*s' *A*s in turn biased *their* *S*s. The correlation between magnitude of *E*s' bias and their respective pair of assistants' bias was .67 ($p = .01$). Apparently, *E*'s hypothesis or expectancy may be communicated to his research assistants without *E*'s ever telling *A*s the nature of his hypothesis or expectancy.

What methodological suggestions remain then which might serve to reduce or eliminate *E* outcome-orientation effects? For those studies in which it is possible to do so, *E* might eliminate himself and his surrogates from the interaction with *S*s. Automated setups make this feasible for some kinds of behavioral research, but not for others. Any technique of instruction of data-collectors by *E* which would eliminate the possibility of the subtle communication of expectancies from *E* to his *A* would be a methodological improvement. This would be no easy matter and no perfect solution. The too frequent failure of the double-blind method in medical research attests to this. It is a failure not of "double-blindness" but of maintenance of "blindness." During the *E-S* interaction each may learn too much about the other to insure "blindness-maintenance."

Not only because of the danger of bias, but also because of the general nature of *E* effect, it would be desirable to employ samples of *E*s drawn as randomly as possible from a relevant population of relevantly uninformed *E*s. Following Brunswik (1956), this would greatly increase the generality of our findings and thus be of benefit even if *no* bias were ever operating. Alternatively, there may be value in employing samples of *E*s with known distributions of bias as Mosteller (1944) has suggested in the case of interviewers. The particular biases, however, need not be pre-existing ones and it may be useful *purposefully* to induce different biases in our sample of *E*s, giving us better control over the nature and degree of experimenters' biases.

Some Substantive Implications of Studies of Experimenter Bias

The findings presented lead us to a consideration of the sociology and psychology of science and of scientists. But perhaps the most compelling and the most general conclusion to be drawn is that human beings can engage in highly effective and influential unprogrammed and unintended communication with one another. The subtlety of this communication is such that casual observation of human dyads is unlikely to reveal the nature of this communication process. Sound motion pictures may provide the necessary opportunity for more leisurely, intensive, and repeated study of subtle influential communication processes. We have obtained sound motion picture records of 28 experimenters each interacting with several subjects. Preliminary analyses have given us cause to hope that we may be able to learn something of consequence about the mediation of the *E* bias phenomenon in particular, and about subtle communication processes in general. In these films, all *E*s read identical words to their *S*s so that the burden of communication falls on the gestures, expressions, and intonations which accompany the highly programmed aspects of *E*s' inputs into the *E-S* interaction.

The study of the mediation of *E*'s bias via his research assistants has particularly interesting implications for social psychology. It appeared from that study that in a two-party interaction (*A-S*) there may be a non-present third party (*E*) who operates through one of the participants but without necessarily having simply made that participant a surrogate for himself. The participant serving as "carrier" for the non-present influencer may still be able to exert his own influence in a manner additive to the influence of the non-present participant. Furthermore, interpersonal influence once-removed does not appear to be an all-or-none phenomenon. The more a person is able to influence others subtly, the more he seems able to make others carriers of his subtle influence (Rosenthal, Persinger, *et al.*, in press).

Some interesting practical questions arise from these considerations. When an experienced physician or psychotherapist tells the neophyte therapist that the neophyte's patient has a poor or good prognosis, is the experienced clinician only assessing or is he actually "causing" the poor or good prognosis? When the master teacher tells his apprentice that a pupil appears to be a slow learner, is this prophecy then self-fulfilled? When the employer tells the employee that a task cannot be accomplished, has the likelihood of its accomplishment thereby been reduced? More subtly, might these phenomena occur even if the supervisors never verbalized their beliefs? The experiment cited suggests that they may.

Summary

In the normal course of behavioral research, different experimenters (*E*s) often obtain different data from comparable groups of subjects. We may call this *E* effect or gross "error" (Eckler and Hurwitz, 1957). That portion of *E* effect which is non-canceling or specifically predictable we regard as *E* net error or bias. In this paper we have summarized a program of research on *E* bias. We presented experimental evidence for the occurrence and generality of the phenomenon. *E* expectation and motivation were shown to be partial determinants of the results of behavioral research. Verbal conditioning was rejected as a necessary mode of mediation of the biasing phenomenon. *E* and *S* correlates of *E* biasing and *S* susceptibility to *E* bias were summarized. Methodological and substantive implications of our research program were discussed.

REFERENCES

Beecher, H. K. *Measurement of subjective responses: quantitative effects of drugs.* New York: Oxford Univ. Press, 1959.

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

Brunswik, E. *Perception and the representative design of psychological experiments.* Univ. Calif. Press, 1956.

Ebbinghaus, H. *Memory: a contribution to experimental psychology.* (1885) Translated by Ruger, H. A. & Bussenius, Clara E., New York: Teachers College, Columbia Univ., 1913.

Eckler, A. R., & Hurwitz, W. N. Response variance and biases in censuses and surveys. Unpublished manuscript, U. S. Bureau of the Census, June, 1957.

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

FODE, K. L. The effect of experimenters' and subjects' level of anxiety and need for approval on experimenter outcome-bias. Unpublished doctoral dissertation, Univ. No. Dakota, 1963.

FODE, K. L. The effect of non-visual and non-verbal interaction on experimenter bias. Unpublished master's thesis, Univ. No. Dakota, 1960.

FODE, K. L., Rosenthal, R., Vikan-Kline, Linda L., & Persinger, G. W. Susceptibility to influence in a verbal conditioning situation. Unpublished data, Univ. No. Dakota, 1961.

GOFFMAN, E. *The presentation of self in everyday life.* Univ. Edinburgh Soc. Sci. Res. Centre, Monogr. 2, 1956.

GRIFFITH, R. Rorschach water percepts: a study in conflicting results. *Amer. Psychologist*, 1961, *16*, 307–311.

GRUENBERG, B. C. *The story of evolution.* New York: Van Nostrand, 1929.

HYMAN, H. H., COBB, W. J., FELDMAN, J. J., HART, C. W., STEMBER, C. H. *Interviewing in Social Research.* Chicago: Univ. Chicago Press, 1954.

LEVIN, S. M. The effects of awareness on verbal conditioning. *J. exp. Psychol.*, 1961, *61*, 67–75.

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

MASLING, J. The influence of situational and interpersonal variables in projective testing. *Psychol. Bull.*, 1960, *57*, 65–85.

MATARAZZO, J. D., SASLOW, G., & Pareis, E. N. Verbal conditioning of two response classes: some methodological considerations. *J. abnorm. soc. Psychol.*, 1960, *61*, 190–206.

McGUIGAN, F. J. The experimenter: A neglected stimulus object. Paper based on a contribution to the symposium: *On the Social Psychology of the Psychological Experiment*, Henry W. Riecken, Chairman, Amer. Psychol. Assn., New York: Sept., 1961.

MOSTELLER, F. Correcting for interviewer bias. In Cantril, H. *Gauging public opinion.* Princeton Univ. Press, 1944, 286–288.

MULRY, R. C. The effects of the experimenter's perception of his own performance on subject performance in a pursuit rotor task. Unpublished master's thesis, Univ. No. Dakota, 1962.

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.

PERSINGER, G. W. The effect of acquaintanceship on the mediation of experimenter bias. Unpublished master's thesis, Univ. No. Dakota, 1962.

PFLUGRATH, J. Examiner influence in a group testing situation with particular reference to examiner bias. Unpublished master's thesis, Univ. No. Dakota, 1962.

REICHENBACH, H. *The rise of scientific philosophy.* Berkeley & Los Angeles: Univ. Calif. Press, 1951.

RIECKEN, H. W. A program for research on experiments in social psychology. In *Decisions, values and groups.* Vol. II. New York: Pergamon Press, in press.

ROSENTHAL, R. The hypothesis of unconscious experimenter bias. In: an attempt at the experimental induction of the defense mechanism of projection. Unpublished doctoral dissertation, UCLA, 1956, pp. 65–72.

ROSENTHAL, R., & FODE, K. L. The effect of experimenter bias on the performance of the albino rat. Unpublished manuscript, Harvard Univ., 1960.

ROSENTHAL, R & FODE, K. L. The problem of experimenter outcome-bias. In Ray, D. P. (ed.) *Series research in social psychology.* Symposia studies series no. 8, Wash., D. C. Natl. Inst. Soc. Behav. Sci., 1961.

ROSENTHAL, R., FODE, K. L., FRIEDMAN, C. J., & VIKAN-KLINE, LINDA L. Subjects' perception of their experimenter under conditions of experimenter bias. *Percept. mot. Skills*, 1960, *11*, 325–331.

ROSENTHAL, R., FODE, K. L., & VIKAN-KLINE, LINDA, L. The effect on experimenter bias of varying levels of motivation of *E*s and *S*s. Unpublished manuscript. Harvard Univ., 1960.

ROSENTHAL, R., FODE, K. L., VIKAN-KLINE, LINDA, L., & PERSINGER, G. W. The role of verbal conditioning in the mediation of experimenter outcome-bias. Unpublished manuscript, Harvard Univ., 1961.

ROSENTHAL, R., FRIEDMAN, C. J., JOHNSON, C. A., FODE, K. L., SCHILL, T. R., WHITE, C. R., & VIKAN-KLINE, LINDA L. Variables affecting experimenter bias in a group situation. *Genet. Psychol. Monogr.*, in press.

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. Unpublished manuscript, Harvard Univ., 1961.

ROSENTHAL, R. & PERSINGER, G. W. Let's pretend: subjects' perception of imaginary experimenters. *Percept. mot. Skills*, 1962, *14*, 407–409.

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 L., & FODE, K. L. The effect of early data returns on data subsequently obtained by outcome-biased experimenters. Unpublished manuscript, Harvard Univ., 1961.

ROSENTHAL, R., PERSINGER, G. W., VIKAN-KLINE, LINDA L., & FODE, K. L. The effect of experimenter bias and subject set on awareness in verbal conditioning. Unpublished manuscript, Harvard Univ., 1961a.

ROSENTHAL, R., PERSINGER, G. W., VIKAN-KLINE, Linda L., & MULRY, R. C. The role of the research assistant in the mediation of experimenter bias. *J. Pers.*, in press.

VIKAN-KLINE, LINDA L. The effect of an experimenters' perceived status on the mediation of experimenter bias. Unpublished master's thesis, Univ. No. Dakota, 1962.

WHITE, C. R. The effect of induced subject expectations on the experimenter bias situation. Unpublished doctoral dissertation, Univ. No. Dakota, 1962.

WILSON, E. B. *An introduction to scientific research.* New York: McGraw Hill, 1952.

APPENDIX

Sample Instructions to Es and Ss[6]

Instructions to Es. You have been asked to participate in a research project developing a test of empathy. The reason for your participation in this project is to standardize results of experiments of this type. There is the problem in psychological research of different examiners getting somewhat different data on the same tests as a function of individual differences. Therefore, to standardize the tests it is better methodological procedure to use *groups* of experimenters.

You will now be asked to run a series of *S*s and obtain from each ratings of photographs. The experimental procedure has been typed out for you and is self-explanatory.

According to preceding research of this nature, the type of subjects that you will be using have averaged a +5* rating. Therefore, the *S*s you are running should also average about a +5* rating.

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

GOOD LUCK!

Experimental Procedure for Es. In front of you, you will find the instructions you are to read to your *S*s, a sheet of paper for recording each *S*'s rating for each photo and a set of 20 numbered photos.

[6] Instructions varied from study to study. These are presented as fairly typical.
* For half the Es this read "−5."

After recording data from each subject at the top of the recording sheet and reading instructions to the *S*, you are ready to begin.

Take photo #1 and say: "This is photo #1" and hold it in front of the *S* until he tells you his rating, which you will write down on the recording sheet. Continue this procedure through the 20 photos. Do not let any *S* see any photo for longer than 5 seconds.

After each subject, total the ratings of the 20 photos and find the average (mean).

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

We are in the process of developing a test of empathy. This test is designed to show how well a person is able to put himself into someone else's place. I will show you a series of photographs. For each one I want you to judge whether the person pictured has been experiencing success or failure. To help you make more exact judgments you are to use this rating scale. As you can see, the scale runs from -10 to $+10$. A rating of -10 means that you judge the person to have experienced extreme failure. A rating of $+10$ means that you judge the person to have experienced extreme success. A rating of -1 means that you judge the person to have experienced mild failure while a rating of $+1$ means that you judge the person to have experienced mild success. You are to rate each photo as accurately as you can. Just tell me the rating you assign to each photo. All ready? Here is the first photo. (No further explanation may be given although all or part of the instructions may be repeated.)