JOURNAL OF APPLIED BEHAVIOR ANALYSIS 1978, 11, 513-521 NUMBER 4 (WINTER 1978)

THE EFFECT OF WITNESSING CONSEQUENCES ON THE BEHAVIORAL RECORDINGS OF EXPERIMENTAL OBSERVERS

FRANCIS C. HARRIS AND ANTHONY R. CIMINERO[1]

UNIVERSITY OF GEORGIA

The cueing effects of interviewer praise contingent on a target behavior and expectation of behavior change were examined with six observers. Experiment I investigated the effect of cues in conjunction with expectation. Experiment II assessed the relative contributions of cues and expectation, and Experiment III examined the effect of cues in the absence of expectation. The frequencies of two behaviors, client eye contact and face touching, were held constant throughout a series of videotaped interviews between an "interviewer" and a "client". A within-subjects design was used in each experiment. During baseline conditions, praise did not follow eye contact by the client on the videotape. In all experimental conditions, praise statements from the interviewer followed each occurrence of eye contact with an equal number of praises delivered at random times when there was no eye contact. Three of the six observers dramatically increased their recordings of eye contact during the first experimental phase, but these increases were not replicated in a second praise condition. There were no systematic changes in recorded face touching. Witnessing the delivery of consequences, rather than expectation seemed to be responsible for the effect. This potential threat to the internal validity of studies using observational data may go undetected by interobserver agreement checks.

DESCRIPTORS: witnessing consequences, observer bias, reliability, interobserver agreement, human observers

Recent research has shown several ways that observational data may be inaccurate. Possible influencing factors include observer "drift" (Wildman, Erickson, and Kent, 1975), knowledge of experimental hypotheses in conjunction with contingent feedback on recordings congruent with the hypotheses (O'Leary, Kent, and Kanowitz, 1975), observer cheating (O'Leary and Kent, 1973), knowledge of agreement checking (Reid, 1970; Romanczyk, Kent, Diament, and O'Leary, 1973; Taplin and Reid, 1973), predictability of behavior (Mash and McElwee, 1974), complexity of behavior (Mash and Makohoniuk, 1975), and subject reactivity (Mash and Hedley, 1975; Roberts and Renzaglia, 1965; Surratt, Ulrich, and Hawkins, 1969; Zegiob and Forehand, 1978). Under some circumstances, however, subject reactivity has not been found (Dubey, Kent, O'Leary, and Broderick, 1977; Hagen, Craighead, and Paul, 1975; Johnson and Bolstad, 1973; Martin, Gelfand, and Hartmann, 1971; Mercatoris and Craighead, 1974). Such biased data may threaten internal validity and produce results that may be due to observational variables, rather than independent variables. Kazdin (1977) and Kent and Foster (1977) provided comprehensive reviews of the relevant literature on bias.

A potential type of bias that has not received attention concerns the influence of extraneous cues indicating the occurrence of target behaviors. The most obvious examples of these types of cues are the consequences often delivered immediately following target behaviors to the persons being observed. These cues could prompt an observer to score behaviors that would not otherwise have been scored, as in a previous baseline

[1]Reprints may be obtained from Anthony R. Ciminero, Department of Psychology, Universitiy of Georgia, Athens, GA 30602. Appreciation is expressed to Alice L. Harris and I. Daniel Turkat for their assistance in preparing the videotapes.

condition when no consequences were delivered, thus making data reported in experimental conditions artificially inflated or deflated.

The magnitude of this type of bias would, of course, differ from study to study, but if it were operating it could introduce one form of systematic bias. The danger of biased conclusions would be greatest in studies reporting small but systematic differences in behavior between experimental conditions.

The primary purpose of the present series of investigations was to determine the effects of witnessing the delivery of consequences on the recording behavior of observers. Since there is some evidence (O'Leary *et al.*, 1975) that knowledge of experimental hypotheses in conjunction with feedback to the observer can bias behavioral recordings, even though expectation alone does not seem to cause bias (Kent, O'Leary, Diament, and Dietz, 1974), expectancy effects were also investigated.

Experiment I investigated the effect of cues associated with the delivery of consequences in conjunction with expectation. Experiment II assessed the relative contributions of cues and expectation, and Experiment III examined the effect of cues in the absence of expectation.

GENERAL PROCEDURES

Videotape Preparation

A confederate "client" was provided with a written scenario describing several complaints typically associated with sleeping problems. She was instructed to use the scenario to guide her responses during the subsequent interviews.

Taping sessions were conducted in a 4-m by 3-m room furnished with two chairs. The interviewer and client sat facing each other, with the camera (Sony, AVC-3200) positioned behind the interviewer such that the back of his head and the face of the client were in the picture. The interviewer's face was not visible on the tapes to prevent cueing on his mouth movements. The experimenter sat out of camera range and operated the videotape recorder (Sony, AV-8600).

For each 10-min interview, eight 15-sec intervals were randomly designated for the occurrence of each of the following behaviors: (1) client touches her face; (2) client makes eye contact with the interviewer, followed by a praise statement from the interviewer; and (3) interviewer praises for eye contact in the absence of client eye contact. During the taping of each interview, the experimenter used an earphone to listen to an audiocassette on which the numbers 1 to 40 had been recorded at 15-sec intervals. Whenever the number for any of the designated intervals was heard, the experimenter immediately held up signs dictating the behavior of the participants. The client had been instructed previously to move her head about continuously and not touch her face except when instructed to do so. All the praise statements included a description of the behavior being praised (*e.g.*, "I'm glad you just looked at me."). The interviewer did not comment on the client's face-touching behavior.

After taping all 15 interviews, a copy of each tape was made using another videotape recorder (Sony, VO-1600) and a television set that monitored the original recording. None of the praise statements was transferred to the copy. This was accomplished by carefully viewing each original tape and recording the words spoken immediately before a praise statement. As the copy was being made, the recording volume on the machine being recorded onto was turned off just before each praise. It was turned back on immediately after each praise.

Recording System

Observations were conducted using a recording form divided into 40 numbered blocks, each representing a 15-sec interval. Symbols (EC and FT) were printed in each block. An all-or-none recording system was used such that if one or more occurrences of a target behavior was observed, the observer simply slashed through the appropriate symbol. Movement from block to blocked was cued by prerecorded numbers on the audiocassette previously used in the preparation

of the videotapes. Before each observation session, the cassette was rewound as far as possible. The starting of the cassette player was synchronized with a spoken number at the beginning of each videotape to ensure uniform recording conditions across all sessions.

Eye contact was scored when the client positioned her face and head such that she appeared to make eye contact with the interviewer for approximately two or more consecutive seconds. The 2-sec period was approximated by instructing the observers to count, "One thousand one, one thousand two." The trainer modelled this at an approximately standard rate during initial training of each observer. Face touching was scored when any part of the client's hand touched any part of her head between her jaw and hairline.

Interobserver agreement between the criterion observers and between the experimental observers was calculated in two different ways for each session. Overall agreement was determined by taking total agreements (occurrence and nonoccurrences) and dividing by the number of observation intervals. Occurrence agreement was determined by taking occurrence agreements and dividing by occurrence agreements plus occurrence disagreements.

Check on Content of Videotapes

The tapes were made such that the number of eye contacts and face touches was constant across all sessions. To demonstrate this, a pair of experienced criterion observers, one male and one female, who were both experienced in making behavioral observations, observed a complete set of the tapes on which there were no praise statements. Thus, any systematic changes in the experimental observers' recorded levels of the target behaviors could be attributed to witnessing praises and/or expecting a change in behavior.

The criterion observers were given 15 min to study the instructions and definitions. All questions regarding recording procedures and definitions were answered by having the observer reread the appropriate section of the instructions. They were told that their recordings would be compared to each other's to check for agreement. They then observed the first 10-min interview. If 0.80 occurrence agreement was obtained on each behavior they observed the next tape. If not, the observer studied the instructions for 15 additional minutes and rescored the tape. This procedure was continued until all three training tapes had been scored with 0.80 or better occurrence agreement for each behavior. The criterion observers then scored the 12 experimental tapes from which the praise statements had been removed.

The criterion observers' recordings were used to prepare a criterion protocol for each tape. An interval was scored on the protocol for each of the two behaviors if the criterion observers agreed on the behavior's occurrence or nonoccurrence. Intervals in which there was disagreement were excluded from the protocol for the appropriate behavior.

Across sessions, overall agreement between the criterion observers for eye contact ranged from 0.98 to 1.00, with a mean of 0.99. Means for each condition were 1.00, 1.00, 1.00, and 0.98, respectively. Occurrence agreement for eye contact ranged from 0.88 to 1.00, with a mean of 0.98. Means for each condition were 1.00, 1.00, 1.00, and 0.92, respectively. Overall agreement for face touching ranged from 0.95 to 1.00, with a mean of 0.99. Means for each condition were 0.98, 0.99, 1.00, and 0.98, respectively. Occurrence agreement for face touching ranged from 0.75 to 1.00, with a mean of 0.97. Means for each condition were 0.86, 0.99, 1.00, and 0.98, respectively.

Figure 1 shows the mean percentage of intervals scored for each target behavior by the primary criterion observer during each condition. A coin toss determined the choice of the "primary" criterion observer. This distinction was made merely for convenience in plotting the data. Eye contact and face touching were scored at consistent levels, 17.5% to 22.5% for eye contact and 15% to 25% for face touching, across all

sessions. Eye-contact means for each condition were 20%, 22%, 20%, and 20% respectively. Face-touching means were 17%, 23%, 19%, and 23%, respectively.

Experimental-Observer Training

The first three 10-min tapes, with the praises removed, were arbitrarily designated as observer training tapes. All the experimental observers were trained exactly as the criterion observers had been, except that their recordings were compared with the protocol instead of each other's recordings. The experimental observers were told once, at the start of training, that their recordings would be checked against the recordings of several other observers. During training, the experimental observers viewed the three training tapes, in turn, until 0.80 occurrence agreement with the protocol was achieved for each behavior on a tape seen for the first time. If criterion was not met on any tape, that tape was rescored until criterion was met before moving on. Observers 1 to 6 required, respectively, 3, 6, 4, 4, 3, and 4 observations of the first training tape before reaching 0.80 occurrence agreement on both behaviors. All six experimental observers met criterion on the first viewing of the second training tape.

Setting and Standard Procedures

Experimental observation sessions were conducted on 12 consecutive weekdays. A solid partition separated the adjacent observation stations. The experimental observers viewed the tapes in the same sequence in which they were presented to the criterion observers.

EXPERIMENT I

METHOD

Design

An A-B-A-B design was used to assess the effects of interviewer praise contingent on client eye contact in conjunction with the expectation of client improvement on observers' recording behavior. In each baseline (A) phase, experimental observers viewed three of the interviews from which the praises had been removed and recorded occurrences of client face-touches and eye contact with the interviewer. In each expectation plus praise (B) condition interview, the interviewer delivered praise following each occurrence of eye contact and at several other random times when there was no eye contact. During the B conditions, they were told that praise had been effective in increasing eye contact in previous research.

Subjects

Two female undergraduate psychology majors, 22 and 25 yr old, served as experimental observers. They were randomly chosen from a group of four students assigned to participate in the present project as part of their commitment to a summer institute in psychology at the University of Georgia. Students attending the institute received a stipend. None of the four students working on this project had previous experience as a behavioral observer.

Procedure

The experimenter announced each condition of which expectation was a component by stating that there was substantial experimental evidence that contingent praises had been successful in increasing a variety of behaviors, and that the interviewer would be using them to increase the client's eye contact. Subsequent sessions in the expectation plus praise condition were preceded by a shortened version of the original announcement.

RESULTS

Across sessions, overall agreement on eye contact between the experimental observers ranged from 0.80 to 1.00, with a mean of 0.90. Means for each condition were 0.93, 0.87, 0.90, and 0.90, respectively. Occurrence agreement for eye contact ranged from 0.41 to 1.00, with a mean of 0.64. Means for each condition were 0.72, 0.47, 0.64, and 0.72, respectively. Overall agreement on face touching ranged from 0.93 to

1.00, with a mean of 0.97. Means for each condition were 0.98, 0.97, 0.98, and 0.95, respectively. Occurrence agreement on face touching ranged from 0.70 to 1.00, with a mean of 0.87. Means for each condition were 0.89, 0.85, 0.92, and 0.80, respectively.

Figure 1 shows the percentage of intervals scored for each target behavior by each experimental observer. Observer 1 recorded eye contact at a mean frequency of 21% during the baseline condition. There was an immediate increase to a mean of 38.3% during the first expectation plus praise condition. This recovered to the previous level during the second baseline condition. Reinstitution of the expectation plus praise condition did not produce another increase in recorded eye contact. Observer 1 recorded face touching at a consistent level, 17.5% to 24%, across all sessions. Observer 2 recorded eye contact and face touching consistently, 15% to 23% and 15% to 25%, respectively, across all sessions.

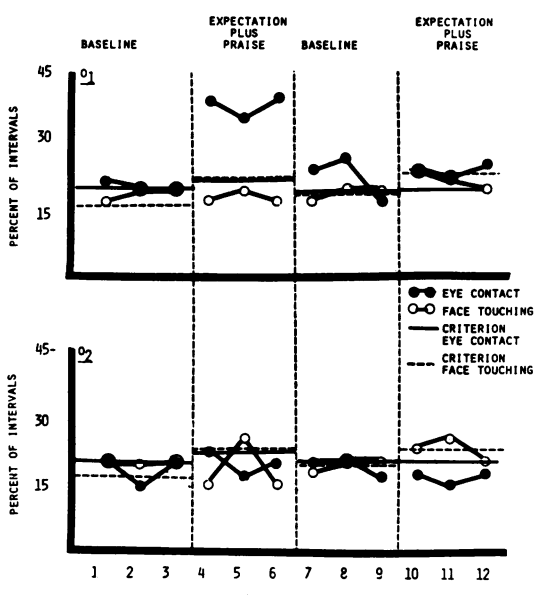

Fig. 1. Percentage of intervals in which Observers 1 and 2 scored eye contact (closed circles) and face touching (open circles). The horizontal (broken and solid) lines represent the mean percentages of intervals in each condition scored by the primary criterion observer.

EXPERIMENT II

METHOD

Design

Based on the results of Experiment I, Experiment II was designed to assess the relative contributions of expectation and witnessing praises to the increased recording of eye contact. An A-B-BC-B design was used to determine if expectation alone would be sufficient to increase recorded eye contact and if witnessing praises would increase recordings over the levels recorded during expectation alone. During the A condition, the observers viewed tapes from which all praises had been removed. During the B conditions, they also observed tapes with no praises, but were given the expectation that the client's eye contact would increase. During the BC condition, they were given the same expectation as in the previous B condition and they observed the tapes that contained correct and incorrect praises.

Subjects

The two undergraduate students participating in the summer institute who were not selected for the previous experiment served as experimental observers. One was a 23-yr-old male and the other a 24-yr-old female.

Procedure

The videotapes, recording system, observer training strategy, and general procedures were the same as those used in Experiment 1. When expectation was used alone, the observers were told that the client had experienced a praise session immediately before the observation session. Before each session in the expectation plus praise condition, the observers were told that the praises would be given during the session instead of before it.

RESULTS

Across sessions, overall agreement on eye contact between the experimental observers ranged from 0.85 to 1.00, with a mean of 0.91. Means

for each condition were 0.93, 0.89, 0.88, and 0.92, respectively. Occurrence agreement for eye contact ranged from 0.40 to 1.00, with a mean of 0.62. Means for each condition were 0.72, 0.50, 0.64, and 0.63, respectively. Overall agreement on face touching ranged from 0.93 to 1.00, with a mean of 0.97. Means for each condition were 0.99, 0.97, 0.99, and 0.94, respectively. Occurrence agreement on face touching ranged from 0.73 to 1.00, with a mean of 0.88. Means for each condition were 0.96, 0.85, 0.96, and 0.78, respectively.

Figure 2 shows the percentage of intervals scored by each experimental observer for each behavior. The data collected by the original criterion observers are shown again in this figure. Except for eye contact in the first session, Observer 3 recorded the target behaviors at consistent levels, 17.5% to 22.5% for eye contact and 17.5% to 25% for face touching, across all sessions. Observer 4 recorded face touching at a consistent level across all sessions. He recorded eye contact at a mean frequency of 17% during baseline.

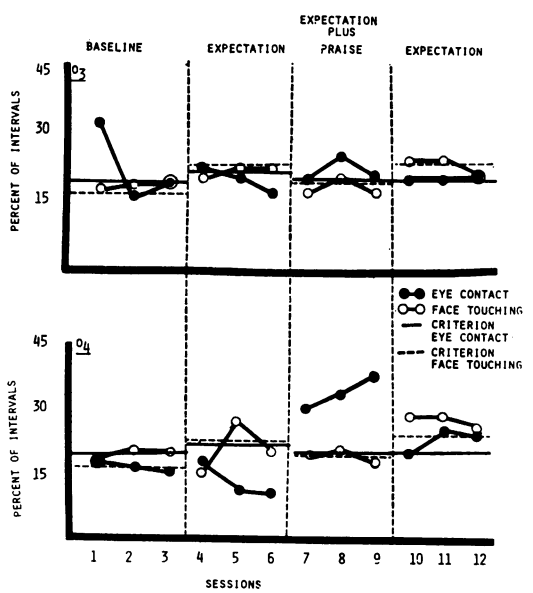

Fig. 2. Percentage of intervals in which Observers 3 and 4 scored eye contact (closed circles) and face touching (open circles). The horizontal (broken and solid) lines represent the mean percentage of intervals in each condition scored by the primary criterion observer.

Although there was a slight change in the direction opposite the expectation for Observer 4, there did not appear to be any systematic change as a result of the expectation procedures. However, there was an immediate increase to a mean of 29% during the subsequent expectation plus praise condition. This recovered to the previous expectation-only level in the final condition in which no praises were witnessed.

EXPERIMENT III

METHOD

Design

Based on the results of the previous experiments, Experiment III was designed to determine if witnessing praises in the absence of outcome expectations would be sufficient to increase the recording of eye contact. An A-B-A-B design was used. During the A conditions, the observers viewed the tapes from which all praises had been removed. During the B conditions, they viewed the tapes on which there were correct and incorrect praise statements.

Subjects

Two male undergraduate psychology majors enrolled in an introductory psychology course were the experimental observers. Each was 22 yr old. Neither had previous experience as an observer.

Procedure

The videotapes, recording system, and observer training strategy were the same as those used in Experiments I and II.

RESULTS

Across sessions, overall agreement on eye contact between the experimental observers ranged from 0.68 to 1.00, with a mean of 0.89. Means for each condition were 0.96, 0.82, 0.88, and 0.90, respectively. Occurrence agreement on eye contact ranged from 0.40 to 1.00, with a mean of 0.61. Means for each condition were 0.78, 0.44, 0.46, and 0.53, respectively. Overall agree-

ment on face touching ranged from 0.93 to 1.00. with a mean of 0.96. Means for each condition were 0.99, 0.97, 0.96, and 0.96, respectively. Occurrence agreement on face touching ranged from 0.70 to 1.00, with a mean of 0.87. Means for each condition were 0.96, 0.85, 0.87, and 0.83, respectively.

Figure 3 shows the percentage of intervals scored by each experimental observer and the criterion observers' data. Observer 5 recorded eye contact at a mean 16.5% level during baseline. This increased to 43.6% during the first praise condition and returned to the previous baseline level during the second baseline condition. Reinstitution of the praise procedures, however, did not produce another increase in recorded eye contact. Observer 5 recorded face touching at a consistent level across all conditions. Observer 6 recorded eye contact and face touching at consistent levels across all conditions.

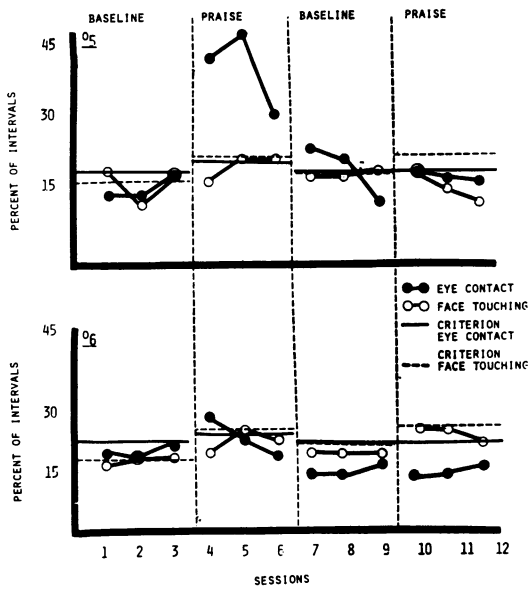

Fig. 3. Percentage of intervals in which Observers 5 and 6 scored eye contact (closed circles) and face touching (open circles). The horizontal (broken and solid) lines represent the mean percentage of intervals in each condition scored by the primary criterion observer.

GENERAL DISCUSSION

The present data suggest that witnessing consequences can influence the behavioral recordings of some observers. This was demonstrated by eye contact above the 20% level being recorded only during praise conditions. Behavior-specific effects of witnessing consequences are suggested by the absence of any systematic changes in recorded face touching. Expectation did not play any apparent role in the altered recordings. This is consistent with previous research (Kent *et al.*, 1974), suggesting that merely expecting a change in behavior does not lead to altered recordings of specific behaviors.

Even though the recordings of only three of the six experimental observers were influenced, it is important to note that the effect on those observers was quite obvious. Inflated eye-contact recordings occurred only during the initial praise condition for the affected observers. Although interval validity was not demonstrated with the affected observers by the variations of the A-B-A-B design, some internal validity was indicated by the fact that the control target behavior, face touching, was not influenced.

The failure to replicate the effect for either subject who showed an initial bias merits elaboration. Since half the consequences were delivered at random times during the interviews, they often occurred when the client was quite obviously looking away from the interviewer and not meeting the definition for eye contact. As time went by, the discrepancy between client behavior and interviewer comments was so great, it probably became obvious and rendered the praises ineffective in changing recording behavior. Informal discussions with the subjects after the study corroborated this speculation. Future research could investigate this phenomenon by having consequences delivered following close or distant approximations to the target behavior.

These results suggest that the interobserver agreement checks might not always detect the biasing effect of witnessing cues. For all three experimental observer pairs, all the condition

means for overall eye contact agreement were above 0.80, even though one member of each pair recorded considerably more behavior than the other during each initial praise condition. This suggests the importance of plotting both observers' data before drawing conclusions and/or using conservative methods (occurrence or nonoccurrence agreement) to determine interobserver agreement (Hawkins and Dotson, 1975). In cases where only one observer was affected by extraneous cues, more conservative methods of calculating interobserver agreement probably would yield differentially low scores in certain conditions and cue the investigator and reader that a problem existed. For example, in the present series of investigations, occurrence agreement was systematically lower in conditions when one member of the pair was affected. In cases such as these when only one member of an observer pair is affected, occurrence agreement would be sensitive to the bias effect. The danger would arise when the primary observer and agreement checker were *both* affected. In that case, the data would be reliable (in terms of agreement scores) but inaccurate in reflecting the levels of the behavior under study.

The reduced interobserver agreement scores following the initial baseline condition for all three observer pairs appear to have been an indirect result of the experimental manipulation, rather than of a qualitative difference between the first three tapes and all the others. All observers were initially trained to 0.80 occurrence agreement with a criterion protocol. Occurrence agreement scores were above 0.70 for the initial baseline conditions but decreased substantially during each initial experimental condition. This did not occur with the criterion observers, thus making it unlikely that there was a difference in actual eye-contact behavior between the first three tapes and all the others. The decrease in occurrence agreement scores following the first baseline condition for all three pairs of experimental observers appears to be a result of only one member of each pair being affected by the experimental manipulation.

It also appears unlikely that inadequate training procedures could have accounted for the reduced interobserver agreement scores following the initial baseline condition. The observer training procedures in this series of experiments appear to be no less rigorous than are commonly reported in the literature. The observers viewed the initial training tape a mean of four times before achieving 0.80 occurrence agreement with the protocol on both behaviors. They all achieved 0.80 occurrence agreement with the protocol on the first viewing of the second training tape. This makes it likely that the present results are due to the independent variables, rather than to poorly trained observers.

The present results suggest that there might be important individual differences in observers with respect to their susceptibility to being influenced by extraneous cues regarding target behavior occurrence. This appears to contradict the widely held assumption (Baer, 1977) that the vast majority of observers, given the same behavioral definition, will be able to agree on the occurrence or nonoccurrence of the behavior regardless of any other factors. Future research could focus on identifying the relevant individual differences.

In many previous studies, no attention has been given to this particular form of bias caused by witnessing consequences. The present results might cast some doubt on conclusions of those studies demonstrating small but consistent differences in behavior between experimental conditions.

Until the parameters of observer reactivity are investigated using different consequences, target behaviors, observers, and experimental subjects, researchers might minimize it by: (1) cautioning observers to attend only to the behavior under study and disregard any extraneous cues; (2) incorporating training in correctly discriminating target behavior occurrence in the presence of consequences during initial observer training and subsequent recalibration sessions; or (3) if feasible, designing the experiment such that observers are never exposed to extraneous cues that

a behavior has occurred. These suggestions are obviously tentative and await empirical validation to document their effectiveness in reducing this type of bias.

REFERENCES

Baer, D. M. Reviewer's comment: just because it's reliable doesn't mean that you can use it. *Journal of Applied Behavior Analysis,* 1977, **10,** 117-119.

Dubey, D. R., Kent, R. N., O'Leary, S. G., and Broderick, J. E. Reactions of children and teachers to classroom observers: A series of controlled investigations. *Behavior Therapy,* 1977, **8,** 887-897.

Hagen, R. L., Craighead, W. E., and Paul, G. L. Staff reactivity to evaluative behavioral observations. *Behavior Therapy,* 1975, **6,** 201-205.

Hawkins, R. P. and Dotson, V. A. Reliability scores that delude: An Alice in Wonderland trip through the misleading characteristics of inter-observer agreement scores in interval recording. In E. Ramp and G. Semb (Eds), *Behavior analysis: areas of research and application.* Englewood Cliffs, New Jersey: Prentice-Hall, 1975. Pp. 359-376.

Johnson, S. M. and Bolstad, O. D. Methodological issues in naturalistic observation: Some problems and solutions for field research. In L. A. Hamerlynck, J. Handy, and D. A. Nash (Eds), *Behavior change: Methodology, concepts, and practice.* Champaign, Illinois: Research Press, 1973. Pp. 7-67.

Kazdin, A. E. Artifact, bias, and complexity of assessment: the ABCs of reliability. *Journal of Applied Behavior Analysis,* 1977, **10,** 141-150.

Kent, R. N. and Foster, S. L. Direct observational procedures: Methodological issues in naturalistic settings. In A. R. Ciminero, K. S. Calhoun, and H. E. Adams (Eds), *Handbook of behavioral assessment.* New York: John Wiley & Sons, 1977. Pp. 279-328.

Kent, R. N., O'Leary, K. D., Diament, C., and Dietz, A. Expectation biases in observational evaluation of therapeutic change. *Journal of Consulting and Clinical Psychology,* 1974, **42,** 774-780.

Martin, M. F., Gelfand, D. M., and Hartmann, D. P. Effects of adult and peer observers on boys' and girls' responses to an aggressive model. *Child Development,* 1971, **42,** 1271-1275.

Mash, E. J. and Hedley, J. Effect of observer as a function of prior history of social interaction. *Perceptual and Motor Skills,* 1975, **40,** 659-669.

Mash, E. J. and Makohoniuk, G. The effects of prior information and behavioral predictability on observer accuracy. *Child Development,* 1975, **46,** 513-519.

Mash, E. J. and McElwee, J. D. Situational effects on observer accuracy: Behavior predictability, prior experience, and complexity of coding categories. *Child Development,* 1974, **45,** 367-377.

Mercatoris, M. and Craighead, W. E. Effects of non-participant observation on teacher and pupil classroom behavior. *Journal of Educational Psychology,* 1974, **66,** 512-519.

O'Leary, K. D. and Kent, R. N. Behavior modification for social action: Research tactics and problems. In L. A. Hamerlynck, L. C. Handy, and E. J. Mash (Eds), *Behavior change: Methodology, concepts, and practice.* Champaign, Illinois: Research Press, 1973. Pp. 69-96.

O'Leary, K. D., Kent, R. N., and Kanowitz, J. Shaping data collection congruent with experimental hypotheses. *Journal of Applied Behavior Analysis,* 1975, **8,** 43-51.

Romanczyk, R. G., Kent, R. N., Diament, C., and O'Leary, K. D. Measuring the reliability of observational data: a reactive process. *Journal of Applied Behavior Analysis,* 1973, **6,** 175-184.

Reid, J. B. Reliability assessment of observation data: A possible methodological problem. *Child Development,* 1970, **41,** 1143-1150.

Roberts, R. R. and Renzaglia, G. A. The influence of tape recording on counseling. *Journal of Counseling Psychology,* 1965, **12,** 10-16.

Surratt, P. R., Ulrich, R. E., and Hawkins, R. P. An elementary student as a behavioral engineer. *Journal of Applied Behavior Analysis,* 1969, **2,** 85-92.

Taplin, P. S. and Reid, J. B. Effects of instructional set and experimenter influence on behavior reliability. *Child Development,* 1973, **44,** 547-554.

Wildman, B. G., Erickson, M. T., and Kent, R. N. The effect of two training procedures on observer agreement and variability of behavior ratings. *Child Development,* 1975, **46,** 520-524.

Zegiob, L. E., Arnold, S., and Forehand, R. An examination of observer effects in parent-child interactions. *Child Development,* 1975, **46,** 509-512.

Zegiob, L. and Forehand, R. Parent-child interactions: Observer effects and social class differences. *Behavior Therapy,* 1978, **9,** 118-123.

Received 20 December 1977.
(Final Acceptance 13 July 1978.)