# COMMENTARY

## On the Psychology of Watching Birds: the Problem of Observer-Expectancy Bias

David F. Balph and Martha Hatch Balph[1]

Studies have shown that observational data, so basic to much of avian research, are subject to systematic error (reviewed by Rosenthal 1966). A type of systematic error that is particularly important for those of us who document the behavior of birds is observer-expectancy bias. Here, we discuss some of what is known about this form of bias and suggest that the problem is complex and difficult to resolve. We hope that our commentary stimulates thought and discussion about what could be true for most of us who collect observational data: "I wouldn't have seen it if I hadn't believed it" (Foster et al. 1975).

The influence upon data of an observer's preconceived notions or wishes has been recognized and investigated at least since the beginning of this century. Much of the research has been in educational and social psychology and has focused upon the psychological environment in which experiments are conducted. Typical of such investigations is Cordaro and Ison's (1963) study of observers who documented the activity of planaria. Some observers were given to believe that their planaria would move and turn frequently, whereas others expected their animals to move and turn infrequently. The high-activity-expectancy group recorded an average of 18 moves and 49 turns during 100 trials, whereas the low-activity-expectancy group recorded an average of only 1 move and 10 turns in 100 trials. The authors concluded that this difference was produced by the groups' differing expectancies. Rosenthal (1969), in reviewing 94 studies of observer-expectancy bias from seven areas of behavioral science, found that about 70% of observers obtained results in the direction of their expectancy. This figure rose to 100% among observers of animal (as opposed to human) subjects. Rosenthal (1969: 236) speculates that in the experimental literature of the behavioral sciences, the effects of the experimental variable are not impressively larger than the effects of observer-expectancy bias.

The magnitude of observer-expectancy bias is determined in part by the kind of observations being made (reviewed by Rosenthal 1969, Salvia and Meisel 1980). If variables (1) are vaguely defined, (2) require subjective assessment, (3) are difficult to perceive, or (4) are psychologically important to the observer, the potential for bias is high.

Interactions between researcher and observer also influence the potential for observer-expectancy bias (reviewed by Rosenthal 1969). For example, if the researcher is perceived as a confident, authoritarian figure, observers are more likely to bias observations in the direction of the researcher's expectancies than if the researcher is not such a figure. Discussions between researchers and observers, or among observers during experiments, about patterns in the data may be particularly influential in creating bias. Kazdin (1977) mentions several studies that did not show observer-expectancy bias until such discussions occurred.

Tests of hypotheses pose a particular dilemma that is not easily resolved. Hypothesis testing in many disciplines is equated with quality research. Although we believe hypothesis testing is a necessary part of progress in science (see Romesburg 1981), this technique nevertheless creates conditions that produce observer-expectancy bias. When the reputation or research support of a researcher depends upon the outcome of the test of a hypothesis and he or she has taken no precautions against bias, the test may be nothing more than a self-fulfilling prophecy—a point occasionally made in the biological literature. The situation is similar to telling students conducting an experiment that they will receive an "A" if they observe their animals behaving in one way and an "F" if they observe something else. Common sense dictates that this situation is likely to promote subjectivity. One must wonder to what extent bias is responsible for the fact that relatively few "serious" hypotheses (as opposed to "straw" or "nuisance" hypotheses) are rejected in the literature. One must also wonder to what degree the expectancies of researchers are responsible for what many of us believe to be true in areas of science (such as ecology and behavior) where hypothesis testing is common, the pressure to obtain desired results is high, and the likelihood of observer-expectancy bias is ignored.

It has been suggested that tests of alternative hypotheses afford protection against personal bias (Chamberlin 1965). We do not entirely agree with this assessment. Researchers usually favor one hypothesis over another or believe that one is more likely to be correct than another, and this preference or belief creates the potential for bias. It also has been suggested that finding agreement between replications, either in experiments (Wiens 1981) or in observers (Salvia and Meisel 1980), attests to the reliability of the work. We strongly support these procedures but believe that repeated tests by researchers who favor a given hypothesis are not independent tests. We agree with Rosenthal's (1966:

[1] Department of Fisheries and Wildlife and the Ecology Center, Utah State University, Logan, Utah 84322 USA.

15) statement that "... replicated observations made under similar conditions of anticipation, instrumentation, and psychological climate may, by virtue of their intercorrelation, all be in error with respect to some external criterion ...."

One might think that the best way to address the problem of bias caused by observer expectancy is to make researchers more aware of the phenomenon. There is some evidence to suggest, however, that this approach will not solve the problem. Salvia and Meisel (1980) assessed the research literature in an area of the behavioral sciences where researchers are aware of the need for precautions against observer-expectancy bias. They concluded that almost 50% of 329 studies had a high potential for such bias and that only 10% took all appropriate precautions.

Such results are not surprising when one considers the socio-psychological conditions under which scientific researchers work. We as researchers are something less than a community of scholars selflessly and cooperatively searching for truth. If we were such a community, we should be willing to publish all of our findings anonymously. Rather, we are individuals caught up in a system that rewards individual productivity—productivity measured not in units of truth, but in such terms as number of publications produced, amount of research money obtained, and amount of prestige garnered. Most of us in the research community must be attentive to these measures, and to varying degrees we may pursue them directly (see Broad 1981).

The motivational state of most researchers is, however, more complex than the simple pursuit of measured productivity. Researchers certainly are taught to appreciate truth in science and learn that it is unpleasant to be discovered in error. The researcher plays what approaches a "zero-sum game," however, in which an increase in effort better to assure truth means a decrease in productivity. Under these conditions, it is unrealistic to expect a researcher voluntarily to increase his or her research standards beyond what is expected by readers (both reviewers and general readers) in the discipline, and currently the ornithological community does not seem concerned about observer-expectancy bias.

One might suppose that, once readers are made aware of observer-expectancy bias, they will insist that researchers address the problem. This may not hold true, however. The problem lies in what we, as readers, find pleasurable. We like to read about results that exhibit patterns, especially simple or "neat" patterns. We readily accept correlation as indicating a causal relationship if the relationship fits an expectancy. We especially enjoy learning of findings that are consistent with our own beliefs. This is why papers or textbooks that point out inconsistencies and present alternative or conflicting views may be less popular than those that present with authority a simple, consistent, and currently accepted view of reality. Unfortunately, reading about truth *per se* is not necessarily rewarding. Thus, the kinds of results one reads about probably are more important in dictating what one accepts as true than the manner in which the results were obtained.

What is perceived to be immediately gratifying, however, may not be to one's benefit over the long term. The reader who forgoes the pleasure of believing an attractive but suspect biological story for skepticism will be rewarded should future research find the work in error. A reader who refuses to be party to building a house of cards in certain areas of ornithology is rewarded by being part of a more vigorous and respected discipline. The acceptance of this broader perspective of self-interest should lead readers to become the critics and skeptics advocated by Wiens (1981). It is such discrimination on the part of readers that will force researchers to be more sensitive to the problem of observer-expectancy bias.

The potential for observer-expectancy bias can be reduced by various means (discussed by Borg and Gall 1979). For example, data collected automatically or in situations in which the observer has no knowledge of the experiment's purpose are likely to be relatively free of observer-expectancy bias. An analogous situation occurs when a hypothesis is tested using data collected for a purpose other than that of testing the hypothesis in question (see Caryl 1979). The potential for observer-expectancy bias also is low in experiments where the variables are well-defined and easily recorded. When an observer is collecting data for an experimenter, bias can be reduced if the experimenter and observer interact as little as possible. Finally, film records may facilitate the detection of bias if present. Comprehensive reviews of ways to reduce observer expectancy and other biases are available in research methodology textbooks in the behavioral sciences (e.g. Tuckman 1978, Borg and Gall 1979). Some of these methods can be applied easily to a variety of avian studies, although others may be difficult, costly, or perhaps impossible to use.

The difficulties of countering the potential for observer-expectancy bias inherent in much of avian research, especially hypothetico-deductive research, will require some patience and discrimination on the part of critical readers. The goal should be an increase in the quality of avian research, not its destruction by unreasonable demands. Fortunately, there is plenty of room for compromise. For example, it may be unreasonable to expect a graduate student to conduct a "blind" experiment by hiring and training observers to collect data for a project about which they have no knowledge; but it might be reasonable to expect that the test variables be clearly defined. It may be necessary for a researcher to record a subjectively assessed variable under poor lighting conditions, but the results may be sufficiently conclusive to override the influence of observer-expectancy bias. At the very least, readers could expect authors to include statements about the risk of observer-expectancy bias when reporting on observational data and to qualify their conclusions accordingly.

One point that should promote understanding and compromise in all of this is that many readers are also researchers. What we decide to accept as the former, we may eventually have to live with as the latter.

We thank Alan C. Kamil and an anonymous reviewer for their helpful comments on the manuscript.

## LITERATURE CITED

BORG, W. R., & M. D. GALL. 1979. Educational research—an introduction, third ed. New York, Longman, Inc.

BROAD, W. J. 1981. The publishing game: getting more for less. Science 211: 1137-1139.

CARYL, P. G. 1979. Communication by agonistic displays: what can game theory contribute to ethology? Behaviour 68: 136-169.

CHAMBERLIN, T. C. 1965. The method of multiple working hypotheses. Science 148: 754-759.

CORDARO, L., & J. R. ISON. 1963. Psychology of the scientist: X. Observer bias in classical conditioning of the planarian. Psychol. Repts. 13: 787-789.

FOSTER, G. G., J. YSSELDYKE, & J. REESE. 1975. I wouldn't have seen it if I hadn't believed it. Exceptional Children 41: 469-473.

KAZDIN, A. E. 1977. Artifact, bias, and complexity of assessment: the ABC's of reliability. J. Appl. Behav. Anal. 10: 141-150.

ROMESBURG, H. C. 1981. Wildlife science: gaining reliable knowledge. J. Wildl. Mgmt. 45: 293-313.

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

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

SALVIA, J. A., & C. J. MEISEL. 1980. Observer bias: a methodological consideration in special education research. J. Special Educ. 14: 261-270.

TUCKMAN, B. W. 1978. Conducting educational research, second ed. New York, Harcourt Brace Jovanovich, Inc.

WIENS, J. A. 1981. On skepticism and criticism in ornithology. Auk 98: 848-849.

*Received 2 April 1982, accepted 21 March 1983.*