(3rd ed., Vol. 1, pp. 441–86). New York: Random House.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasiexperimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago, IL: Rand-McNally. Reprinted as *Experimental and quasi-experimental designs for research.* Chicago, IL: Rand-McNally, 1966.

Crano, W. D., & Brewer, M. B. (1986). *Principles and methods of social research.* Boston, MA: Allyn & Bacon.

WILLIAM D. CRANO

experimenter effects The term experimenter effect refers to the effect of the experimenter on the results of the research. Experimenter effects are *artifacts* or threats to the validity of the study.

Certain experimenter effects such as observer, interpreter, and intentional effects occur without investigators actually affecting their subjects' responses to the experimental task. These effects occur in the mind, the eye, or the hand of the investigator. Other effects such as biosocial, psychosocial, situational, MODELING, and EXPECTANCY EFFECTS are interactional. These effects occur when some attributes or behavior of the experimenter actually affect subjects' responses to the experimental task.

OBSERVER EFFECTS

Observer effects, unintentional inaccuracies in the observation and recording of the events being studied, were first detected by astronomers. Early in the nineteenth century, Bessel, an astronomer at Konigsberg studied the observations of stellar transits made by senior astronomers. He found that differences in observation occurred surprisingly often (Boring, 1950).

Things do not seem to have changed much in the present century: in an analysis of 21 studies involving 314 observers and 139,000 observations, about 1 percent of the observations were found to be in error. Further, approximately 66 percent of these errors were found to occur in the direction of the experimenter's hypothesis (Rosenthal, 1991). Observer effects can be reduced by exercising great care in data collection. Such effects might also be reduced by the use of multiple independent observers so that observers' errors cancel each other.

INTERPRETER EFFECTS

Interpreter effects occur when data are inaccurately interpreted. Interpreter effects, however, might be less serious than observer effects because the former are public whereas the latter are private. Researchers can agree or disagree on the interpretation of observations (and they often do, as can be seen from the debates in technical journals) but they generally do not have access to the observations themselves.

Some common interpreter effects in the social sciences can be reduced by improved research and data-analytic methodology. For example, many people doubted the effectiveness of psychotherapy until Smith and Glass (1977) demonstrated on the basis of a quantitative, comprehensive analysis of the literature on psychotherapy that there is strong, convincing evidence for the effectiveness of psychotherapy. Until recently, researchers could impose a wide range of interpretations on vast bodies of literature depending on their own personally preferred theory. But now a variety of meta-analytic procedures allow for systematic and quantitative summaries of vast research domains that tend to increase the objectivity of the conclusions drawn (Rosenthal, 1991; *see* META-ANALYSIS).

INTENTIONAL EFFECTS

Intentional effects include dishonest reporting and fabrication of data. Although instances of fraud and dishonesty are probably relatively rare, intentional effects should nevertheless be regarded as effects of the investigators themselves. Self-serving errors in citations of the literature also fall under the category of intentional effects.

Intentional effects have been demonstrated in many of the sciences (Broad & Wade, 1982). Perhaps the most notorious example of

intentional
is the case
separate rep
50 pairs of
exactly 0.7
ally, a virt
cannot be e
ated his da
in recordin
Intention
cisely beca
plementatio
ethical trai
prevent su

BIOSOCIA
Biosocial e
tributes as
investigato
the direct e
investigato
these attri
tigator. Tl
acteristics.
certain m
result of
searchers
toward th
evidence
conduct t
ently anc
results. M
subject ca
which m
the subje
possible l
studies a
strategy
from dil
biosocial

PSYCHC
Psychos
ality of
characte
need fc
warmth
differen
perimer
obtain

intentional effects in the behavioral sciences is the case of the late Cyril Burt. In three separate reports of over 20, over 30, and over 50 pairs of twins, he reported a correlation of exactly 0.771 for all three studies: statistically, a virtually impossible finding! But we cannot be entirely sure whether Burt fabricated his data or whether he was just careless in recording the data (an observer error).

Intentional errors are hard to prevent precisely because they are planned. Careful implementation of ethical codes of conduct and ethical training of researchers might help to prevent such errors.

BIOSOCIAL EFFECTS

Biosocial effects are associated with such attributes as the age, gender, and race of the investigator. It is difficult to disentangle the direct effect of the biosocial attributes of the investigator on subjects from the influence of these attributes on the behavior of the investigator. Thus, because of their biosocial characteristics, researchers might be *perceived* in a certain manner by their subjects, or as a result of their biosocial characteristics, researchers might *actually behave* differently toward their subjects. For example, there is evidence that male and female experimenters conduct the "same" experiment quite differently and therefore might obtain different results. Moreover, biosocial attributes of the subject can affect the experimenter's behavior which might in turn affect the responses of the subject. Researchers should be alert to possible biosocial effects when they plan their studies and conduct the research. A good strategy might be to compare results obtained from different experimenters to check for biosocial effects.

PSYCHOSOCIAL EFFECTS

Psychosocial effects result from the personality of the experimenter. Experimenters' characteristics such as their level of ANXIETY, need for approval, their status, and their warmth all influence the results of studies in different, unpredictable ways. Thus, experimenters who are higher in status tend to obtain more conforming responses from subjects and warmer examiners tend to obtain better intellectual performance on standardized tests of intelligence than do cooler examiners (Rosenthal, 1969). As in the case of biosocial effects, psychosocial effects should be considered when studies are being planned and should be checked for by comparing the results obtained from different experimenters when data are being analyzed.

SITUATIONAL EFFECTS

Many situational effects are possible, e.g., effects of the level of experience of the investigator or effects of the degree of acquaintance between the subject and the experimenter. Thus, experienced experimenters may get different results from inexperienced ones. Similarly, experimenters who know their subjects sometimes get different responses from experimenters who do not know their subjects. Further, events that happen during the course of the experiment can lead to changes in subjects' responses. Researchers should be alert to the possibility of situational effects when they plan their studies and should also examine their data for possible situational effects (e.g., by comparing results obtained from experimenters with different levels of experience).

MODELING EFFECTS

There is some evidence that in a variety of research contexts research subjects tend to respond to a task in the same way their experimenter would respond to it. As in the case of biosocial, psychosocial, and situational effects, investigators should consider possible modeling effects in planning their research and, to minimize such effects, should, if possible, use multiple experimenters.

EXPECTANCY EFFECTS

Last, experimenter expectancy effects occur when investigators' hypotheses and expectations regarding how the research will turn out influence the way in which the research does turn out. The hypothesis held by the experimenter leads unintentionally to behavior toward the subjects which, in turn, increases the likelihood that the hypothesis will be confirmed. This phenomenon is also called a SELF-FULFILLING PROPHECY (Rosenthal, 1966).

A classic case

One of the earliest demonstrations of experimenter effects occurred in the case of Clever Hans (Pfungst, 1911). Hans was the horse of Mr Von Osten, a German mathematics teacher. By tapping his foot, Hans could add, subtract, multiply, divide, spell, read, and solve problems of musical harmony. Unlike other performing animals, Hans had not been trained to perform by his owner. And unlike other performing animals, he was able to perform consistently well even in the absence of his owner. Intrigued by Hans' skills, Pfungst and his colleague Stumpf undertook a program of systematic research to discover the secret of Hans' talent. They discovered that Hans could not answer questions that his questioners could not answer. Further, Hans could not answer questions if he could not see the questioner. Based on these observations, Pfungst and Stumpf reasoned that Hans must be picking up cues from the questioner as to when to begin and when to stop tapping his foot. Indeed, a forward inclination of the head would start Hans tapping and as the questioner straightened up Hans would stop tapping – even a slight motion of the eyebrows or nostrils was sufficient to stop Hans from tapping. Clearly, Hans was able to pick up the expectations of his questioners and thereby fulfill their prophecies.

Experiments with human subjects

In the first demonstration of experimenter expectancy effects in the laboratory, ten advanced undergraduate and graduate students of psychology served as the experimenters. Each experimenter was assigned 20 introductory psychology students as his or her subjects. The experimenter showed a series of ten photographs of people's faces to each subject individually and asked the subject to rate the degree of success or failure shown in the face of each person pictured in the photograph. The photos had been selected so that on the average they were quite neutral. All ten experimenters were given identical instructions regarding the task and identical instructions to read to their subjects. They were told not to deviate from these instructions. They were also told that the purpose of the experiment was to see how well they could duplicate well-established experimental results. Half the experimenters were told that the "well-established" finding was that their subjects should rate the photos as being of successful people and half the experimenters were told that the finding was that the subjects should rate the photos as being of unsuccessful people. Results showed that experimenters expecting "successful" photo ratings did obtain higher photo ratings than did the subjects expecting unsuccessful photo ratings (Rosenthal & Fode, described in Rosenthal, 1966). Because all the experimenters read the same instructions to the subjects it was not clear how exactly the experimenters were able to influence their subjects. But an ingenious experiment provided an explanation for this subtle influence process.

Adair and Epstein (described in Rosenthal, 1969) replicated the basic photo rating experiment on the self-fulfilling effects of experimenters' hypotheses. But during the replication experiment, they tape-recorded experimenters' instructions to their subjects. They then conducted a second experiment using the tape-recordings of experimenters' voices instead of real experimenters. They found that when the tape-recorded instructions had originally been read by experimenters expecting failure perception by their subjects, the tape-recordings evoked greater failure perceptions from the subjects. Apparently the experimenters' tone of voice influenced subjects to perceive success or failure in correspondence with the experimenters' own expectations.

Experiments with animal subjects

Experimenter expectancy effects have also been found with infra-human subjects. For example, Rosenthal and Fode (described in Rosenthal, 1966) told 12 students in an experimental psychology class that studies had shown that maze-brightness and maze-dullness could be developed in strains of rats by successive inbreeding of well performing and poorly performing maze runners. Half the experimenters were told that their rats were

maze-brigl
their rats
assigned f
to run tc
elevated T
changeabl
equally of
results re
better per
formers. A
ers' rating
experimen
better pe
brighter,
than expe
expect w
perimente
ported th
more gen
expecting

Expectanc

The majo
pectancy
laboratory
revealed t
real world
earliest fi
experimer
ary schoo
ligence c
"blooming
children
chosen at
condition
teachers.
childrens'
tual bloo
make sign
ance in t
8 months
the childr
test and t
lieved to
greater g
children
Jacobson,
The cu
provide u
ence of ir
Rosentha
have imp

maze-bright and the other half were told that their rats were maze-dull. Each student was assigned five rats. The animals had to learn to run to the darker of two arms of an elevated T-maze. The two arms were interchangeable and the rewarded arm was located equally often on the right as on the left. The results revealed that animals believed to be better performers actually became better performers. After the experiment, experimenters' ratings of their animals indicated that experimenters who had been led to expect better performance rated their animals as brighter, more pleasant, and more likeable than experimenters who had been led to expect worse performance. Further, experimenters expecting better performance reported that they had handled their rats in a more gentle manner than did experimenters expecting poor performance.

Expectancy effects beyond the laboratory. The majority of studies on interpersonal expectancy effects have been conducted in the laboratory. But a number of studies have revealed that the phenomenon exists in the real world as well. For example, in one of the earliest field experiments, the "Pygmalion" experiment, all of the children in an elementary school were administered a test of intelligence disguised as a test to predict "blooming." Approximately 20 percent of the children from each of 18 classrooms were chosen at random to be in the experimental condition and their names were given to their teachers. The teacher was told that these childrens' performance on a test of "intellectual blooming" indicated that they would make significant gains in intellectual performance in the next 8 months of school. After 8 months, at the end of the school year, all the children were retested with the same IQ test and the children whom the teachers believed to be "bloomers" showed significantly greater gain in performance than did the children in the control group (Rosenthal & Jacobson, 1968).

The cumulative results of over 450 studies provide unequivocal evidence for the existence of interpersonal expectancy effects (e.g., Rosenthal & Rubin, 1978). These findings have important substantive implications for

Table 1

Strategies for the Control of Experimenter Effects

1. Increasing the number of experimenters
   - decreases learning of influence techniques
   - helps to maintain blindness
   - minimizes effects of early data returns
   - increases generality of results
   - randomizes expectancies
   - permits the method of collaborative disagreement
   - permits statistical correction of expectancy effects
   - decreases the likelihood of biosocial, psychosocial, situational, and modeling effects
2. Observing the behavior of the experimenter
   - sometimes reduces expectancy effects
   - permits correction for unprogrammed behavior
   - facilitates grater standardization of experimenter behavior
3. Maintaining blind contact (see table 2)
   - minimizes expectancy effects
4. Minimizing experimenter–subject contact (see table 3)
   - minimizes biosocial, expectancy, modeling, psychosocial, and situational effects
5. Employing expectancy control groups
   - permits assessment of expectancy effects
6. Analyzing experiments for order effects
   - permits inferences about changes in experimenter behavior
7. Analyzing experiments for computational errors
   - reduces effects of interpreter effects
   - permits inference about expectancy effects
8. Developing selection procedures
   - permits prediction of expectancy effects
   - minimizes biosocial, psychosocial effects
9. Developing training procedures
   - permits prediction of expectancy effects
   - minimizes situational and modeling effects
10. Developing a new profession of psychological experimenter
    - maximizes applicability of controls for all experimenter effects
11. Conducting meta-analyses
    - minimizes interpreter, observer, and intentional effects.

investigators. Perhaps the most striking implication is that we engage in highly influential unintended COMMUNICATION with one another and that this process of unintentional influence can affect our experimental results. Consequently, individuals who conduct experiments, such as research assistants, should be kept blind to the hypotheses under investigation so that their expectations will not influence the results of the research.

A number of strategies to prevent experimenter effects including expectancy effects are presented in table 1. Table 2 suggests methods to promote blind contact in order to control for experimenter expectancy effects. Table 3 suggests methods to minimize contact between experimenters and subjects.

Table 2

Blind Contact as a Control for Expectancy Effects

A. Sources of breakdown of blindness
   1. Principal investigator
   2. Subject.

B. Procedures for facilitating maintenance of blindness
   1. The "total blind" procedure
   2. Avoiding feedback from the principal investigator
   3. Avoiding feedback from the subject

Table 3

Minimized Contact as a Control for Expectancy Effects

A. Automated data collection systems
   1. Written instructions
   2. Tape-recorded instructions
   3. Filmed instructions
   4. Televised instructions
   5. Telephoned instructions
   6. Computerized instructions

B. Restricting unintended cues to subjects and experimenters
   1. Interposing screen between subject and experimenter
   2. Contacting fewer subjects per experimenter
   3. Having subjects or machines record data

CONCLUSION

It has been our task to describe the nature of experimenter effects as a source of artifacts in behavioral research. The recognition of these sources of artifact should not result in readers feeling discouraged. We agree with Herbert Hyman's comment that the demonstration of systematic error may well mark an advanced state of science.

All scientific inquiry is subject to error, and it is far better to be aware of this, to study the sources in an attempt to reduce it, and to estimate the magnitude of such errors in our findings, than to be ignorant of the errors concealed in the data. One must not equate ignorance of error with the lack of error. The lack of demonstration of error in certain fields of inquiry often derives from the nonexistence of methodological research into the problem and merely denotes a less advanced stage of that profession (Hyman, 1954, p. 4).

See also: COMMUNICATION; META-ANALYSIS; SELF-FULFILLING PROPHECIES.

BIBLIOGRAPHY

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

Broad, W., & Wade, N. (1982). *Betrayers of the truth*. New York: Simon & Schuster.

Hyman, H. H. (1954). *Interviewing in social research*. Chicago; IL: University of Chicago Press.

Pfungst, O. (1911). *Clever Hans (The horse of Mr. Von Osten)*. New York: Henry Holt. (Reissued 1965).

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

——(1969). Interpersonal expectations. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 181–277). New York: Academic Press.

——(1991). *Meta-analytic procedures for social research* (rev. ed.). Newbury Park, CA: Sage.

——& Jacobson, L. (1968). *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston.

——& Rubin, D. B. (1978). Interpersonal expectancy effects: The first 345 studies. *Behavioral and Brain Sciences, 3*, 377–86.

Smith, M. I
analysis o
American

eyewitness
liability of
"pillars" of
century (*see*
Wells & Lo
Kassin, 198
able contrib
century, the
terest in th
with the ap
and the first
efforts.

If an eye
and become
have to co
reliability
testimony.
the one ha
igated perc
relate to th
report. Ap
other hand
improveme
Both grou
cerned wit
present en
will focus
facilitation
verge on sc

PERCEPTI
An accura
first instan
ating an e
attend to t
at the actu
distance of
ent in the
(e.g., sens
differentia
impairmen
entailing a
ate the d

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist, 32*, 752–60.

NALINI AMBADY
ROBERT ROSENTHAL

eyewitness testimony Research into the reliability of eyewitnesses has been one of the "pillars" of legal psychology for almost a century (*see* Lösel, Bender, & Bliesener, 1992; Wells & Loftus, 1984; Wrightsman, Willis, & Kassin, 1987). Following a first peak of notable contributions in the early decades of the century, the late sixties brought renewed interest in the field. The turning point came with the application of experimental methods and the first tentative theories and integrative efforts.

If an eyewitness report acquires legal status and becomes trial evidence, jurors and judges have to come to a decision concerning the reliability as well as the veracity of the testimony. Experimental psychologists on the one hand have for the main part investigated perceptual and memory processes that relate to the accuracy and completeness of a report. Applied legal psychologists on the other hand have been more interested in the improvement of interrogation techniques. Both groups have been and are still concerned with the problem of deception. The present entry follows that division. First it will focus on memory processes, then discuss facilitation of recall and finally it will converge on some aspects of lie detection.

PERCEPTION

An accurate witness report requires in the first instance reliability of perception. Evaluating an eyewitness account therefore has to attend to the interaction of situational factors at the actual time of the episode (e.g., light, distance of target) with the limitations inherent in the sensory equipment, be they general (e.g., sensory thresholds, adaptation time) or differential (e.g., nearsightedness, hearing impairment). An example of a perceptual law entailing an error is the tendency to overestimate the duration of an observed sequence. Constructive processes can also play an important part. For instance a spatial and temporal coincidence of stimulus events can lead to the inference of causality and the attribution of intention.

MEMORY

Analyses of MEMORY processes in eyewitness research usually employ the tripartite system of an acquisition, retention and retrieval stage.

ACQUISITION/ENCODING

Witness factors such as prior knowledge or special encoding strategies (e.g., depth of processing, giving verbal labels to actions or persons) have instigated considerable research. Among the event characteristics that are most likely to improve encoding accuracy are long exposure time and familiarity with the situation, complexity of the situation and frequency of the observed act. Some hold that a high stress level and/or a high level of crime seriousness are detrimental to accuracy, due to a reduced attention span (as in a state of "weapon focus," in which the witness fails to perceive the offender's face or other relevant information, because his or her attention is constricted to a threatening object). Among the significant characteristics of the target person are attractiveness, sex and race (CROSS-RACIAL FACIAL IDENTIFICATION being more difficult).

RETENTION/STORAGE

An assumption of memory decay obviously entails an emphasis on storage time. It has consistently been shown that witnesses remember more details when interrogated immediately than after a delay. Another research focus has been the effect of memory stemming from postevent information. If it originates in an external source, such postevent information amounts to leading questions or suggestive communications. Loftus (1979) was among the first to investigate such influences on memory during the storing phase. She used a series of coloured slides depicting an accident. In different experimental conditions one group of subjects saw a yield sign,