Psychological Review
1986, Vol. 93, No. 2, 216-229

0033-295X/86/$00.75

Under What Conditions Does Theory Obstruct Research Progress?

Anthony G. Greenwald
Ohio State University

Anthony R. Pratkanis
Carnegie-Mellon University

Michael R. Leippe
Adelphi University

Michael H. Baumgardner
Burke Marketing Services
Cincinnati, Ohio

Researchers display confirmation bias when they persevere by revising procedures until obtaining a theory-predicted result. This strategy produces findings that are overgeneralized in avoidable ways, and this in turn hinders successful applications. (The 40-year history of an attitude-change phenomenon, the sleeper effect, stands as a case in point.) Confirmation bias is an expectable product of theory-centered research strategies, including both the puzzle-solving activity of T. S. Kuhn's "normal science" and, more surprisingly, K. R. Popper's recommended method of falsification seeking. The alternative strategies of condition seeking (identifying limiting conditions for a known finding) and design (discovering conditions that can produce a previously unobtained result) are result centered; they are directed at producing specified patterns of data rather than at the logically impossible goals of establishing either the truth or falsity of a theory. Result-centered methods are by no means atheoretical. Rather, they oblige resourcefulness in using existing theory and can stimulate novel development of theory.

Imagine looking at a projected photographic image that is so badly focused that identification is impossible. The picture is gradually focused until it is just slightly blurred, at which point you are asked to guess what it is. Without the gradual focusing, you might identify the slightly blurred picture about 75% of the time. However, with prior exposure to the more blurred image, you can correctly identify it only about 25% of the time. Interpreting this finding, both Wyatt and Campbell (1951) and Bruner and Potter (1964) suggested that subjects' preliminary hypotheses, formed on the basis of early, poor data, interfered with effective interpretation of later, better data.

The Wyatt–Campbell and Bruner–Potter findings provide striking illustrations of the pervasive phenomenon of confirmation bias—the tendency for judgments based on new data to be overly consistent with preliminary hypotheses. The range of demonstrations of confirmation bias includes primacy effects in impression formation (Asch, 1946; Luchins, 1957) and persuasion (Lund, 1925; Miller & Campbell, 1959); delayed discovery of simple problem solutions (Luchins, 1942; Mynatt, Doherty, & Tweney, 1977; Wason, 1960); expectancy biasing of pupil achievement (Rosenthal & Jacobson, 1968); perseverance of belief in discredited hypotheses (Anderson, 1983; Nisbett & Ross, 1980; Ross, Lepper, & Hubbard, 1975); and selective retrieval of information that confirms one's hypotheses (Snyder & Uranowitz, 1978), one's opinions (Pratkanis, 1984; Ross, McFarland, & Fletcher, 1981), or one's self-concept (Mischel, Ebbesen, & Zeiss, 1976; Swann & Read, 1981). In addition, confirmation bias is manifest in the many effects that can be described in terms of Merton's (1948) concept of self-fulfilling prophecy (Darley & Fazio, 1980; Festinger, Riecken, & Schachter, 1956; Snyder & Swann, 1978). As is indicated by the variety of these illustrations, confirmation bias is a very general phenomenon: One's preliminary hypotheses have a decided advantage in the judgment process.

In contrast with this picture of a widespread human trait of confirmation bias, consider the familiar stereotype of the scientist as an impartial observer whose hypotheses stand or fall according to the blind justice of objective data. This stereotype may have a kernel of truth, but the evidence indicates that scientists, like other humans, frequently operate in confirmation-biased fashion (e.g., Armstrong, 1982; Brush, 1974; Feynman, 1985; Gould, 1981; Kuhn, 1970; Lakatos, 1976; Mahoney, 1977; Mitroff, 1974; Orne, 1969; Popper, 1934/1959; Rosenthal, 1966; Westfall, 1973). The scientist's confirmation bias is the main subject of the present article, in which we (a) document the obstructive potential of researchers' confirmation biases, (b) examine the psychological and social functions that (nevertheless) maintain those biases, and (c) identify research strategies that can remedy these problems of confirmation bias in science.

After a prefatory note on theory, we give an overview of the history of an empirical phenomenon, the sleeper effect. The authors' involvement in research on the sleeper effect motivated the methodological analysis that follows; that analysis, in turn, provides the basis for research strategy recommendations that are presented and analyzed in the remaining sections.

Preparation of this report was in part supported by Grants SOC 74-13436, BNS 76-11175, and BNS 82-17006 from the National Science Foundation and by Grant MH 323117 from the National Institute of Mental Health, all entitled "Research in Persuasive Communication."

For their comments on an earlier draft, the authors thank Robert P. Abelson, Craig A. Anderson, Norman H. Anderson, J. Scott Armstrong, Steven J. Breckler, Jack W. Brehm, Donald T. Campbell, Lee J. Cronbach, Reid Hastie, Edward E. Jones, William J. McGuire, Judson Mills, and Susumu Yamaguchi.

Correspondence concerning this article should be addressed to Anthony G. Greenwald, Ohio State University, 404C West 17th Avenue, Columbus, Ohio 43210.

Prefatory Note on Theory

These preliminary comments on theory are intended to forestall some unintended interpretations of the authors' views on the usage of theory and the usefulness of theories.

Although some confine the term to formal axiom-based systems, *theory,* as used here, includes also much more modest *statements that express relationships among concepts.* Concepts are distinguished from *operations,* which are explicit procedures for fabricating or measuring events. Concepts and operations are linked by *rules of correspondence,* which associate a single concept to multiple operations. Research is conducted by means of operations but commonly (indeed, necessarily) described in terms of corresponding concepts—that is, in terms of theories. For example, "presentation of reward contingent on a response increases the probability of that response" should be understood as a highly theoretical statement in terms of concepts of reward, contingency, probability, and response. In contrast, "presentation of food to a pigeon 0.2 s after its depression of a key increases the rate of its depression of the key" is an attempt to describe the corresponding operations in a much less theoretical fashion. Verbal descriptions of procedures necessarily include concepts. However, the concepts can be chosen—as in the latter of the two preceding statements—to reduce surplus meaning (implied generalization). A statement much longer than the latter one, referring to characteristics of the apparatus, the pigeon, the food, and the time and location of the events in question, would have been still less theoretical. In present terms, then, a statement is theoretical to the extent that it generalizes beyond the operations that support it. Theory necessarily entails the risk of overgeneralization; two main concerns of this article are the common occurrence of theoretical overgeneralizations and the possibility of avoiding them.

This article does conclude, as advertised by its title, that theory can obstruct research progress. However, the authors also believe that theory is necessary to the progress (indeed, the process) of research. These views are not inconsistent. It will be argued that theory is likely to obstruct research progress when the researcher's primary goal is to test the theory. In testing a theory, the theory can dominate research in a way that blinds the researcher to potentially informative observations. When existing theory is less sacred, because the researcher's goal is other than theory testing, the researcher is more likely (a) to attend to the operations on which research findings depend and, consequently, (b) to discover theories of increasing power.

History of the Sleeper Effect

The First 35 Years (1943–1978)[1]

A sleeper effect is an effect that takes some time to become apparent, in other words, a delayed-action effect. One particular sleeper effect, a delayed persuasive impact following a communication accompanied by a discounting cue (i.e., information indicating that the communication is untrustworthy), has received particular research attention and is often referred to as *the* sleeper effect. This ("the") sleeper effect was first identified and named by Hovland, Lumsdaine, and Sheffield (1949) on the basis of their research on opinion changes produced by the *Why We Fight* films that were made and used by the U.S. Army during World War II. (Reviews of the broader group of delayed persuasive impact studies can be found in McGuire, 1985, and Pratkanis, 1981.)

The sleeper effect is one of a number of findings that play a special role in psychology texts and courses—the role of the interesting quirk. An interesting quirk, to start with, is counterintuitive, and it therefore attracts the student's attention. When the instructor or writer manages to explain this oddity in terms of more general principles, the student is duly impressed with the power of psychology to go beyond common sense.

Given its pedagogic function, one is tempted not to subject an interesting quirk to close scrutiny. The sleeper effect, indeed, was not examined very closely for about two decades after it had secured a place in psychology texts. However, it did receive the attention of several researchers during the 1970s. That attention revealed that the sleeper effect is an empirically troublesome phenomenon and that its initial evidential basis was unsatisfactory.

Discovery of the sleeper effect. Hovland et al. (1949 [research done in early 1943]) investigated the effects of *The Battle of Britain,* a film that was intended to increase the confidence of U.S. Army recruits in the fighting ability of their British allies. Hovland et al. reported that when the effect of this film was assessed 9 weeks after it was shown, some of the opinion items showed significantly more improvement in regard for the British than had been apparent at an earlier posttest that was given just 5 days after the film. Hovland et al. called this finding a "sleeper effect."

Explanation and acceptance. As part of a program of research at Yale University that grew out of the Hovland group's World War II research efforts, laboratory attention was given to the sleeper effect. Three published experiments (Hovland & Weiss, 1951; Kelman & Hovland, 1953; Weiss, 1953) contributed to a conclusion in favor of a cue/content dissociation explanation, one of four interpretations that had earlier been proposed by Hovland et al. (1949). The dissociation interpretation is based on two assumptions: (a) The association between the communication and discounting cue weakens with time, and (b) the association of discounting cue to topic weakens more rapidly than does the association of message content to topic.

In their influential 1953 volume, *Communication and Persuasion,* Hovland, Janis, and Kelley reviewed the Yale sleeper-effect studies and advocated the dissociation interpretation. Despite a suggested alternative interpretation by Festinger (1955), this support for the dissociation interpretation prepared the sleeper effect to play the interesting-quirk role, which it played in most social psychology texts published between 1955 and 1975.

1971–1978: The sleeper effect questioned. While one of the present authors was engaged in a frustrating attempt to use the sleeper effect to test a cognitive response theoretical analysis of persuasion (eventually published in Gillig & Greenwald, 1974), Cook (1971) circulated a review of research on the sleeper effect. In that review, Cook introduced a distinction between an absolute sleeper effect (a significant increase in a communication's impact after a delay) and a relative sleeper effect (significantly slower decay of the impact of a message accompanied by a discounting

[1] For a more thorough review of this history, see Pratkanis (1981).

cue than of one accompanied by an accepting cue or no cue). Also, Cook made the startling observation that there really had never been any satisfactory evidence for the (absolute, or proper) sleeper effect; the widely cited Yale studies had demonstrated only the relative sleeper effect. (See Capon & Hulbert, 1973, for a similar assessment of the early sleeper effect literature.)

On encountering Cook's observations about the inadequacy of the sleeper effect evidence, Gillig and Greenwald (1974) converted their investigation into an attempt to find the (absolute) sleeper effect. The Gillig and Greenwald findings, based on seven separate experimental tests, produced clear evidence for the relative sleeper effect but none for the sleeper effect proper.

Recent History of the Sleeper Effect (1978–1984)

Use of new methods to test the sleeper effect. Perhaps Gillig and Greenwald's experiments had not found a sleeper effect because their procedures—even though more powerful than those of previous sleeper-effect experiments—were nevertheless not powerful enough. That observation led to some efforts at technical development. Ronis, Baumgardner, Leippe, Cacioppo, and Greenwald (1977; see also Baumgardner, Leippe, Ronis, & Greenwald, 1983) reported an initial series of experiments using a laboratory computer to construct a message-dense environment that made it possible to observe decay of persuasive messages' impacts within a single laboratory session. Initial applications of this method to the search for the sleeper effect, conducted in 1977, were unequivocally negative. In four replications, the relative sleeper effect was repeatedly found, but there was clearly no absolute sleeper effect. This initial research is reported in more detail as part of a report that includes studies conducted at Ohio State University over a 7-year period (Pratkanis, Greenwald, Leippe, & Baumgardner, 1985).

1978: A reliable sleeper effect reported. Concurrently with the research being done at Ohio State University, Cook and his colleagues had also undertaken a search for the sleeper effect. Their search was based on technical improvements and conceptual analysis that were first described in Cook's 1971 paper. Cook's analysis specified the following requirements for a sleeper effect (quoted from Gruder et al., 1978):

> (a) a persuasive message [must have] substantial initial impact on attitudes; (b) this change is totally inhibited by a discounting cue; (c) the cue and message are dissociated over time; and (d) the cue and message are dissociated quickly enough so that the message by itself still has some impact when dissociation occurs. (p. 1074)

Gruder et al.'s (1978) study obtained a sleeper effect and, importantly, replicated it. They concluded that when the above conditions are implemented, one obtains a sleeper effect.

However, it remained disturbing that some of the studies that appeared to have satisfied Cook's conditions for a test of the sleeper effect had not produced the effect. In particular, it was not clear why the seven tests by Gillig and Greenwald or the four initial tests by Pratkanis et al. (1985) did not find a sleeper effect. In trying to solve this puzzle, the present authors' attention was drawn to an aspect of Gruder et al.'s procedure that differed from most previous sleeper effect studies: Gruder et al. had presented their discounting cue (the information that led subjects to distrust the message) *after* subjects had read the message. That might have been a critical aspect of their procedure. Accordingly, the present authors conducted further studies with the message-dense procedure, varying whether the discounting cue information was presented before or after the message.

Empirical resolution: Timing of discounting cue is critical. Figure 1 synthesizes into one graph the results of 16 experiments with the message-dense procedure, conducted between 1977 and 1984, that included 26 discounting-cue treatments. (These studies are reported by Pratkanis et al., 1985.) The experiments summarized in Figure 1 tested several procedural variations for their possible impact on the sleeper effect. Only one variation made a difference: whether the discounting cue information came before or after the persuasive message. Of course, there were many differences between the multimessage, one-session, computer-controlled procedure of the studies in Figure 1 and the single-message, two-session procedure with which Gruder et al. (1978) had produced a sleeper effect. Accordingly, conclusive evidence for the importance of timing of source information in producing the sleeper effect might be obtained by replicating Gruder et al.'s procedure with an added condition in which the discounting cue precedes the message. Pratkanis et al. (1985) conducted such a replication, adding to Gruder et al.'s design the critical condition in which the discounting cue preceded the persuasive message. Gratifyingly, Pratkanis et al. replicated Gruder et al.'s sleeper

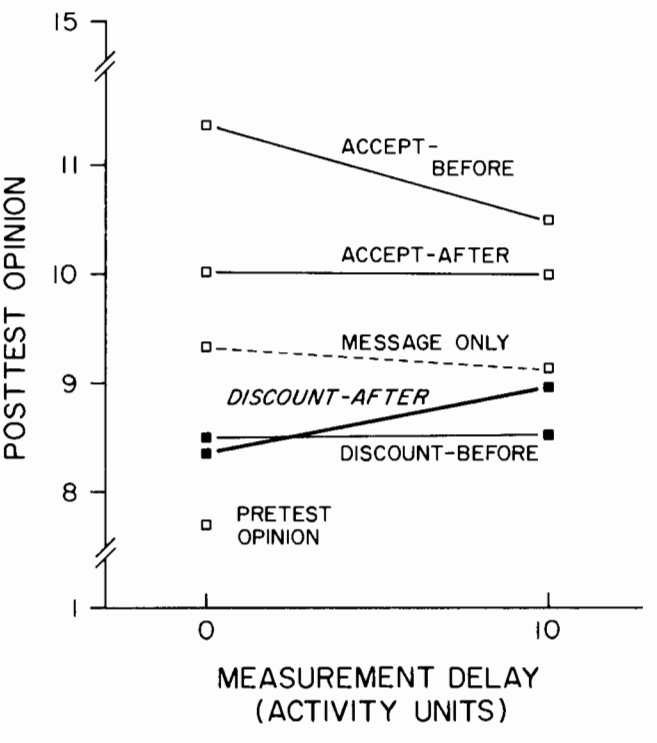

Figure 1. Summary figure for 16 sleeper-effect experiments (Pratkanis et al., 1985). (The units of the abscissa are activities such as reading messages or responding to opinion items on other topics, which intervened between presentation of a message and assessment of its effect on opinion. The time interval for 10 such units is typically less than 10 min. Each experiment included several or all of the delay intervals between 0 and 10 units. The plotted data are average linear regression slopes, weighted by the number of observations in each experiment in which the condition was used.)

effect with their cue-after-message procedure but obtained no sleeper effect when the discounting cue instead preceded the message.

Theoretical interpretation: Differential decay of persuasion. Pratkanis et al. proposed a differential decay interpretation to account for the dependence of the sleeper effect on timing of the discounting cue (see Figure 2). According to this interpretation, the cue-after-message treatment produces the sleeper effect when conditions are such that (a) the message has an initial strong persuasive impact and (b) the discounting cue immediately and effectively opposes this impact but (c) this countereffect of the discounting cue decays more rapidly than does the impact of the message. The sleeper effect is thus produced by a decaying impact of persuasive messages, not by an increasing impact. Importantly, though, there are two impacts, which oppose one another and which decay at different rates. In contrast, when the discounting cue precedes the message, the initial interpretation of the message is assumed to be strongly affected by the cue, such that (a) initial persuasive impact is considerably attenuated, (b) the message and source are likely to form a unit in memory, and (c) therefore the analysis of Figure 2, in terms of separate decaying impacts, does not apply.

Theory as an Obstruction to Research

History of the Sleeper Effect in Retrospect

There were several instances of confirmation bias in the sleeper effect research. First, researchers and textbook writers prematurely adopted a single explanation (the cue-content dissociation hypothesis) to the exclusion of other alternatives, such as the several suggested by Hovland et al. Second, the empirical criterion of the effect was switched to the relative sleeper effect pattern, a change that helped to preserve the dissociation interpretation's viability. Third, until the Cook (1971) review, researchers and reviewers emphasized theory-confirming findings and tended to overlook null results. Fourth, the dependence of Gruder et al.'s (1978) replicated sleeper effect on an unusual procedure (i.e., on presenting the discounting cue after the persuasive message) was initially not detected.

The research done at Ohio State University was not confirmation biased in the usual sense. (Indeed, it might be suspected of being falsification biased.) It was nevertheless fixated (through 1982) on a single theoretical hypothesis, the dissociation hypothesis. In the course of the research at Ohio State University, there was an attempt to improve methods of persuasion research, while continuing to test the dissociation hypothesis. These developments in method were useful. However, their inadequacy as a solution to the sleeper effect problem was made clear by Gruder et al.'s (1978) discovery of a reliable sleeper effect.

In retrospect, the problem was less in the methods being used to answer the question of the sleeper effect's existence than in the way the question itself had been asked. In particular, both the group of researchers at Ohio State University and the Cook–Gruder group had been asking "Does the sleeper effect exist?" and were identifying the answer to that question with an evaluation of the dissociation theory. The post-1978 sleeper-effect research has made it apparent that a much superior question would have been "Under what conditions can the sleeper effect be obtained?" This latter question guided the final phase of the present authors' research on the sleeper effect. Had this question been asked sooner, the discovery that the sleeper effect can be produced by presenting the discounting cue after the message might well have been achieved much sooner.

Thus, the latest in the series of methodological developments stimulated by research on the sleeper effect (Cook, 1971; Cook, Gruder, Hennigan, & Flay, 1979; Greenwald, 1975, 1976; Ronis et al., 1977) is an analysis of researchers' methods of posing their research questions. Research questions can be—and often are—stated in a theory-centered fashion that, as will presently be shown, is prone to avoidably overgeneralized conclusions.

Overgeneralization and the Theory-Confirming Research Strategy

Perhaps the most generally admired research strategy in any scientific discipline is that of testing theories. However, this admirable strategy is easily misused to produce nearly useless research conclusions. To appreciate this, consider the enterprise of testing a theory in terms of a series of tactical choices by the researcher (see Figure 3).

The theory-testing approach runs smoothly enough when theoretically predicted results are obtained. However, when predictions are not confirmed, the researcher faces a predicament that can be called the *disconfirmation dilemma* (Greenwald & Ronis, 1981). This dilemma is resolved by the researcher's choosing between proceeding (a) as if the theory being tested is incorrect (e.g., by reporting the disconfirming results), or (b) as if the theory is still likely to be correct. The researcher who preserves faith in the theory's correctness will persevere at testing the theory—perhaps by conducting additional data analyses, by collecting more data, or by revising procedures and then collecting more data.

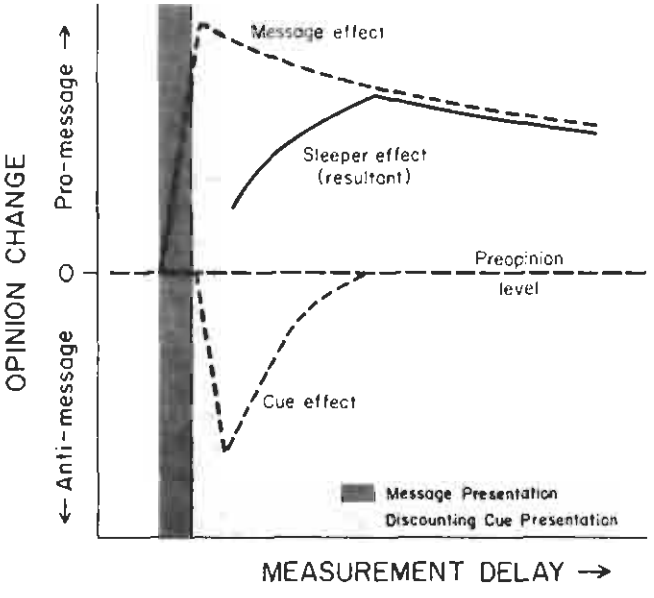

Figure 2. Differential decay interpretation of the sleeper effect. (The discounting cue has immediate counterpersuasive impact, but this negative effect decays more rapidly than does the positive effect of the message.)

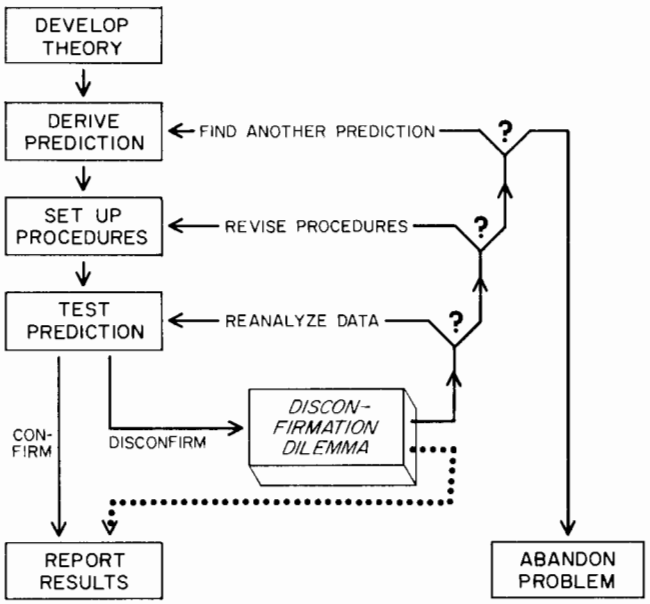

Figure 3. Theory-centered research method, showing the disconfirmation dilemma. (Question marks indicate choices subsidiary to resolving the disconfirmation dilemma; the resolution leading toward "report results" is dotted to indicate its infrequent use. When the researcher repeatedly resolves the disconfirmation dilemma by retesting the prediction rather than reporting results, the theory-centered strategy is one of theory-confirming, rather than theory-testing.)

It is possible for the disconfirmation dilemma to be resolved repeatedly by preserving faith in the guiding theory. When the researcher's faith in the theory cannot be shaken by disconfirming data, it is inappropriate to describe the research strategy as theory testing. Rather, the strategy is effectively one of theory confirming.[2]

A theory-confirming researcher perseveres by modifying procedures until prediction-supporting results are obtained. Particularly if several false starts have occurred, the resulting confirmation may well depend on conditions introduced while modifying procedures in response to initial disconfirmations. However, because no systematic empirical comparison of the evolved (confirming) procedures with earlier (disconfirming) ones has been attempted, the researcher is unlikely to detect the confirmation's dependence on the evolved details of procedure. Although the conclusions from such research need to be qualified by reference to the tried-and-abandoned procedures, those conclusions are often stated only in the more general terms of the guiding theory. Such conclusions constitute avoidable overgeneralizations.

Examples of (Avoidably?) Overgeneralized Conclusions

Published reports rarely provide the information needed to determine whether a researcher was engaged in theory confirming rather than theory testing. However, in Greenwald's (1975) survey, many researcher–respondents did report practices that correspond to the theory-confirming approach, such as (a) persevering by modifying procedures when an initial hypothesis test produces a disconfirming result and (b) being reluctant to prepare reports of disconfirming findings for publication.[3] The histories of many research problems show periods during which it appears (at least in the light of hindsight) that researchers labored to produce avoidably overgeneralized conclusions. The following cases, starting with the sleeper effect, come from diverse areas of research in order to make clear that no subdiscipline of psychology has a monopoly on the theory-confirming strategy. (Nor is there any reason to believe that theory-confirming research and avoidably overgeneralized conclusions are confined to psychology; see Brush, 1974, and Feynman, 1985, pp. 340 ff., for examples from physics.)

The sleeper effect. The differential decay interpretation of the sleeper effect could have surfaced as early as 1959, when Miller and Campbell used a similar principle to analyze persistence functions for the impact of a sequence of two opposing persuasive communications. (See Ronis, 1980, for a similar analysis.) As suggested previously, the 25-year delay in reaching the present understanding can be credited to researchers' collective fixation on attempts to test the dissociation interpretation of discounting cue effects.

The insufficient justification effect. It was predicted from cognitive dissonance theory (Festinger, 1957) that one should like a course of action more if one has chosen to follow it with a small (and seemingly insufficient) inducement rather than a large (and certainly adequate) inducement. After the report of a confirming result by Festinger and Carlsmith (1959), it took about 13 years of very active research controversy (see, e.g., Janis & Gilmore, 1965; J. M. Nuttin, 1975; Rosenberg, 1965) to establish severe limiting conditions for this insufficient justification effect (Calder, Ross, & Insko, 1973; Collins & Hoyt, 1972). The 13 years between discovery and establishment of limiting conditions included more concentrated research effort and journal space than have been commanded by any other social psychological phenomenon, before or since. In retrospect, that research was of great value in establishing a new and influential account of the effects of incentives on human motivation and cognition. Nevertheless, there was a high ratio of wasted to effective effort on the insufficient justification problem as (a) researchers formed into several theoretical camps (dissonance, learning theory, self-perception, and impression management); (b) each camp published own-theory-confirming results; (c) many findings remained unpublished and uninterpreted because they did not easily fit with any of these theoretical positions; and (d) consequently the social psychology community was delayed in discovering the limiting conditions of published findings.

The law of effect. Thorndike (e.g., 1933) was the first to claim support for the theory—the law of effect—that rewards have an automatic effect of strengthening the stimulus–response relations on which they are contingent. For several decades the law of effect was widely treated as *the* basic principle of the psychology

[2] This situation was discussed by Greenwald (1975) as a case of "prejudice against the null hypothesis."

[3] It is tempting to credit researchers' reluctance to prepare reports of disconfirming findings to journal editors' prejudices against publishing such reports. However, it seems that authors rarely provide the occasion for those prejudices to be exercised.

of learning. The overgeneralizations that supported faith in the law of effect as a principle of human learning were uncovered only through the intensely persevering research efforts of a few critics. Human learning results that were once credited to automatic effects of rewards were eventually attributed to a variety of other factors (see Buchwald, 1969; J. R. Nuttin, 1953; Postman, 1963). After falling from favor as a principle of human learning, the law of effect nevertheless persisted in the domain of animal learning for another decade or so. Eventually, however, the "standard" reward effects, which had been obtained with bar pressing by rats and key pecking by pigeons, were demonstrated to have been inappropriately overgeneralized to apply to any (arbitrarily selected) response that an organism can generate. A more accurate view of the effects of rewards is that they work reliably only for limited categories of stimuli and responses that vary from species to species (Bolles, 1975).

Imitative aggression. Starting with the work of Bandura and Walters (1963), many laboratory studies demonstrated that observation of aggression is likely to produce imitative aggression. These consistent results, coupled with field studies that found correlations of aggressive behavior with frequency of viewing violent television programs, prompted the U.S. Public Health Service (1972; National Institute of Mental Health, 1982) to conclude that violence on television is a significant contributor to antisocial aggressive behavior. Remarkably, some elaborately conducted field experiments provided no support for this conclusion (e.g., Feshbach & Singer, 1971, Milgram & Shotland, 1973). Perhaps these field experiments did not replicate laboratory findings because the laboratory studies typically included procedures that had no counterparts in ordinary exposures to media violence. (Most notably, the laboratory studies usually provided the subject–viewer an opportunity to act aggressively in a *permissive* setting very shortly after exposure to televised violence.) The once widely accepted principle that observation of aggression typically causes imitative aggression appears now to be an overgeneralized conclusion from the relevant research. (See Freedman, 1984, for a recent review that supports a conclusion of, at most, very limited imitative effects of witnessing violence. Of course, even very limited imitation of observed violence can be a cause for great social concern.)

Other problem areas. There are many other problem areas in which it is plausible that avoidable overgeneralizations have the status of well-established conclusions. In seeking to identify such areas, one might look especially to (a) topics for which there is a strongly expected pattern of findings and (b) topics for which the publication of some nonreplications makes clear that researchers do not yet understand the conditions needed to produce the phenomenon in question.

Literature reviews. Avoidable overgeneralization can occur among literature reviewers who deal with others' findings, just as much as among researchers in regard to their own findings. Consider, for example, that in the second paragraph of this article, the authors were at no pains to take note of empirical phenomena inconsistent with the principle of confirmation bias. (Have we strengthened or weakened our thesis by observing that we omitted evidence inconsistent with it?) Reviews that have attempted to document race differences, sex differences, and genetic transmission of traits are often assumed to have overgeneralized their conclusions. Of course, reviews that have argued against those very same phenomena are also likely to have overstated their cases in confirmation-biased fashion.

Inevitability of Overgeneralized Conclusions

Whenever a conclusion based on replicable empirical findings is stated, there will necessarily be alterable conditions on which that conclusion depends. That is, any research finding, when summarized as a conclusion, is necessarily overgeneralized. This inevitability of overgeneralization follows jointly from (a) the necessary incompleteness of the researcher's control or observation of variables in the research setting and (b) the impossibility of describing, in a summary statement, the levels of all possibly relevant variables to which it applies.

Figure 4 illustrates the reasons for the inevitability of overgeneralized conclusions. As shown in the figure for the case of a sleeper-effect experiment, any experiment involves an indefinite number of contextual variables. Many of these are left uncontrolled by the experimenter, and many others are controlled, most by being set (often arbitrarily) at fixed levels and some by strategies such as randomization or counterbalancing. In contrast, only a few variables are focal, in the sense of being deliberately varied between treatments (independent variables) or measured as properties of the experimental units (dependent variables).[4]

A complete description of the results of an experiment would include a description of the values of all focal *and* contextual variables. Such a description is, of course, not possible. Typically, the description is confined to an observed configuration of data for the focal (independent and dependent) variables, perhaps mentioning one or a few prominent contextual variables. Such a conclusion statement implicitly generalizes across all levels of the (indefinite number of) unmentioned contextual variables. This generalization across contextual variables is inevitably in error. That is, there must be conditions, among the infinity of unexamined contextual variations, under which the result does not hold.

History of the Sleeper Effect in Prospect

An important aspect of scientific progress consists of giving increasingly complete accounts of the conditions that limit known findings. In the case of the sleeper effect, progress after Hovland et al. (1949) consisted of limiting the sleeper effect to the condition of accompanying the communication with discounting information (Hovland & Weiss, 1951), further limiting it to the condition of a message that was strong enough to have residual impact by the time the discounting cue's impact had dissipated (Gruder et al., 1978), and most recently limiting it further to the condition in which the discounting information is presented after the persuasive message (Pratkanis et al., 1985). Prospectively, it is certain that the conditions for obtaining the sleeper effect can be further limited. (And, for reasons that will be developed in the next section, it is equally certain that the conditions for obtaining the sleeper effect can be expanded.)

---

[4] For simplicity this accounting of variables does not separately recognize contextual variables that are confounded with focal variables. These might also be understood as misidentifications or construct invalidity of the focal variables (cf. Cook & Campbell, 1979; Cronbach & Meehl, 1955).

CONTEXTUAL VARIABLES		FOCAL VARIABLES	
UNCONTROLLED	CONTROLLED	INDEPENDENT	DEPENDENT
CURRENT EVENTS CLIMATE TIME OF DAY • • • • • • • •	SUBJECT CHARACTERISTICS (randomized) ROOM CUES EXPERIMENTER MESSAGE TOPIC AND CONTENT PROCEDURE "DETAILS" • •	DISCOUNTING CUE (present vs. absent) MEASUREMENT DELAY (short vs. long)	OPINION ON MESSAGE TOPIC

Figure 4. Conceptual domain of the sleeper-effect experiment. (The assumption that an observed relationship among the focal [independent and dependent] variables has no particular dependence on the levels of controlled or uncontrolled contextual variables is inevitably in error.)

Should/Can Anything Be Done?

Is the Social System of Research Adequately Self-Correcting?[5]

Perhaps one need not be concerned about researchers' biases toward confirming their preferred theories. After all, scientific disciplines are social systems; their participants can be expected to have complementary and, therefore, mutually compensating theoretical biases (see Campbell, 1986, for a discussion of this topic).

Unfortunately, however, theoretical competition cannot be counted on to produce empirical resolution. Consider, as an example of a topic for which there is much theoretical controversy but no empirical resolution, the topic of psychic (extrasensory) phenomena. Most laboratory researchers on psychic phenomena are advocates.[6] Not surprisingly, then, the great majority of empirical publications on psychic phenomena report confirming results. More generally, fringe topics—those that are marginal to accepted, established domains—are researched almost exclusively by advocates who are grinding theoretical axes (e.g., speed reading, lie detection, biorhythms, and subliminal influence). Additionally, there is an abundance of nontraditional procedures that are claimed to have therapeutic effect (e.g., proposed techniques for weight control, anxiety reduction, pain relief, or amelioration of drug abuse) for which virtually all research is done by the technique's small group of proponents.

Even within established, accepted domains of research, greater rewards of research support, career advancement, and public recognition tend to fall to the researcher who investigates and confirms a new prediction than to the one who pursues a more familiar idea (see discussions by Armstrong, 1982; Fishman & Neigher, 1982; Greenwald, 1975). Editors may similarly be more ready to allocate journal space to results that confirm novel predictions than to studies that do the more mundane mopping up of the trail left by these novel developments. This pattern of incentives to the researcher would be justified if results that supported novel theoretical predictions were generally informative. However, if the novel result has been obtained within the context of a theory-confirming strategy, its conclusion has been effectively foreordained and is, therefore, not very informative.

Sad Fate of Suggested Remedies

Previous critics, in the process of making observations about researchers' strategies, have noted that it is difficult to take much published research at face value (e.g., Meehl, 1967). Further, they have suggested various remedies, which include (a) using a more stringent alpha criterion than the customary .05, to combat the effective "alpha inflation" that is incurred by repeated serial testing of the same prediction (Selvin & Stuart, 1966; Sterling, 1959); (b) standardly reporting measures of magnitude of effect, because null-hypothesis rejection can be obtained with trivially small effects when enough observations are made (Hays, 1981); (c) basing editorial decisions on reports of just methods (not results) in order to avoid bias against publishing null results (Walster & Cleary, 1970); and (d) developing statistical procedures for drawing conclusions in form other than rejection of null hypotheses—specifically, using interval estimation or Bayesian techniques (Bakan, 1966; Edwards, Lindman, & Savage, 1963; Grant, 1962; Greenwald, 1975). Although one can find occasional demonstration uses of these procedures, it is remarkable that the suggested remedial techniques are not much used (even by their advocates!).

The nonacceptance of remedial measures indicates that existing practices, faulty though they may be, have a powerful appeal. The present analysis identifies that appeal as their theory-confirming function. The remedies suggested by previous critics may remain unused precisely because they uniformly reduce prospects for obtaining results that confirm researchers' theoretical predictions. (Never mind that it is the least justifiable theory confirmations that the alternative procedures should selectively avoid!)

This is a harsh verdict. To state it bluntly: Researchers' dispositions to confirm hypotheses support their use of methods that are demonstrably prone to misinterpretation and, because of that, obstruct scientific progress.

What Is the Appeal of Confirmation-Biased Methods?

Confirmation bias delays finding support for beliefs other than those presently held. Why have confirmation-biased methods not been supplanted, long ago, by more objective methods? We suggest three answers. Although these answers are speculative, they have the virtue of suggesting ways to overcome the problems of confirmation bias.

First, *confirmation bias can be a useful heuristic.* That is, even though they lack objectivity, confirmation-biased methods may often work well (see Klayman & Ha, 1984). Confirmation bias

[5] The extent to which the present social system of research is self-correcting is further considered below under the heading "Is the Condition-Seeking Strategy Already in Use?"

[6] Theoretical opponents have well recognized the futility of attempting to base an empirical case against psychic phenomena on laboratory research. Instead, they have operated as detectives, attempting to expose error or fraud in purported demonstrations of psychic phenomena (e.g., Randi, 1982).

is, indeed, problematic only when there exists an alternative hypothesis that might be demonstrated to be superior to the researcher's hypothesis. When no superior alternative is available, confirmation bias helps to provide support for the best (or only) available hypothesis. (Against this, of course, it can be argued that—at least in principle—a better alternative always exists and that confirmation bias will delay its discovery.)

Second, *confirmation bias reinforces itself.* The experience of having one's predictions confirmed should increase or maintain a researcher's self-esteem and support the researcher's expectations of similar future successes. These components of an efficacious self-image are known to contribute to sustained, persevering activity (Bandura, 1977; Greenwald, 1980). Thus, researchers who most effectively confirm their predictions will be the ones who are most likely to sustain or to expand their research activity. If confirmation-biased researchers are thus more likely to persist in research activity than are unbiased ones, then it follows that active researchers are likely to be confirmation biased. In effect, confirmation bias is a trait that tends to be selected (in the evolutionary sense) by the disciplinary environment in which research occurs.

Third, *confirmation bias helps to strengthen the "establishment" that practices it.* By its effect in delaying support for novel hypotheses, confirmation bias is a conservative force that maintains a theoretical status quo. If veteran researchers pass on both specific theoretical beliefs and confirmation-biased strategies to their students, the theories of the senior practitioners will tend to be maintained by researchers of the next generation (cf. Tolman's remark: "Once set up, a [theoretical] system . . . serves as a sort of sacred grating behind which each novice is commanded to kneel in order that he may never see the real world, save through its interstices" [1932, p. 394]). To all those in the school of a theory, there are tangible rewards of status and resources for perpetuating the theory via collective, confirmation-biased methods.

These interpretations suggest that (confirmation-biased) theory-confirming research strategies have great survival value. In order to compete with confirmation-biased strategies, alternative strategies must have similar self-sustaining potential; they must permit researchers to succeed sufficiently to sustain their participation in research and in the training of future researchers.

Relation to the Views of Philosopher–Historians of Science

Present norms of scientific method have been strongly influenced by Popper's (1934/1959) contrast between dogmatic (confirmation seeking) and critical (disconfirmation seeking, or falsification) methods. In Popper's analysis, exclusive use of confirmation-seeking methods is not science, but "pseudoscience." The empirical knowledge content of a discipline can grow, according to Popper, only by use of critical, falsification-seeking methods.

Translating our argument into Popper's terms, we have observed that researchers who set out on the high (critical/falsification) road of testing theoretical hypotheses may nevertheless be effectively drawn onto the low (dogmatic) road of theory confirming. The difference between these two paths is subtle. A critical research investigation becomes a dogmatic one somewhere in the course of the researcher's perseverance in testing theoretical predictions that have produced disconfirming results.

In contrast to Popper, Kuhn (1970) considers that ordinary scientific activity thrives on confirmation seeking—solving puzzles within the existing paradigm. Although Kuhn notes the virtue of steady progress that comes from this "normal science," he also observes that such tradition-directed activity does not succeed indefinitely. Anomalous results accumulate until only a major theoretical reorganization (scientific revolution) can accommodate them.

Our description of researchers' theory-confirming activities fits better with Kuhn's conception of puzzle solving than with Popper's "sophisticated falsificationism" (as characterized by Lakatos, 1970). Although the normal science that Kuhn described does achieve steady gains, still (in our view) these gains may come at a needlessly slow pace, and valuable resources are consumed in the production of avoidably overgeneralized conclusions. By contrast, Popper's falsification-seeking methods can in principle overcome confirmation bias. However, the evidence of history is that the falsification method receives much lip service but little use. Many researchers who would describe their activity as falsification seeking are effectively engaged, instead, in the subtly different activity of theory confirming.

Result-Centered Research Strategies

Opposed although they are in other respects, both Popper and Kuhn maintain that theory testing is the central pole around which scientific activity revolves. In contrast, our analysis suggests that theory should not play so pivotal a role. The researcher who sets out to test a theory is likely to become ego-involved with a theoretical prediction, to select procedures that will lead eventually to prediction-confirming data, and thereby, to produce avoidably overgeneralized conclusions.

In this section we describe two research strategies that move theory away (but not far away) from the center of the research process. In evaluating these strategies, we shall attempt to justify the conclusion—which we see as a radical one—that testing theory should not routinely be the central goal of research.[7]

Method of Condition-Seeking

One method of avoiding needlessly overgeneralized findings is to set out deliberately to *reduce* the generalizability of an existing finding. That is, one can seek to discover which, of the many conditions that were confounded together in procedures that have obtained a finding, are indeed necessary or sufficient. This is the method of condition seeking.

Theory-testing researchers start with a theory, from which a predicted result (a relationship among focal variables), *R*, is generated. The aim of ensuing research is, in effect, to answer the question "Does *R* occur?" In contrast, the question that follows from the condition-seeking strategy is of the form "*Under what conditions* does *R* occur?" (For the sleeper-effect example, the

[7] It may be useful to remind the reader here that we do not advocate an atheoretical approach to research. Rather, we see theory as essential to research. The necessity of theory is discussed further at the end of this section and in the following, concluding section.

condition-seeking question is "Under what conditions does a communication accompanied by a discounting cue produce delayed persuasive impact?") The condition-seeking researcher deliberately seeks to produce a *qualified* conclusion (such as "A persuasive message accompanied by a discounting cue produces delayed impact *if* the cue is presented after an effective message, but *not if* it is presented before the message").

Increasing precision of conclusions. Sustained use of the condition-seeking method generates a progression of research questions that, if pursued empirically, yield increasingly precise conclusions. In abstract, the sequence is as follows:

Question 1: Under what conditions does *R* occur?
Answer 1: Under Condition C1 (and not in the absence of C1).
Question 2: Under what conditions does C1 produce *R*?
Answer 2: Under Condition C2 (and not in the absence of C2).
Question 3: Under what conditions does (C1 + C2) produce *R*? (etc.).

In summary, when a research question is asked in condition-seeking form, the answer must specify conditions under which a given result is not obtained, as well as specifying conditions under which it is obtained. The difference between unqualified and qualified conclusions corresponds to that between statistical main and interaction effects. The condition-seeking method can therefore be understood as an interaction-effect-seeking method.

Design Approach

In contrast with the method of condition-seeking, which discovers conditions on which an existing finding depends, the design approach aims to specify conditions that can produce a presently *un*obtainable result. The design approach is an attempt to engineer a desired result.

As with the method of condition seeking, research questions based on the design approach can be phrased to start "Under what conditions . . . ?" One use of the design approach is to discover conditions that can produce results that are of known practical value—for example, "Under what conditions do people who are inebriated decide that they should not drive?" or "Under what conditions do adolescents resist pressure from peers to smoke cigarettes?"

Another use of the design approach is to generate results that constitute significant reversals of familiar theory-based findings—for example, "Under what conditions does viewing televised violence *reduce* subsequent tendencies for the viewer to act aggressively?" or "Under what conditions does a child's liking for an activity *increase* directly with the magnitude of incentive to engage in it?" Because it is, in effect, well known how not to obtain the sought result, the outcome of an experiment using the design strategy need not be in the form of an interaction effect.

The condition-seeking and design strategies form a complementary pair. The method of condition seeking aims to discover conditions on which an already obtainable finding depends, whereas the design approach seeks conditions on which a presently unobtainable finding depends. The condition-seeking method reduces the range of generality of a finding, whereas the design approach increases (from nil) the range of generality of a finding.

Is the Condition-Seeking Strategy Already in Use?

Some readers of a preliminary version of this article suggested that the advocated condition-seeking strategy, even if not explicitly identified by this (or any other) label, is already a widely used strategy. This comment occurred in two forms: (a) that there is a long-standing tradition among researchers whereby predicted statistical interaction effects are considered to be a particularly valuable form of result and (b) that condition seeking occurs effectively as the collective product of theoretical competition in a research community, even if it is not the goal of individual researchers or teams. Although we agree with both of these observations, we cannot agree that contemporary research practice is therefore free of the ills of confirmation bias.

Interaction-effect-seeking in practice. There appears to be at least a limited informal tradition favoring interaction-effect research designs, as indicated by responses to a query sent to several innovators in social psychological research method.[8] At the same time, explicit discussion of the importance of interaction effects is almost nonexistent in psychological methods texts. The following statement by Mills (1969) was the most detailed that we could find in an examination of prominent psychological methodology texts of the past three decades (cf. Carlsmith, Ellsworth, & Aronson, 1976, p. 248; McGuire, 1969, p. 140).

> Perhaps the most valuable kinds of experimental studies in social psychology are ones that find a statistical interaction between two independent variables in their effect on a dependent variable. To show that a phenomenon varies as a function of certain conditions usually helps a great deal to narrow down the possible explanations for the phenomenon. (p. 437)

The lack of general use of condition-seeking methods is indicated both by the discussion of the problem of avoidable overgeneralizations, earlier in this article, and by McGuire's (1973, 1983) criticisms of researchers' inattention to conditions on which their findings depend. (McGuire's views are considered in more detail below.) In sum, there appears to have been an informal discovery (perhaps independently among several researchers) of the value of a condition-seeking strategy. However, this informal discovery has not been accompanied by any strong or widespread advocacy of the method.

Examples of collective condition seeking. It is possible to find many examples of condition seeking emerging as the informal, collective product of competitive theory-centered individual research efforts. As previously noted, controversy over the insufficient justification effect was resolved by the progressive, collective discovery of qualifying conditions (see further discussion, below). Another problem that occupied several theoretically competing groups of social psychologists was the risky shift effect, the finding that group decisions were typically more risky than those made by individuals (Stoner, 1961; see the discussion of scientific progress in the context of this problem by Cartwright, 1973). The risky shift controversy was eventually resolved by the discovery of conditions that produced the opposite effect, a conservative shift (Brown, 1965; Myers & Lamm, 1975). A problem

[8] Personal communications from Elliot Aronson (May 17, 1984), Jack W. Brehm (January 15, 1985), Edward E. Jones (January 28, 1985), and Judson Mills (January 14, 1985).

that occupied many cognitive psychologists for about a decade was the set-size effect, the finding of an increase in latency of judging that an item is a member of a defined target set, as a monotonic function of the number of items in the target set (Sternberg, 1967). Although interpretation of the set-size effect is still at issue, research activity on the problem subsided greatly upon establishment of conditions under which the effect did not occur (Shiffrin & Schneider, 1977).

Comparative efficiency of between-groups versus within-group condition seeking. The insufficient justification, risky shift, and set-size effects are just a few of the many cases in which active empirical disputes achieved resolution upon discovery of limiting conditions for previously overgeneralized conclusions. These examples make clear that, indeed, condition seeking does often occur as the collective product of a community of researchers who are competitively pursuing theory-centered strategies. But do these examples justify the existing practice of researchers pursuing theory-centered (and often confirmation-biased) strategies? To the contrary, the examples given might better be taken as indications that progress cannot be postponed indefinitely! There is typically a great deal of wasted effort in the competitive resolution of an empirical controversy. This inefficiency is largely due to researchers' suppressed attention to results that did not come out as hoped. The consequence is that valuable information—about conditions under which a predicted effect fails to occur—is not generally available to the community of interested researchers. Careful analysis might reveal that ultimate resolutions of empirical controversies are more likely to be produced by theoretically impartial bystanders who were warmed to action by the flames of controversy than by theoretically committed antagonists.

Existing Uses of the Design Strategy

Even though the design strategy has not previously been identified or advocated as a distinct method, occasional examples of its use provide compelling evidence for its value. Consider, for example, the effort by Aronson, Stephan, Sikes, Blaney, and Snapp (1978) to enable desegregated schools to undo the educational disadvantage of minorities. When nationwide school desegregation was originally implemented in the United States, it was not anticipated that majority–minority contact might adversely affect the self-esteem and school performance of minorities. This oversight was, in effect, the consequence of overgeneralization of conclusions from prior research—a point made in reviews by Stephan (1978) and Gerard (1983; see also S. W. Cook, 1984). Aronson and his colleagues identified this adverse effect and designed a classroom strategy that was expected, on theoretical grounds, to reverse it. Their successful plan, the "jigsaw classroom" (based partly on the famous Robbers Cave experiment by Sherif, Harvey, White, Hood, & Sherif, 1961), made use of group tasks that required cooperative, equal-status interaction in heterogeneous (majority plus minority) groups.

Some other examples of the design strategy are J. R. Nuttin's (1953) creation of conditions under which punished responses were remembered better than rewarded ones; Moscovici's (1976) demonstration of situations in which a minority influences a majority, reversing the usual conformity effect; and Orne's construction of situations in which normal subjects exhibited phenomena that were previously believed to be exclusively associated with the special states of sensory deprivation (Orne & Scheibe, 1964) and hypnosis (Orne & Evans, 1965). Such findings not only have considerable surprise value, but more importantly, they dramatically call attention to the need for improved theoretical understanding.

Scientific/Theoretical Progress via the Condition-Seeking Method

The condition-seeking method produces a progression of qualifying conditions on existing findings. This may suggest a dreary view of progress: an accumulation of increasingly precise results that are limited to ever-shrinking domains, as interactions with previously unexamined variables are discovered. (Consider Cronbach's comment: "Once we attend to interactions, we enter a hall of mirrors that extends to infinity. However far we carry our analysis—to third order or fifth order or any other—untested interactions of a still higher order can be envisioned" [1975, p. 119].)

However, the suggestion that progress via condition seeking occurs in the form of increasingly trivial results is based on a too-narrow view, one that ignores the vital role of theory in scientific progress. The need for theory to guide empirical progress, along with the stimulus to theory that is provided by new findings, rescues the condition-seeking method from being just a means of cultivating empirical trivialities. In order to accommodate the increasing precision of findings obtained by condition seeking, existing theories are stressed; theories must grow in their power in order to permit economical descriptions of the new findings.

An example of the increasing economy of statement that can accompany progressive discovery of limiting conditions comes from the history of research on the insufficient justification problem. Statement 1, below, is worded in terms of the theoretical analysis originally provided by cognitive dissonance theory; it bulges with appended limiting conditions that are needed to accommodate the findings of 13 years of research. Statement 2 gives a more economical summary in terms of an evolved theoretical account.

> 1. *Multiply qualified original result.* Dissonance between prior belief and a belief-discrepant statement is reduced, by change of belief toward agreement with the discrepant statement, more when a small (than a large) incentive has been offered for making the statement (Festinger & Carlsmith, 1959) if (a) the person freely chooses to make the statement (Brehm & Cohen, 1962) and (b) is publicly committed to the statement (Brehm & Cohen, 1962), when (c) the effect of the statement is to produce aversive consequences (Calder, Ross, & Insko, 1973) that (d) were foreseeable when the person agreed to make the statement (Wicklund & Brehm, 1976).
>
> 2. *Theoretically evolved statement.* Belief change toward agreement with a belief-discrepant statement occurs to the extent that making a disbelieved statement implies a trait (such as dishonesty or stupidity) that is strongly inconsistent with the subject's self-image (Aronson, 1968; Deutsch, Krauss, & Rosenau, 1962; Greenwald & Ronis, 1978).

Statement 1 gives the undesirable appearance of progress in the form of a theory with ever-shrinking domain. By contrast, Statement 2 is powerfully more general, subsuming the (qualified) insufficient justification effect of Statement 1 under an analysis that integrates dissonance reduction with a principle of self-es-

teem-maintaining cognitive change (cf. Greenwald, 1980; Schlenker, 1982; Tesser & Campbell, 1983). Statement 2 is, for example, powerful enough to suggest a fifth qualification of the insufficient justification effect: It should not occur among persons whose self-concept includes traits that are consistent with making disbelieved statements. (Nevertheless, further progress at condition seeking should lead toward an interpretation even more powerful than Statement 2.)

McGuire's Analysis of the Relation of Theory to Research

Our conception of the relationship of research to theory has been preceded and heavily influenced by Willim J. McGuire's (1973, 1983) writings. This influence can be seen most readily through some quotations from McGuire's (1983) description of the methodological approach that he calls *contextualism.* (These quotations are from McGuire's section entitled "Corrupting Effects of the 'Hypothesis-Testing' Method.")

> It can be taken for granted that some set of circumstances can be found to confirm any expressible relationship, provided that the researcher has sufficient stubbornness, stage management skills, resources, and stamina sooner or later to find or construct a situational context in which the predicted relationship reliably emerges. (pp. 15–16)
>
> A researcher ingenious enough and persistent enough will at last design an experiment that comes out "right" and it is only this sanitized experiment . . . that becomes the sole portion of the research program revealed in the manuscript submitted to the dissertation committee or the journal. (p. 15)
>
> The more valuable information obtainable through empirical confrontation emerges from the pattern of contexts in which the predicted relationship obtains as contrasted with those in which the contrary relationship or none at all obtains. (p. 16)
>
> Contextualism maintains that a theory (like knowledge on any other level) is an oversimplified and distorted representation of any situation. . . . A hypothesis or its contradictory is each adequately true in a few appropriate contexts and each is dangerously false in many others. . . . [Contextualism regards] empirical confrontation, not as a test to determine whether the hypothesis is true or not, but rather as a continuing discovery process to disclose the hypothesis's full meaning by revealing its hidden assumptions and so specifying in which contexts its misrepresentations are tolerable and in which seriously misleading. (pp. 7–8)

We agree very fundamentally with the premises of McGuire's analysis. Further, our suggested condition-seeking and design strategies overlap with some of the research-strategy suggestions McGuire makes (1983, pp. 26–27). We exceed McGuire's recommendations primarily in concluding that theory testing should often be displaced from its status as a central goal of research.

Conclusion: The Role of Theory in Research

Science in the Middle Ground Between Truth and Falsity

Contemporary philosophical analyses of scientific method demand a skeptical regard toward theory: No theory can be proven true by empirical data. And, just as it is impossible to prove a theory true, so also is it impossible to prove one false. Any theory can be patched, by ad hoc addition of assumptions, to fit with existing data. The joint impossibility of either proving or disproving theories might appear to justify dispensing with theories entirely. However, such a skepticism is difficult to maintain. The ability to benefit from experience is greatly facilitated by converting experience into conceptual knowledge (theory) that can serve as a basis for action, even if that knowledge rests on an uncertain epistemological foundation.

Fortunately, there is an immense middle ground between the philosophically unjustifiable positions of regarding theories as empirically provable or disprovable. And it is the work of science to evaluate theories within this middle ground. Fragile and misdirected although they are, theories are the essential containers of scientific knowledge and the necessary vehicles of scientific progress. As suggested by the metaphors of containers and vehicles, criteria such as storage capacity and speed of progress—criteria that are appraisable without having to speak of proof and disproof, or of truth and falsity—are most appropriate for evaluating theories. The work of science may best be regarded as *ap*proving and dis*ap*proving of theories, rather than as proving and disproving them.

Evolution of Theory in Response to Empirical Goals

Perhaps the greatest virtue of researchers is their perseverance, their willingness to continue data collection until a sought pattern materializes. When this perseverance is applied in the context of a theory-centered method, the ill-defined (because logically impossible) activity of theory testing is often converted into the better defined activity of producing the pattern of results predicted by the theory. In order to produce a theory-predicted pattern, stress is put on procedure. The consequence is that the persevering researcher finds a procedure that will produce the desired relationship. But when, as is often the case, the finished research product provides no contrast between the theory-predicted pattern and its absence, there may be no solid gain in understanding of the conditions on which the empirical pattern depends. More specifically, the researcher may not be able to describe, in a published report, the procedures that were critical to obtaining the reported prediction-confirming finding. Such a finding is difficult to replicate, and the theory that predicted it is difficult to apply.

By contrast, the condition-seeking and design strategies do not allow the often-commendable trait of perseverance to corrupt (as McGuire puts it) the research process. By explicitly seeking empirical interaction-effect patterns, these strategies put stress on theory, along with procedure. This stress obliges theory to increase its precision and thereby to become increasingly useful.

The observations just made can be recast in terms of the functioning of scientific research within an evolving system of knowledge. Alternative research strategies can be understood in terms of the manner in which they allow theories to act as evolutionary variations and the pressures that they apply to select among those variations. It would certainly be efficient for such selection pressures to operate directly on theories. However, it is an article of faith among scientists (and perhaps a chief defining characteristic of scientific activity) that theory is to be evaluated primarily by the indirect means of empirical test. In an empirical test, the conceptual elements of a theory are given operational form, and

this procedural proxy faces the survival test of confirmation versus disconfirmation.

In principle, theory-centered methods operate by subjecting the empirical incarnations of theories to crucial experimental tests (cf. Chamberlin, 1890/1965; Platt, 1964). However, when the researcher becomes an ego-involved theory advocate, this falsification-seeking strategy is converted to a theory-confirming strategy (see Figure 3), at which point the philosophical impossibility of disproving a theory provides effective license to preserve the theory come what may. The theory-confirming researcher allows the rule-of-correspondence link between theory and operations to be loose and variable. It is then this link between theory and procedure that becomes exposed to the confirmation–disconfirmation test, rather than the theory itself. In this theory-confirming context, survival of theories is governed more by political selection criteria than by empirical ones. It is not surprising, then, that a process deserving of Kuhn's (1970) label, *revolution,* may be needed to replace a well-established theory.

For result-centered methods (such as condition seeking and design), again research procedures provide the variations on which selection operates directly. The selection criterion is agreement versus disagreement of an obtained data pattern with one specified in advance. If a procedural variant that has been suggested by a theory does not produce the desired pattern, it is replaced, and importantly, the theory that suggested it is at least mildly discredited. The survival of a theory is a function of its ability to generate effective procedures. Importantly, the result-centered researcher works with designs that can pinpoint critical procedural variations (e.g., by producing interaction-effect findings).

The supporting role that theory plays in result-centered methods is by no means a minor one. Result-centered research goals such as producing an explosion based on nuclear fission, destroying malignant (but not nonmalignant cells), or creating computerized simulations of intelligent performances place great demands on theory to suggest possibly effective procedures. The present argument is that the selection pressures of result-centered research methods achieve more rapid evolution of effective theory than do those of theory-centered methods.

Under What Conditions Does Theory Obstruct Research Progress?

The identifying mark of our recommended condition-seeking and design strategies is a research question that begins with the words "Under what conditions . . . ." We conclude by summarizing our answer to the question posed in this article's title: Under what conditions does theory obstruct research progress? Theory obstructs research progress when testing theory is taken as the central goal of research, if (as often happens) the researcher has more faith in the correctness of the theory than in the suitability of the procedures that were used to test it. In other words, theory obstructs research progress when the researcher is an ego-involved advocate of the theory and may be willing to persevere indefinitely in the face of prediction-disconfirming results. Regrettably, because the social systems of many scientific disciplines encourage researchers to be ego-involved theory advocates, the conditions under which theory is likely to obstruct scientific progress are ones that occur commonly.

Philosophers of science since Francis Bacon, along with methodologists of more recent eras, have been aware of the potential for research activity to be misdirected. Accordingly, scientists have repeatedly been urged toward greater objectivity of method. Perhaps research will eventually become the objective, impartial, non-ego-involved enterprise that, over the years, methodologists have advocated. That day, however, is not yet at hand. In the meantime, recommendations for research strategy should take into account the reality that researchers work most perseveringly when ego-involved—in other words, that ego-involvement is an effective, albeit fallible, motivator of research activity. The result-centered methods that we recommend permit the researcher to be ego-involved, but with the goal of producing a desired pattern of results rather than that of confirming a particular theory. It remains to be seen whether this alternative heuristic—which we argue to be a more efficient one—can be as psychologically compelling as the confirmation-biased heuristic that it would replace.

References

Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology, 19,* 93–108.

Armstrong, J. S. (1982). Research on scientific journals: Implications for editors and authors. *Journal of Forecasting, 1,* 83–104.

Aronson, E. (1968). Dissonance theory: Progress and problems. In R. P. Abelson et al. (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally.

Aronson, E., Stephan, C., Sikes, J., Blaney, N., & Snapp, M. (1978). *The jigsaw classroom.* Beverly Hills, CA: Sage.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258–290.

Bakan, D. (1966). The test of significance in psychology research. *Psychological Bulletin, 66,* 432–437.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191–215.

Bandura, A., & Walters, R. H. (1963). *Social learning and personality development.* New York: Holt, Rinehart & Winston.

Baumgardner, M. H., Leippe, M. R., Ronis, D. L., & Greenwald, A. G. (1983). In search of reliable persuasion effects: II. Associative interference and persistence of persuasion in a message-dense environment. *Journal of Personality and Social Psychology, 45,* 524–537.

Bolles, R. C. (1975). Learning, motivation, and cognition. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 1, pp. 249–280). Hillsdale, NJ: Erlbaum.

Brehm, J. W., & Cohen, A. R. (1962). *Explorations in cognitive dissonance.* New York: Wiley.

Brown, R. W. (1965). *Social psychology.* New York: Free Press.

Bruner, J. S., & Potter, M. C. (1964). Interference in visual recognition. *Science, 144,* 424–425.

Brush, S. G. (1974). Should the history of science be rated X? *Science, 183,* 1164–1172.

Buchwald, A. M. (1969). Effects of "right" and "wrong" on subsequent behavior: A new interpretation. *Psychological Review, 76,* 132–143.

Calder, B. J., Ross, M., & Insko, C. A. (1973). Attitude change and attitude attribution: Effects of incentive, choice, and consequences. *Journal of Personality and Social Psychology, 25,* 84–99.

Campbell, D. T. (1986). Science's social system of validity-enhancing collective belief change and the problems of the social sciences. In D. W. Fiske & R. A. Shweder (Eds.), *Pluralisms and subjectivities in the social sciences.* Book in preparation.

Capon, N., & Hulbert, J. (1973). The sleeper effect—an awakening. *Public Opinion Quarterly, 37,* 333–358.

Carlsmith, J. M., Ellsworth, P. C., & Aronson, E. (1976). *Methods of research in social psychology.* Reading, MA: Addison-Wesley.

Cartwright, D. (1973). Determinants of scientific progress: The case of research on the risky shift. *American Psychologist, 28,* 222–231.

Chamberlin, T. C. (1965). The method of multiple working hypotheses. *Science, 148,* 754–759. (Original work published 1890)

Collins, B. E., & Hoyt, M. F. (1972). Personal responsibility-for-consequences: An integration and extension of the "forced compliance" literature. *Journal of Experimental Social Psychology, 8,* 558–593.

Cook, T. D. (1971). *The discounting cue hypothesis and the sleeper effect.* Unpublished manuscript, Northwestern University.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation.* Chicago: Rand McNally.

Cook, T. D., Gruder, C. L., Hennigan, K. M., & Flay, B. R. (1979). History of the sleeper effect: Some logical pitfalls in accepting the null hypothesis. *Psychological Bulletin, 86,* 662–679.

Cook, S. W. (1984). The 1954 Social Science Statement and school desegregation: A reply to Gerard. *American Psychologist, 39,* 819–832.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116–127.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Darley, J. M., & Fazio, R. H. (1980). Expectancy confirmation processes arising in the social interaction sequence. *American Psychologist, 35,* 867–881.

Deutsch, M., Krauss, R. M., & Rosenau, N. (1962). Dissonance or defensiveness? *Journal of Personality, 30,* 16–28.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242.

Feshbach, S., & Singer, R. D. (1971). *Television and aggression.* San Francisco: Jossey-Bass.

Festinger, L. (1955). Social psychology and group processes. In C. P. Stone & Q. McNemar (Eds.), *Annual review of psychology* (Vol. 6, pp. 187–216). Stanford, CA: Annual Reviews.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58,* 203–210.

Festinger, L., Riecken, H., & Schacter, S. (1956). *When prophecy fails.* Minneapolis: University of Minnesota Press.

Feynman, R. P. (1985). *Surely you're joking, Mr. Feynman.* New York: Norton.

Fishman, D. B., & Neigher, W. D. (1982). American psychology in the eighties: Who will buy? *American Psychologist, 37,* 533–546.

Freedman, J. L. (1984). Effect of television violence on aggressiveness. *Psychological Bulletin, 96,* 227–246.

Gerard, H. B. (1983). School desegregation: The social science role. *American Psychologist, 38,* 869–877.

Gillig, P. M., & Greenwald, A. G. (1974). Is it time to lay the sleeper effect to rest? *Journal of Personality and Social Psychology, 29,* 132–139.

Gould, S. J. (1981). *The mismeasure of man.* New York: Norton.

Grant, D. A. (1962). Testing the null hypothesis and the strategy and tactics of investigating theoretical models. *Psychological Review, 69,* 54–61.

Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin, 82,* 1–20.

Greenwald, A. G. (1976). Within-subjects designs: To use or not to use? *Psychological Bulletin, 83,* 314–320.

Greenwald, A. G. (1980). The totalitarian ego: The fabrication and revision of personal history. *American Psychologist, 35,* 603–618.

Greenwald, A. G., & Ronis, D. L. (1978). Twenty years of cognitive dissonance: Case study of the evolution of a theory. *Psychological Review, 85,* 53–57.

Greenwald, A. G., & Ronis, D. L. (1981). On the conceptual disconfirmation of theories. *Personality and Social Psychology Bulletin, 7,* 131–137.

Gruder, C. L., Cook, T. D., Hennigan, K. M., Flay, B. R., Alessis, C., & Halamaj, J. (1978). Empirical tests of the absolute sleeper effect predicted from the discounting cue hypothesis. *Journal of Personality and Social Psychology, 36,* 1061–1074.

Hays, W. L. (1981). *Statistics for psychologists* (3rd ed.). New York: Holt, Rinehart & Winston.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change.* New Haven, CT: Yale University Press.

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. (1949). *Experiments on mass communications.* Princeton, NJ: Princeton University Press.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15,* 635–650.

Janis, I. L., & Gilmore, J. B. (1965). The influence of incentive conditions on the success of role playing in modifying attitudes. *Journal of Personality and Social Psychology, 1,* 17–27.

Kelman, H. C., & Hovland, C. I. (1953). "Reinstatement" of the communicator in delayed measurement of opinion change. *Journal of Abnormal and Social Psychology, 48,* 327–335.

Klayman, J., & Ha, Y. (1984). *Confirmation, disconfirmation, and information in hypothesis-testing* (Working Paper No. 119). Chicago: University of Chicago, Center for Decision Research, Graduate School of Business.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd. ed.). Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). London: Cambridge University Press.

Lakatos, I. (1976). *Proofs and refutations.* Cambridge, England: Cambridge University Press.

Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs, 54,* 1–95.

Luchins, A. (1957). Primacy-recency in impression formation. In C. I. Hovland et al. (Eds.), *The order of presentation in persuasion* (pp. 33–61). New Haven, CT: Yale University.

Lund, F. (1925). The psychology of belief. IV. The law of primacy in persuasion. *Journal of Abnormal and Social Psychology, 20,* 183–191.

Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research, 1,* 161–175.

McGuire, W. J. (1969). The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 3, pp. 136–314). Reading, MA: Addison-Wesley.

McGuire, W. J. (1973). The yin and yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology, 26,* 446–456.

McGuire, W. J. (1983). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 16, pp. 1–47). New York: Academic Press.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 233–346). New York: Random House.

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34,* 103–115.

Merton, R. K. (1948). The self-fulfilling prophecy. *Antioch Review, 8,* 193–210.

Milgram, S., & Shotland, L. (1973). *Television and social behavior: Field experiments.* New York: Academic Press.

Miller, N., & Campbell, D. T. (1959). Recency and primacy in persuasion

as a function of timing of speeches and measurements. *Journal of Abnormal and Social Psychology, 59,* 1–9.

Mills, J. (1969). *Experimental social psychology.* London: Macmillan.

Mischel, W., Ebbesen, E. B., & Zeiss, A. M. (1976). Determinants of selective memory about the self. *Journal of Consulting and Clinical Psychology, 44,* 92–103.

Mitroff, I. I. (1974). Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. *American Sociological Review, 39,* 579–595.

Moscovici, S. (1976). *Social influence and social change.* New York: Academic Press.

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental Psychology, 29,* 85–95.

Myers, D. G., & Lamm, H. (1975). The polarizing effect of group discussion. *American Scientist, 63,* 297–303.

National Institute of Mental Health. (1982). *Television and behavior: Ten years of scientific progress and implications for the eighties.* Washington, DC: U.S. Government Printing Office.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings.* Englewood Cliffs, NJ: Prentice-Hall.

Nuttin, J. M. (1975). *The illusion of attitude change: Toward a response contagion theory of persuasion.* London–New York–Leuven: Academic Press–Leuven University Press.

Nuttin, J. R. (1953). *Tâche, réussite, et échec.* Louvain, Belgium: Publications Universitaires.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. L. Rosnow (Eds.) *Artifact in behavioral research,* (pp. 143–179). New York: Academic Press.

Orne, M. T., & Evans, F. J. (1965). Social control in the psychological experiment: Antisocial behavior and hypnosis. *Journal of Personality and Social Psychology, 1,* 189–200.

Orne, M. T., & Scheibe, K. E. (1964). The contribution of nondeprivation factors in the production of sensory deprivation effects: The psychology of the "panic button." *Journal of Abnormal and Social Psychology, 68,* 3–12.

Platt, J. R. (1964). Strong inference. *Science, 146,* 347–353.

Popper, K. R. (1959). *The logic of scientific discovery.* New York: Basic Books. (Original work published 1934)

Postman, L. (1963). Rewards and punishments in human learning. In L. Postman (Ed.), *Psychology in the making* (pp. 331–401). New York: Knopf.

Pratkanis, A. R. (1981). *A hunt for the sleeper effect in persuasion.* Unpublished master's thesis, Ohio State University.

Pratkanis, A. R. (1984). *Attitudes and memory: The heuristic and schematic functions of attitudes.* Unpublished doctoral dissertation, Ohio State University.

Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1985). *In search of reliable persuasion effects: III. The sleeper effect is dead. Long live the sleeper effect.* Manuscript submitted for publication.

Randi, J. (1982). *Flim-flam.* Buffalo, NY: Prometheus.

Ronis, D. L. (1980). Repetition and agreement with opposing arguments: A delayed action effect. *Journal of Experimental Social Psychology, 16,* 376–387.

Ronis, D. L., Baumgardner, M. H., Leippe, M. R., Cacioppo, J. T., & Greenwald, A. G. (1977). In search of reliable persuasion effects: I. A computer-controlled procedure for studying persuasion. *Journal of Personality and Social Psychology, 35,* 548–569.

Rosenberg, M. J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology, 1,* 28–42.

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

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

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880–892.

Ross, M., McFarland, C., & Fletcher, G. J. O. (1981). The effect of attitude on the recall of personal history. *Journal of Personality and Social Psychology, 40,* 627–634.

Schlenker, B. R. (1982). Translating actions into attitudes: An identity-analytic approach to the explanation of social conduct. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 15, pp. 193–247). New York: Academic Press.

Selvin, H. C., & Stuart, A. (1966). Data-dredging procedures in survey analysis. *American Statistician, 20,* 20–23.

Sherif, M., Harvey, O. J., White, B. J., Hood, W., & Sherif, C. (1961). *Intergroup conflict and cooperation: The Robbers Cave experiment.* Norman: University of Oklahoma.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36,* 1202–1212.

Snyder, M., & Uranowitz, S. W. (1978). Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology, 36,* 941–950.

Stephan, W. (1978). School desegregation: An evaluation of predictions made in *Brown v. Board of Education. Psychological Bulletin, 85,* 217–238.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association, 54,* 30–34.

Sternberg, S. (1967). Two operations in character recognition: Some evidence from reaction time measurements. *Perception and Psychophysics, 2,* 45–53.

Stoner, J. A. F. (1961). *A comparison of individual and group decisions involving risk.* Unpublished master's thesis, Massachusetts Institute of Technology.

Swann, W. B., & Read, S. J. (1981). Self-verification processes: How we sustain our self-conceptions. *Journal of Experimental Social Psychology, 17,* 351–372.

Tesser, A. H., & Campbell, J. (1983). Self-definition and self-evaluation maintenance. In J. Suls & A. G. Greenwald (Eds.), *Psychological perspectives on the self* (Vol. 2, pp. 1–31). Hillsdale, NJ: Erlbaum.

Thorndike, E. L. (1933). A theory of the action of the after-effects of a connection upon it. *Psychological Review, 40,* 434–439.

Tolman, E. C. (1932). *Purposive behavior in animals and men.* New York: Century.

U.S. Public Health Service. (1972). *Television and growing up: The impact of televised violence.* Washington, DC: U.S. Government Printing Office.

Walster, G. W., & Cleary, T. A. (1970). A proposal for a new editorial policy in the social sciences. *American Statistician, 24,* 16–19.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 11,* 92–107.

Weiss, W. (1953). A "sleeper" effect in opinion change. *Journal of Abnormal and Social Psychology, 48,* 173–180.

Westfall, R. S. (1973). Newton and the fudge factor. *Science, 179,* 751–758.

Wicklund, R. A., & Brehm, J. W. (1976). *Perspectives on cognitive dissonance.* Hillsdale, NJ: Erlbaum.

Wyatt, D. F., & Campbell, D. T. (1951). On the liability of stereotype or hypothesis. *Journal of Abnormal and Social Psychology, 46,* 496–500.

Received April 8, 1985
Revision received September 20, 1985 ■