



The Sleeper Effect _An Awakening

Noel Capon; James Hulbert

The Public Opinion Quarterly, Vol. 37, No. 3 (Autumn, 1973), 333-358.

Stable URL:

http://links.jstor.org/sici?sici=0033-362X%28197323%2937%3A3%3C333%3ATSE_A%3E2.0.CO%3B2-D

The Public Opinion Quarterly is currently published by Oxford University Press.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/oup.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

THE SLEEPER EFFECT —AN AWAKENING*

BY NOEL CAPON AND JAMES HULBERT

An examination of research studies that assume the existence of the sleeper effect concept has revealed surprising results: this effect may be observed only under certain restrictive design conditions—with subsets of the population divided on the basis of personality characteristics.

Noel Capon, who was a doctoral student in marketing at the Graduate School of Business, Columbia University, has recently joined the faculty at the Graduate School of Management, University of California, Los Angeles. James Hulbert is Assistant Professor at Columbia University's Graduate School of Business.

THE SLEEPER EFFECT was first named over twenty years ago¹ and has been the subject of numerous subsequent investigations.² Unfortunately, however, close examination of these studies indicates a lack of consensus on the definition of the sleeper effect. Thus, although its existence has been claimed by many researchers, this effect has never been rigorously defined. This article develops a definition that is both practical and consistent with the spirit of most work in the field. Then, on the basis of this common definition, those studies often quoted as supportive of the sleeper effect are reviewed and evaluated. Although investigation of the sleeper hypothesis was not a major focus in a number of these studies, it is only from that perspective that they will be reviewed in this article; other noteworthy findings may well be omitted. The assembled evidence does not support the existence of a sleeper effect as a generalized phenomenon, but some data do suggest the possibility of a weak effect that operates selectively.

DEFINITION

Hovland, Lumsdaine, and Sheffield, originators of the sleeper

* The support of the Faculty Research Fund of the Graduate School of Business, Columbia University, is gratefully acknowledged.

¹ Carl I. Hovland, Arthur A. Lumsdaine, and Fred D. Sheffield, *Experiments in Mass Communication*, New York, Wiley, 1949, pp. 182-200.

² See, for example, Carl I. Hovland and Walter Weiss, "The Influence of Source Credibility on Communication Effectiveness," *Public Opinion Quarterly*, Vol. 15, 1951, pp. 635-650; and Walter Weiss, "A Sleeper Effect in Opinion Change," *Journal of Abnormal and Social Psychology*, Vol. 48, 1953, pp. 173-180.

effect concept, viewed it as a change in opinion over time in the direction advocated by a communication, based on measurements taken shortly after and several weeks after the communication. They observed that “. . . in some cases the effect of time may be to enhance the initial effects of the film. Thus, some of the effects of the film may be ‘sleepers’. . . .”³

Hovland and Weiss’s conception was similar, but ill-defined. It included the subtle but important distinction that a pre-measure served as the datum for measurement of opinion change. In their view, “. . . the ‘sleeper effect’ occurs among the group which initially disagrees with an unreliable source (but subsequently comes to agree with it). . . .”⁴

Kelman and Hovland also agreed that “. . . an individual may at first reject the communicator’s point of view, but after a period of time ‘come around’ to the communicator’s position.”⁵ However, their concept did not relate to a pre-measure but only to post-measures. Greenwald and Gillig employed a similar definition.⁶

Whittaker and Meade did employ a pre-measure, stating that “. . . the sleeper effect occurs when opinion change in the direction of the untrustworthy communicator’s position is larger after a lapse of time than immediately after the communication.”⁷ Stotland, Katz, and Patchen⁸ and Stotland and Patchen⁹ used similar definitions.

For all the studies cited above, a necessary condition for the sleeper effect was that an individual or group of individuals exhibit an opinion change over time in the direction advocated by the communication. However, some used a pre-measure and some an early post-measure as the basis for calculation of opinion change.

³ Hovland, Lumsdaine, and Sheffield, *op. cit.*, p. 213.

⁴ Hovland and Weiss, *op. cit.*, p. 647.

⁵ Herbert C. Kelman and Carl I. Hovland, “Reinstatement of the Communicator in Delayed Measurement of Opinion Change,” *Journal of Abnormal and Social Psychology*, Vol. 48, 1953, p. 327.

⁶ Anthony G. Greenwald and Paulette M. Gillig, “A Cognitive Response Analysis of the ‘Sleeper Effect,’” presented at *American Psychological Association (Div 8) Symposium: “Resistance to Persuasive Communication: Counterarguing Processes,”* Washington, D.C., September 4, 1971, (mimeo).

⁷ James O. Whittaker and Robert D. Meade, “Retention of Opinion Change as a Function of Differential Source Credibility,” *International Journal of Psychology*, Vol. 3, 1968, p. 108.

⁸ Ezra Stotland, Daniel Katz, and Martin Patchen, “The Reduction of Prejudice Through the Arousal of Self-Insight,” *Journal of Personality*, Vol. 27, 1959, pp. 507-531.

⁹ Ezra Stotland and Martin Patchen, “Identification and Changes in Prejudice and in Authoritarianism,” *Journal of Abnormal and Social Psychology*, Vol. 62, 1961, pp. 265-274.

The definition employed by Weiss was quite different. His study required the presence of two groups of communicatees: discounted and non-discounted groups. He considered a sleeper effect present ". . . if following exposure to a communication the loss of effectiveness with the passage of time is less for a group influenced by a discounting factor than for one not receiving this additional treatment."¹⁰ Not only were two groups required but, further, opinion changes in the advocated direction were no longer necessary.

Schulman and Worrall used a communication paradigm employing high and low credibility sources. A sleeper effect was deemed present if ". . . (there was) significantly greater immediate opinion change toward the high credibility source than toward the low credibility source . . . ; at three to four weeks after exposure there was no significant difference . . . ; (and) opinion change in the high credibility source condition showed a drop from immediately after exposure to the delayed measure, while there was a non-significant increase for the low credibility source."¹¹ Despite the fact that this is a recent study, there is inherent confusion in the rambling definition.

The major definitional distinctions, therefore, revolve around (1) whether an individual or group can manifest the effect or whether the effect requires the comparison of two individuals or groups; (2) whether changes in opinion in the advocated direction are necessary; and (3) whether or not the changes are based on opinions held prior to the communication.

Clearly, in order to test for the existence of a sleeper effect, there must be agreement on just what that effect is, or would be. This poses the awkward problem of selecting a single definition with an awareness that subsequent interpretation of the evidence is dependent upon it. The definition that has been selected is, we believe, congruent in spirit with seven of the nine definitions discussed and with most interpretation in the literature.

A sleeper effect may be observed if an individual's agreement with a persuasive communication is greater a long time after exposure to it than immediately thereafter. Further, the final measure of agreement, when compared to the pre-measure, must show a shift in the advocated direction.

This definition implies that a single individual can manifest the effect; that a positive opinion change is required; and the meas-

¹⁰ Weiss, *op. cit.*, p. 173.

¹¹ Gary I. Schulman and Chryssoula Worrall, "Salience Patterns, Source Credibility, and the Sleeper Effect," *Public Opinion Quarterly*, Vol. 34, 1970, p. 371.

ures must be taken at three points in time: before, immediately after, and a long time after the communication.

Now, let O_{t_0} , O_{t_1} , and O_{t_2} be the communicatee's opinion at t_0 , t_1 , and t_2 ; where t_0 and t_1 are immediately before and after the communication and t_2 is later in time. A sleeper effect is recognized if:

$$O_{t_2} > O_{t_1} \quad \text{and} \quad O_{t_2} > O_{t_0}$$

This definition is fundamentally different from Weiss', which posited that a comparison of two individuals or groups, each receiving different treatments, was necessary. The individual—as a target for persuasive communication—could not, therefore, manifest the effect. Such a comparative effect is of interest, but should not be confused with the individual or group effect, which is the basis for most understanding of the sleeper effect. Indeed, the studies of Hovland and Weiss; Kelman and Hovland; Whittaker and Meade; Stotland, Katz, and Patchen; and Stotland and Patchen all employed at least two treatments. Yet these authors based their analysis of the sleeper effect on data from a single treatment. The difference in the definition is highlighted by the hypothetical examples in Figure 1, where I and II imply different treatment groups.

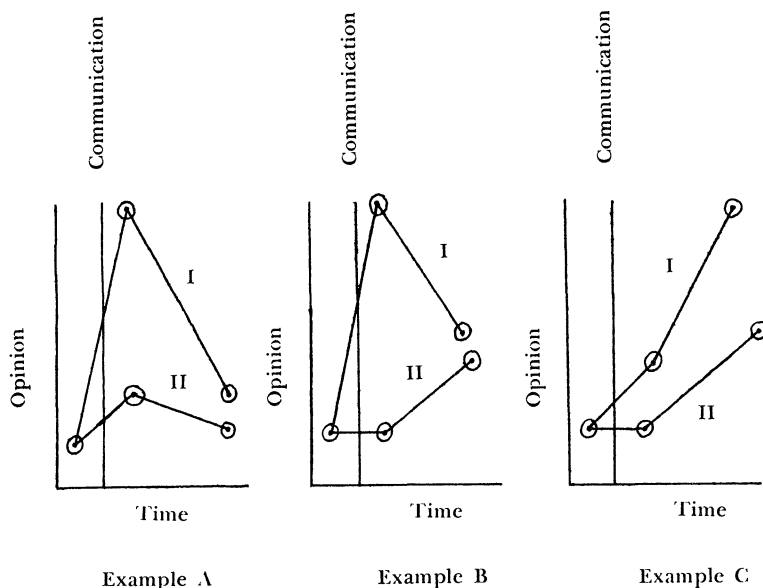
The sleeper effect definition employed in this review implies the existence of a sleeper effect for Group II in example B, and for both Groups I and II in example C. According to Weiss' definition, the required differential effect is found in examples A and B, but not in C. Thus, his definition requires rejection of a result in which positive changes were observed. Such a result is contrary to the spirit of most writing in the field.

RESEARCH DESIGN

Studies of the sleeper effect have had to cope with difficult problems of research design. Thus, not unexpectedly, some of the studies have employed designs that produce inadequate data for testing the definitions employed in this article. However, some of the designs are also inadequate for rigorously testing hypotheses incorporating the author's own definition.

The major design problems are twofold. First, only two of the studies have allowed for the possibility of opinion changes caused by effects extraneous to the experiment. That such effects can be substantial is seen by an examination of data from Hovland, Lumsdaine,

FIGURE 1



Interpretation:

	A	B	C
Reviewers Definition	No Sleeper Effect	Group II Shows Sleeper Effect	Groups I & II Show Sleeper Effect
Weiss Definition	Sleeper Effect	Sleeper Effect	No Sleeper Effect

and Sheffield's and Weiss' studies (Tables 3 and 5, respectively), which employed control groups to alleviate this problem.

The second problem is that all but three of the studies employed before-after designs, which permitted measurement contamination—a potential source of bias due to the subject's sensitization to the experimental procedure. Such contamination may be avoided by taking post-measures on different groups, to ensure that no subject is questioned more than once. The trade-off in such a procedure is the inability to monitor individual changes. However, important though this may be, all but one of the studies employed completely aggregated data in analysis, and the elimination of measurement contamination by use of more groups would have been much preferred.

Table 1 illustrates a procedure that solves both of these design problems. A sleeper effect is confirmed if:

$$(E_{t_2} - C_{t_2}) > (E_{t_1} - C_{t_1}) \quad \text{and} \quad (E_{t_2} - C_{t_2}) > C_{t_0}$$

There are practical difficulties with such a design, however, not the least of which is the requirement for a minimum of five randomly assigned groups. Further, only marginal effects may be isolated, but analysis of internal effects is possible if a more complex design is used.¹²

EXPERIMENTAL STUDIES

The genesis of the sleeper effect lies in Hovland, Lumsdaine, and Sheffield's pioneering investigation of the effects of a wartime propaganda film. They measured servicemen's opinions one week before, one week after, and nine weeks after exposure to the film, using experimental and control groups as shown in Table 2. The authors' percentage scores were converted to frequencies and are shown in Table 3.¹³ The researchers were surprised to find that on eight of fifteen selected opinion items, agreement with the communication was greater after nine weeks than after one week. They christened the phenomenon of increased favorable opinion over time the "sleeper effect," thus adding a new term to the literature.

This initial concept referred only to opinion measures taken *after* exposure. Although pre-measures on nine of the fifteen items were available, these data were neither presented nor employed in the

TABLE 1
A "GOOD" DESIGN FOR COMMUNICATIONS EXPERIMENTS

Treatment	Group				
	Control			Experimental	
	1	2	3	1	2
Questionnaire, t_0	C_{t_0}	—	—	—	—
Communication	—	—	—	X	X
Questionnaire, t_1	—	C_{t_1}	—	E_{t_1}	—
Questionnaire, t_2	—	—	C_{t_2}	—	E_{t_2}

¹² An excellent discussion of these issues appears in Hovland, Lumsdaine, and Sheffield, *op. cit.*, pp. 285-340.

¹³ Table 2 and Table 3 from Hovland, Lumsdaine, and Sheffield, *op. cit.*, p. 183 and pp. 186-187, respectively.

TABLE 2
RESEARCH DESIGN OF HOVLAND, LUMSDAINE, AND SHEFFIELD

Week of Study	Short-time Groups		Long-time Groups	
	Experimental (3 Companies)	Control (3 Companies)	Experimental (2 Companies)	Control (2 Companies)
First Week	"Before" questionnaire	"Before" questionnaire	"Before" questionnaire	"Before" questionnaire
Second Week	Film showing		Film showing	
Third Week	"After" questionnaire	"After" questionnaire		
Eleventh Week			"After" questionnaire	"After" questionnaire

analysis of the sleeper effect. Consequently, the meaning of the results cannot be explored in the context of a definition that employs pre-measures.

However, it is certainly possible to examine the data in the context of the authors' "sleepers." Since Hovland, Lumsdaine, and Sheffield had no *a priori* basis for deciding which opinion items would increase and which would decrease, the null hypothesis of no difference over time should be tested using all of the data, either on

TABLE 3
FREQUENCIES OF AGREEMENT ON OPINION ITEMS FOLLOWING
"BATTLE OF BRITAIN" FILM

Item	Short Time ^a			Long Time ^b		
	Control	Film	Difference	Control	Film	Difference
1	95	203	108	50	68	18
2	243	351	108	115	173	58
3	230	320	90	135	170	35
4	144	207	63	75	100	25
5	279	333	54	168	185	17
6	221	270	49	105	113	8
7	248	293	45	148	155	7
8	216	234	18	105	135	30
9	153	167	14	88	120	32
10	239	248	14	125	155	30
11	320	329	9	153	193	40
12	374	378	4	198	225	27
13	104	108	4	55	80	25
14	50	45	-5	30	55	25
15	252	243	-9	110	138	28
Total difference			566			405

^a Sample size of control and film = 450

^b Sample size of control and film = 250

an aggregated or dis-aggregated basis. Unfortunately the authors did not present all of their data; they chose to omit those items for which there was not at least a ten per cent difference between film and control on one of the two post-measures.

The available data illustrate the difficulties of *post hoc* analysis. If, for example, the scores for these opinion items are merged, the increments and decrements virtually balance, and short-term and long-term effects are similar. Alternatively, when the results are converted into frequencies, summed, and cast into a 2×2 contingency table, Chi square indicates significantly greater long-term than short-term change if the base taken is the number of possible agreements in the control group (see Table 4). Using other bases for contingency tests, however, can indicate lack of significant differences.

In fact, probably the most important finding of this study was that for some items opinion increased while for others it decreased. Hovland, Lumsdaine, and Sheffield devoted considerable effort attempting to explain why this should be so, but were unable to find an explanation that fully satisfied them.¹⁴

Hovland and Weiss investigated the effects of source credibility and the passage of time on opinion change.¹⁵ Their subjects, undergraduate college students, received four different written communi-

TABLE 4
FREQUENCY OF OPINION MEASURE IN AGREEMENT WITH COMMUNICATION

		<i>High Credibility Sources Agreement with Communication</i>		
<i>Topic</i>	<i>Sample Size</i>	<i>Pre- Exposure</i>	<i>Immediately Post-Exposure</i>	<i>Four Weeks Post-Exposure</i>
Anti-Histamines	31	17	24	22
Atomic Submarines	25	9	18	14
Steel Shortage	35	9	17	13
Future of Movies	31	11	15	12
Total	122	46	74	61
		<i>Low Credibility Sources Agreement with Communication</i>		
<i>Topic</i>	<i>Sample Size</i>	<i>Pre- Exposure</i>	<i>Immediately Post-Exposure</i>	<i>Four Weeks Post-Exposure</i>
Anti-Histamines	30	14	18	20
Atomic Submarines	36	12	12	17
Steel Shortage	26	8	7	11
Future of Movies	30	9	14	12
Total	122	43	51	60

¹⁴ *Ibid.*, pp., 190-200.

¹⁵ Hovland and Weiss, *op. cit.*

cations on contemporary issues. Each subject was randomly assigned two high and two low credibility attributions. Opinions (agreement/disagreement with a single question for each topic) were measured before, immediately after, and four weeks after the communications.

On the "immediately after" measure, the proportion of high credibility source responses showing opinion change in the advocated direction was larger than for low credibility source responses for each communication. Four weeks later, however, high credibility source responses showed reduced agreement for each communication. In contrast, low credibility source responses showed increased agreement for three of the four communications. Table 4 shows the relevant data (developed from Hovland and Weiss' Tables 1, 3, and 6).¹⁶

Hovland and Weiss claimed to show the existence of both a source effect and a sleeper effect. The source effect results from the increase, immediately post-exposure, to seventy-four in agreement with the high credibility sources compared to fifty-one for the low credibility sources. Evidence for the sleeper effect was adduced from the results for low credibility sources. The merged agreement scores of forty-three, fifty-one, and sixty indicate increasing agreement with the communications over time, consistent with the definition of the sleeper effect presented in this article.

However, the experimental design exhibited both the major difficulties discussed earlier. Pre- and post-measures were taken on the same subjects so measurement contamination could have occurred. Also, lack of control groups makes it impossible to distinguish between effects due to the experimental treatment and extraneous effects. Since "the topics chosen were of current interest and of a controversial type,"¹⁷ opinion changes could have occurred in the absence of experimental treatment. Thus, the four-week opinion changes from fifty-one to sixty (low credibility group) may merely reflect an increase in base-line opinion. The convergence of the high and low credibility source scores to sixty-one and sixty, respectively, is itself suggestive of a new base-line level.

Despite their claim for a low credibility increment, Hovland and Weiss themselves stated that the results of their study were restricted to group comparisons and that: "Future research must establish an effective 'neutral' base line to answer the question as to the absolute direction of the effects."¹⁸ Thus, while the authors made quite justifiable inferences of intergroup effects, by their own admission the design did not allow the testing of absolute direction. As they

¹⁶ *Ibid.*, pp., 642-645.

¹⁷ *Ibid.*, p. 637.

¹⁸ *Ibid.*, p. 648, footnote 6.

pointed out, testing for opinion increments—the sleeper effect—must await another study.

Weiss' study followed Hovland and Weiss', and did use the control groups that were considered so important in assessing absolute directions of effects. Weiss investigated the existence of a sleeper effect when subjects were explicitly influenced to discount the learned content of a communication.¹⁹ High-school subjects learned to associate the words "true" or "false" with each of eight statements on the effects of smoking. One experimental group, the "non-discounted," received no further treatment. A second experimental group, the "discounted," was immediately exposed to a brief counter-communication designed to undermine the truth or falsity of the eight statements. Opinions (agreement/disagreement with seven opinion items) were measured before treatment, immediately after treatment, three weeks later, and six weeks later.

The experimental design was similar to that of Hovland, Lumsdaine, and Sheffield (Table 2). The use of control groups enabled adjustment for base-line opinion changes. Measurement contamination was minimal since each post-measure—experimental and control—was performed on a different subgroup, although all subjects did receive the "before" questionnaire.

Weiss' results, reproduced in Table 5, indicate immediate increased agreement with the communication for both the "discounted" and "non-discounted" groups.²⁰ However, if Weiss' results are adjusted by changes in control-group opinions, then both experimental groups exhibited continuous *reduction* of agreement with the communication over time. Such decrements are incompatible with virtually all definitions of the sleeper effect, including the revised version presented at the beginning of this article.

Even with respect to Weiss' definition of the "sleeper" effect, which

TABLE 5
OPINION CHANGE SCORES FOR EXPERIMENTAL AND CONTROL GROUPS

<i>Group</i>	<i>Immediate</i>	<i>After 3 Weeks</i>	<i>After 6 Weeks</i>
A) Gross			
Discounted	.233 (N = 27)	.136 (N = 21)	.179 (N = 28)
Non-discounted	.289 (N = 47)	.102 (N = 52)	.105 (N = 53)
Control	-.017 (N = 25)	-.022 (N = 51)	.041 (N = 38)
B) Net of Control			
Discounted	.250	.158	.138
Non-discounted	.306	.124	.064

¹⁹ Weiss, *op. cit.*

²⁰ Weiss., *op. cit.*, p. 176.

requires only differential changes between the two experimental groups, the results are equivocal. First, the lack of significant differences between the opinion-change scores of the experimental groups, at each of the three post-measures, renders doubtful the efficacy of discounting. Second, in testing the opinion-change differences between immediately after and three weeks after treatment, and between immediately after and six weeks after treatment, only the latter was significant at the $p < .05$ level.²¹

In sum, Weiss' study used a sophisticated design that would have left us quite confident of the existence of the sleeper effect if opinion increments had been observed. In the absence of such increments, however, what should be inferred? It would be tempting to conclude that the sleeper effect as a generalized phenomenon of opinion increments does not exist. An equally plausible explanation is that the discounting treatment, which is quite different from Hovland and Weiss' credibility manipulation, was ineffective, so that increments would in no case have been observed.

We must conclude, therefore, that this study offers no evidence for a sleeper effect; we must also recognize the possibility that a more effective manipulation may produce such evidence.

Kelman and Hovland proposed that the sleeper effect was caused by forgetting the source over time.²² Their design required three groups of high-school students who listened to a purported radio recording on the treatment of juvenile delinquents. The communicator—who was either positive, neutral, or negative—took an “extreme, unqualified position in favor of lenient treatment.” The positive communicator was introduced as a respected judge, the neutral communicator as a randomly selected member of the studio audience, and the negative communicator as “an obnoxious, self-centered individual with a shady past and present.”²³

Opinion measures were taken immediately before, immediately after, and three weeks after the communication. Eight multiple-choice questions were asked at all three sessions, but were used only to establish group equivalency. Opinion-change measures were derived from equivalent forms of a twenty-item, Wang and Thurstone, five-point scale, which was administered twice—immediately after and three weeks after the communication. However, just before the three-week measure, one-half of each group had the communicator reinstated by a second exposure to the “broadcast” introduction.

²¹ *Ibid.*

²² Kelman and Hovland, *op. cit.*, p. 327.

²³ *Ibid.*, p. 329.

Kelman and Hovland's major results (developed from their Tables 3 and 4) are shown in Table 6.²⁴ They suggested that the "not quite significant increase" of 0.65 for the negative communicator under non-reinstatement, was evidence for a sleeper effect. Since this increase contrasted with the decrease of 1.29 under reinstatement, they concluded that the sleeper effect was caused by forgetting the source.

Although Kelman and Hovland's study followed Hovland and Weiss', it did not contribute to determining whether or not absolute opinion increments take place with the passage of time. The design did not include the control groups necessary to establish the "neutral" base line; nor did it employ pre-measure data in the analysis of the sleeper effect. In fact, evidence from the study suggests that the negative communicator caused an opinion shift contrary to the one advocated and the observed increase was merely a return to the base-line level: the opinions of the two non-reinstated groups converged to 43.48 and 43.40 and the effects of reinstatement were negative (-1.94) for the negative communicator group and positive (+1.78) for the positive communicator group, suggesting that initial exposure to the negative source might also have been negative. Such a negative effect would imply that the observed increment was merely a return to the base line and that the study offers no support for a sleeper effect.

Whittaker and Meade performed a cross-cultural study of the effects of high and low credibility sources on opinion change, both immediate and over time, using student groups in Brazil, Hong Kong, Lebanon, Rhodesia, and India.²⁵

TABLE 6
EFFECTS OF COMMUNICATION ON OPINION

Group	Treatment	Mean Scale Score		
		Immediately After	Three Weeks After	Change
Positive communicator	Not reinstated	46.70 (N = 97)	43.48 (N = 45)	-3.22
Negative communicator	Not reinstated	42.75 (N = 91)	43.40 (N = 51)	+0.65
Positive communicator	Reinstated	46.70 (N = 97)	45.26 (N = 52)	-1.44
Negative communicator	Reinstated	42.75 (N = 91)	41.46 (N = 38)	-1.29

²⁴ Kelman and Hovland, *op. cit.*, pp. 331-332.

²⁵ Whittaker and Meade. *ob. cit.*

Respondents expressed opinions on ten issues (a single seven-point scale per issue), each of which was relevant to all the nationalities included. Subjects also chose from a list of nine authority groups the ones whose opinions they would most and least respect for each issue. Two weeks later, questions were readministered, but each respondent's questionnaire was precoded to indicate the fictitious opinion of one of his chosen authority groups. On one-third of the items the most respected group was credited, on another third the least respected, and for the remainder, the denoted opinion was ascribed to the majority of students at the respondent's college. Possible order bias was controlled by rotation. Four weeks later, the same questions, with no indication of authority groups, were again administered. Results are summarized in Table 7.²⁶

The authors found no evidence for a sleeper effect: decrements in opinion were observed for both high and low credibility sources in each student group. In any event, the design of this study was less than optimal. Not only were repetitive measures taken on the same group and no control groups employed, but the prior measures of credibility were only relative, not absolute. All of the sources used seem, intuitively, to be of relatively high credibility: physicians, professors, engineers, prominent citizens, lawyers, public officials, majority opinion, businessmen, priests and other clergymen. The lack of clarity in the credibility assignment therefore rendered unlikely the support of an hypothesis linking a sleeper effect to low source credibility.

Although Insko, Arkoff, and Insko did not investigate the sleeper effect, per se, their study is included since opinion increments were observed over time.²⁷ The exposed seventh grade students of both

TABLE 7
IMMEDIATE AND DELAYED MEAN OPINION CHANGE BY COUNTRY

Country	High Credibility			Low Credibility			Majority Source		
	Immed.	Del.	p	Immed.	Del.	p	Immed.	Del.	p
Brazil	1.12	.74	<.01	.69	.63	<.001	.61	.66	ns
Hong Kong	1.54	1.08	ns	.99	.84	ns	1.00	.79	ns
Lebanon	.84	.80	ns	1.03	.85	ns	.92	.73	ns
Rhodesia	1.65	1.17	<.05	1.65	.97	<.01	1.70	1.24	<.05
India	.99	.64	<.05	1.14	.81	ns	.98	.72	ns
Means	1.19	.88	<.01	.96	.80	ns	.89	.76	ns

²⁶ Whittaker and Meade, *op. cit.*, p. 106.

²⁷ Chester A. Insko, Abe Arkoff, and Vera M. Insko, "Effects of High and Low Fear-Arousing Communications upon Opinions Toward Smoking," *Journal of Experimental Social Psychology*, Vol. 1, 1965, pp. 256-266.

sexes and varying academic aptitudes received either high or low fear communications on the effects of smoking. Their "opinion(s) about the effect of smoking on health" and "opinion(s) about future smoking behavior,"²⁸ were measured five weeks before, immediately after, and one week after the recorded communication by means of three and four item, nine-point scales, respectively. The mean opinion-change scores per issue are shown in Table 8.²⁹

Using analysis of variance, the authors detected a number of significant effects involving the time variable. For the smoking and health question, the main effect and a first-order interaction with sex resulted largely from opinion decrements. However, a first-order interaction of time with communication-type on the smoking behavior question resulted, in part, from six increments in the low fear condition. Although the authors had no prior hypotheses for these results, it was their judgement that they had emerged from post-communication, intergroup discussion—which lessened differences among fear arousal conditions. There is no valid reason to reject this argument, so adducing this study as evidence for the sleeper effect seems unwise.

Stotland, Katz, and Patchen attempted to reduce prejudice toward Negroes through the arousal of self-insight.³⁰ One of their hypotheses was that the treatments designed to produce self-insight would be

TABLE 8

MEAN BEFORE-AFTER CHANGE SCORES FOR THE TWO DEPENDENT VARIABLES: OPINION ABOUT FUTURE SMOKING BEHAVIOR AND OPINION ABOUT THE EFFECT OF SMOKING UPON HEALTH

	<i>High Aptitude</i>		<i>Medium Aptitude</i>				<i>Low Aptitude</i>					
	<i>Male</i>		<i>Female</i>		<i>Male</i>		<i>Female</i>		<i>Male</i>		<i>Female</i>	
	<i>High Fear</i>	<i>Low Fear</i>	<i>High Fear</i>	<i>Low Fear</i>	<i>High Fear</i>	<i>Low Fear</i>	<i>High Fear</i>	<i>Low Fear</i>	<i>High Fear</i>	<i>Low Fear</i>	<i>High Fear</i>	<i>Low Fear</i>
Smoking behavior												
Immediate	5.17	3.83	5.83	1.50	5.00	3.62	6.33	2.67	4.33	3.08	2.17	0.83
Delay	6.08	4.17	4.58	2.58	4.00	3.83	4.50	4.50	4.05	-0.42	2.50	1.92
Smoking & Health												
Immediate	4.00	1.75	4.75	3.58	2.17	4.50	4.08	4.17	4.33	1.00	6.58	4.00
Delay	3.92	1.75	1.92	2.42	3.00	3.58	0.92	2.67	2.50	1.17	4.50	4.17

²⁸ *Ibid.*, p. 257.

²⁹ Insko, Arkoff, and Insko, *op. cit.*, p. 262.

³⁰ Stotland, Katz, and Patchen, *op. cit.*

more effective over time than immediately, thus predicting a sleeper effect.

Attitudes of white, female, college students were measured three weeks before, immediately after, and three or four weeks after the experimental treatments. The attitude scales in the first and third sessions were identical—a 21-item Likert-type scale with fifteen useable items. However, to minimize measurement contamination, a different scale of seventeen items, matched with the original and correlating 0.84 in an independent test, was used for the second session. A subscale of four of the twenty-one items of the original scale were repeated in the scale for the second session to provide a constant measure of attitudes.

The four experimental treatments consisted of a case study that gave insight into the dynamics of prejudice. The major experimental treatment included procedures to produce active self-involvement, to appeal to the self-consistency motive, and to make the treatment relevant to prejudiced attitudes toward Negroes. Three additional “control” conditions each omitted one of these procedures. Further, seventeen subjects, not randomly assigned, participated only in the first and third sessions, consequently serving as a partial surrogate for a conventional control group.

The authors’ claim of a sleeper effect was based on the data reproduced in Table 9 (from their Tables 1 and 2).³¹ They cited as evidence the significant changes (on both scales) between the second and third sessions for the relevance control group (-3.70 , -1.22) and major experimental group (-3.15 , -1.64).

Despite design and interpretation problems, this study does provide some genuine support for the existence of a sleeper effect. The design issues relate to the lack of base-line controls, the possibility of measurement contamination with the duplicated items subscale, the reliability of the duplicated items subscale (viz. the following scores represent identical positions on the two scales: $+3.40$ and $+0.70$, -1.65 and $+0.54$, -1.85 and $+0.14$), and interaction between respondents. Furthermore, although the authors claimed sleeper effect support from both the experimental and relevance control groups, the second to third session changes for the relevance control, on the total scale at least, appear to be simply a base-line regression.

Nevertheless, for the experimental group, the results are in the right direction; for both the total and duplicated items scales they are significant, thus providing some evidence suggestive of a sleeper effect. Regrettably, however, although measures were taken of the

³¹ *Ibid.*, pp. 517-519.

TABLE 9
 MEAN CHANGE IN STEREOTYPES BETWEEN SUCCESSIVE SESSIONS AS MEASURED BY CHANGES IN TOTAL STEREOTYPE SCALE AND IN
 DUPLICATED ITEMS SUBSCALE FOR SUBJECTS IN EACH EXPERIMENTAL CONDITION^a

<i>Stereotype Measure</i>	<i>Condition</i>						<i>Total</i>	<i>Control Group</i>
	<i>Experimental</i>	<i>Rationality Control</i>	<i>Relevance Control</i>	<i>Involvement Control</i>				
Total Scale Change	+0.80(66)	-0.10(62)	+3.40(75)	-0.20(78)	+1.00(281)	—	—	
Session 1-2								
Total Scale Change	-3.15(63)	+0.15(60)	-3.70(73)	-1.65(75)	-2.15(271)	—	—	
Session 2-3								
Total Scale Change	-2.45(63)	+0.70(62)	0.00(73)	-1.85(75)	-0.91(273)	+2.90(17)		
Session 1-3								
Duplicated Items Change	+0.34(66)	+0.54(62)	+0.70(75)	-0.44(78)	+0.26(281)	—	—	
Session 1-2								
Duplicated Items Change	-1.64(63)	+0.26(60)	-1.22(73)	+0.54(75)	-0.48(271)	—	—	
Session 2-3								
Duplicated Items Change	-1.54(63)	+1.04(62)	-0.80(73)	+0.14(75)	-0.30(273)	+1.26(17)		
Session 1-3								

^a Positive scores mean more prejudiced stereotype; negative scores mean less prejudiced stereotype.

subjects' initial prejudice, other-directedness, defensiveness, and overall authoritarianism, no second to third session change scores are presented for subgroups segmented along these dimensions. Thus, differential susceptibility to a sleeper effect cannot be determined.

The following study does, however, address the issue of differential susceptibility to the sleeper effect: Stotland and Patchen's experiment used a methodology similar to that of the previous study, except that a rationality scale was included and the control group omitted.³² The major hypothesis was that if an individual considered another person prejudiced toward minority groups for ego defensive reasons, and if he perceived the other person as similar to himself on several objective characteristics, then he would be more likely to reduce his prejudice toward minority groups than if the other person had fewer similar objective characteristics. In addition, it was predicted that the hypothesized decreases in prejudice would occur only after a considerable amount of time had elapsed: a sleeper effect.

The experimental condition consisted of two booklets, each of which had two sections. The first section, common to both booklets, was a five-page selection on emotions and attitudes. The second section was a case history of "Carol," who was either similar (high sim) or dissimilar (low sim) in objective characteristics to the subjects, who were Midwestern female college students. "High sim Carol" was an American girl attending the University of Idaho and "Low sim Carol" was a student at "Canterbury University, located in the agricultural area of New Zealand."³³

The major results of the study are shown in Table 10 (from the authors' Tables 1, 2, and 4).³⁴ In support of the sleeper effect hypothesis, the authors cited significant increases in prejudice between the first and third sessions for subjects low in initial prejudice, defensiveness, and rationality, who were in the high sim condition.

However, as the authors stated, it is the second to third session changes that are important. Although they observed no significant first to second session changes, it does not, of course, follow that the second to third session changes were significant. Nevertheless, they must have been in the right direction—and therefore have deserved attention—even though no control groups were employed, the scales suffered from the same defects as those in the previous study (which were identical), and the results ran counter to the stated hypotheses.

The significance of this study is that the respondents were seg-

³² Stotland and Patchen, *op. cit.*

³³ *Ibid.*, p. 267.

³⁴ *Ibid.*, pp. 268-270.

TABLE 10
 MEAN CHANGE IN TOTAL NEGRO STEREOTYPE SCORE^a FROM FIRST TO THIRD SESSION
 IN RELATION TO:

<i>Variables</i>	<i>Experimental Condition</i>	
	<i>Low Sim</i>	<i>High Sim</i>
a) Original Degree of Prejudice		
Low Prejudice	-1.55(21)	3.65 ^c (30)
High Prejudice	-3.70(16)	-1.25(20)
b) Defensiveness Score: Position on Defensiveness Subscale		
Low	-1.35(18)	7.70 ^d (14)
Medium	-4.45(7)	-.80(25)
High	-2.50(10)	-.25(11)
c) Rationality Score: Position on Rationality Scale		
Low	-3.50 ^b (20)	2.95 ^c (31)
High	-0.65(15)	-0.35(19)

^a Positive scores mean more prejudiced stereotypes; negative scores mean less prejudiced stereotypes.

^b Change significant at .10 level on 2-tailed t test.

^c Change significant at .07 level on 2-tailed t test.

^d Change significant at .05 level on 2-tailed t test.

^e Not shown as significant, but implicitly so from text.

mented into subgroups based upon personality characteristics. Thus, the sleeper effect may not be a generalized effect, strong enough to be found in completely aggregated data like the source effect, but may rather be a weak, subtle, effect existing only in specialized subgroups of the population. Although the authors did not give data for the aggregated "high sim" respondents, collapsing of cells, estimation of subsample variance from the given significance level, and assumption of constant variance indicate that, over-all, the results are non-significant at the .10 level. However, when the "high sim" respondents are categorized along the dimensions of prejudice, defensiveness, and rationality, one subcell in each case shows significance. Thus, despite some methodological weakness, the results of this study suggest the existence of a sleeper effect, though weak and subtle in operation.

Schulman and Worrall have investigated the proposition that the sleeper effect is related to spontaneous dissociation of source characteristics and communication content.³⁵ In their design, groups of students were exposed to written communications, attributed to either high or low credibility sources, that favored adopting the British system of treating drug addiction.

Opinions were measured by a single-item, fifteen-point scale two

³⁵ Schulman and Worrall, *op. cit.*

weeks before, and immediately, four, ten and twenty-six days after the communication. When the "after" measure was taken, subjects were asked to write down anything and everything they could remember about the communication. These data provided a measure of spontaneous association/dissociation of source and content. The experimenters carefully attempted to avoid measurement contamination by ensuring that each subject responded to the "before" question and only one "after" question. However, no control groups were employed. Table 11 shows the results both for the total sample and for those who evidenced prior disagreement with the communication.³⁶ It is clear that, as stated by the authors, no increments in opinion change over time were present. Even when respondent characteristics are considered by taking initial disagreeers and partitioning them into associators and non-associators, as in Table 12,³⁷ only decrements are observed.

Since, as noted previously, Schulman and Worrall's definition of the sleeper effect was the loosest of all those reviewed in this article, it is not surprising that they claimed partial support for their hypothesis. However, it is also clear that, as more conventionally defined, a sleeper effect was non-existent in their study.

Greenwald and Gillig conducted a series of experiments in which they examined the effects of counter-argument and discounting response on persuasive communications.³⁸ They expected to find that discounting responses were less effective over time than counter-arguments in resisting persuasion attempts, and therefore that a sleeper effect would be found in the discounting condition. The discounting condition was achieved by message attribution to a low credibility source, while in the counter-arguing condition subjects

TABLE 11
MEAN OPINION CHANGE AT THREE TIME POINTS FOR HIGH AND LOW CREDIBILITY SOURCES

Source Credibility	Sample	Time Between Exposure and Post-Exposure Measure			
		Immediately After	4 to 10 Days After	26 Days After	4 to 26 Days Combined
High	Total	6.07 (27)	2.25 (69)	3.70 (20)	2.80 (89)
	Initial Disagreeers	7.57 (21)	4.25 (44)	4.92 (12)	4.39 (56)
Low	Total	3.10 (39)	2.20 (65)	2.54 (24)	2.29 (89)
	Initial Disagreeers	4.71 (17)	4.54 (37)	4.15 (13)	4.40 (50)

³⁶ Schulman and Worrall, *op. cit.*, p. 376.

³⁷ Schulman and Worrall, *op. cit.*, p. 377.

³⁸ Greenwald and Gillig, *op. cit.*

TABLE 12

SPONTANEOUS ASSOCIATION OF SOURCE AND CONTENT AND MEAN OPINION CHANGE

Source Credibility	Spontaneous Association	Time Between Exposure and Measurement of Opinion Change and Spontaneous Association			
		Immediately After	4 to 10 Days	26 Days	4 to 26 Days Combined
High	Associators	7.56 (9)	1.17 (7)	4.75 (4)	2.82 (11)
	Non-Associators	7.58 (12)	4.73 (37)	5.00 (8)	4.78 (45)
Low	Associators	3.50 (12)	-.50 (2)	-1.00 (1)	-.67 (3)
	Non-Associators	7.60 (5)	4.83 (35)	4.60 (12)	4.77 (47)

received an additional counter-arguing defense prior to the communication. Discounting subjects received a prior irrelevant communication.

Communications, which were videotaped, addressed health truism topics and were attributed to a low credibility source: "Practicing Nature Therapists."³⁹ (Some experiments were conducted with high credibility attribution but they are not germane and will not be discussed.) Opinions were measured just after and approximately two weeks after each communication, either in the laboratory or by telephone. Four-item, fifteen-point scales were employed.

The results are summarized, in Table 13.⁴⁰ A reduction in the opinion score implies increased agreement with the communication, which is in the right direction for a sleeper effect.

In no case did Greenwald and Gillig find a significant opinion difference to support a sleeper effect, for either the discounting condition or the counter-arguing defense condition. Even though they obtained a right direction effect for the penicillin topic in the "no defense" discounting condition in Experiment 1, they were unable to replicate this. They fared no better with other topics, although one additional right direction effect was obtained as hypothesized. The experimental designs were not optimal, although they improved over time. A control group, which could also function as a pretest, was introduced in Experiment 2, although long-term, post-measures were never taken to allow tracking of the base line. The possibility of measurement contamination, which was present in Experiments 1-4, was eliminated in Experiment 5 by the introduction in each condition of an additional group that received only the long-term opinion questionnaire.

³⁹ *Ibid.*, p. 4.

⁴⁰ Greenwald and Gillig, *op. cit.*, pp. 9-10.

TABLE 13
SUMMARY OF GREENWALD AND GILLIG

Experiment	Topic	Treatment	Mean opinion		Significant Difference in Right Direction	Per Cent of Respondents Changing Opinion in Right Direction
			T ₁	T ₂		
1	Dangers of Pencillin	Defense	10.4	11.3	No	—
		No Defense	10.5	8.6	No	—
2	Undesirability of Annual Medical Checkups	Defense	12.3	12.0	No	—
		No Defense	11.8	12.4	No	—
3	Dangers of Pencillin	Defense	12.6	12.5	No	-25%
		No Defense	11.5	11.5	No	+29
		Control	12.0	—		
4a	Desirability of Large Vita- min C Doses	Defense	12.5	12.5	No	+3
		No Defense	11.6	11.6	No	+9
		Control	12.6	—		
4b	Undesirability of Annual Medical Checkups	Defense	8.7	7.4	No	+27
		No Defense	5.7	4.9	No	+34
		Control	5.0	—		
5a	Dangers of Pencillin	Defense	11.2	11.2	No	—
		No Defense	7.3	10.2	No	—
		Control	11.7	—		
5b	Undesirability of Annual Medical Checkups	Defense	11.3	11.7	No	—
		No Defense	6.9	7.4	No	—
		Control	13.2	—		

The authors' conclusion from the total series of experiments was that "the sleeper effect is, at best, an unreliable phenomenon."⁴¹

One additional piece of evidence is available, however. Greenwald and Gillig have presented some data on the proportion of respondents who changed opinion between T₁ and T₂ (shown in the final column of Table 13). In Experiment 3, the 29 per cent of respondents who changed opinion in the advocated direction in the no-defense condition is "reliably greater than zero."⁴² From an examination of sample size, and assumption of constant variance, it can be inferred that the 27 per cent and 34 per cent of respondents who changed in Experiment 4b are also reliably greater than zero.

Temptingly, this evidence is congruent with the results of Stotland and Patchen's study: a positive effect may not be strong enough to be found over-all, but it may exist for a subset of the population. However, since Greenwald and Gillig did not take premeasures, we can-

⁴¹ *Ibid.*, p. 7.

⁴² *Ibid.*, p. 5.

not say how much of the change is a return to an original position and how much is a true effect. It must be concluded that little, if any, positive evidence for a sleeper effect may be gleaned from this study.

SUMMARY AND CONCLUSIONS

This article has examined a variety of research studies in which the concept of the sleeper effect was addressed. A series of definitional issues have been raised and a common definition, consistent with most of the studies, has been used to evaluate the empirical evidence. The pattern of findings is far from conclusive—either for or against the existence of a sleeper effect. Although a hard-nosed researcher may argue that no unambiguous evidence for a sleeper effect has been uncovered (see Table 13), a more reasoned approach is that such an effect may exist, although it is subtle in operation.

Failure to capture a sleeper effect in a consistent manner is no longer surprising when the various research methodologies employed for the studies are arrayed (see Table 14). The familiar source-message-channel-receiver paradigm illustrates the variety of designs employed. Although college and high-school students were typical audiences, there was no such similarity for source, message, and channel; nor for the dependent variable measure. Five of the studies employed written communications; one was filmed, one personal, two audiotaped (one of these was accompanied by slides), and one videotaped. Message topics ranged from current events to sensitive issues such as racial prejudice. In some designs, treatments involved manipulation of source attribution; in other cases the treatment variable was counter-communication or other forms of message modification. Measurement of opinion was either a head count of respondents in the relevant categories, or it was derived from scale measures. The scales varied from a single-item, fifteen-point scale to a twenty-item, five-point scale.

Not only were the designs employed extremely varied, they also often exhibited methodological flaws: most commonly, the existence of real increments could not be concluded, because there were either no pre-measures or no control groups to measure base-line changes. In addition, many studies employed repetitive measures, thus allowing the possibility of measurement contamination.

However, if, despite the methodological problems, such varied experimental conditions had produced consistent, significant increments between t_1 and t_2 , the impartial observer would infer that there was strong empirical evidence for a sleeper effect. In the absence of

TABLE 14
SUMMARY OF STUDY RESULTS, EXPERIMENTAL DESIGN, AND ANALYSIS

Study Authors	Results				Experimental Design			
	Sleeper Effect Claimed	Any Increment Found $0_t > 0_b$	Significant Increment Found $0_t > 0_b$	Significant Increment Found $0_t > 0_b$, and $0_b > 0_c$	Control Groups	Prior Measure	Post-Measures on Different Groups	Scales Appear Trustworthy
Hovland, Lumsdaine, & Sheffield	Yes	Yes	No	No	Yes	No	Yes	Yes
Hovland & Weiss	Yes	Yes	No ^a	No ^a	No	Yes	No	Yes
Weiss	Yes	No	No	No	Yes	Yes	Yes	Yes
Kelman & Hovland	Yes	Yes	No	No	No	No	No	Yes
Whittaker & Meade	No	No	No	No	No	Yes	No	Yes
Insko, Arkoff, & Insko	—	Yes	Yes	Yes	No	Yes	No	Yes
Stotland, Katz, & Patchen	Yes	Yes	Yes	Yes	(No) ^b	Yes	No	No
Stotland & Patchen	Yes	Yes	No	No	No	Yes	No	No
Schulman & Worrall	Yes	No	No	No	No	Yes	Yes	Yes
Greenwald & Gillig	No	Yes	No	No	No	No/Yes	No/Yes	Yes

^a Not tested by authors.

^b Control group present but non-random and results not used.

TABLE 14 CONTINUED

<i>Study Authors</i>	<i>Analysis</i>					<i>Unambiguous Evidence For Sleeper Effect?</i>
	<i>Random Phenomena</i>	<i>Base Line Change</i>	<i>Inadequate Scales</i>	<i>Measurement Contamination on Post-Measures</i>		
Hovland, Lumsdaine, & Sheffield	Yes	Yes	No	No	No	No
Hovland & Weiss	Yes	Yes	No	Yes	Yes	No
Weiss	—	—	—	—	—	No
Kelman & Hovland	Yes	Yes	No	Yes	Yes	No
Whittaker & Meade	—	—	—	—	—	No
Insko, Arkoff, & Insko	Yes	Yes	No	Yes	Yes	No
Stotland, Katz, & Patchen	No	Yes	Yes	Yes	Yes	No
Stotland & Patchen	Yes	Yes	Yes	Yes	Yes	No
Schulman & Worrall	—	—	—	—	—	No
Greenwald & Gillig	Yes	Yes	No	Yes	Yes	No

TABLE 15
SUMMARY OF SOURCES, MESSAGES, CHANNELS, RECEIVERS, AND MEASURES OF OPINION AND OPINION CHANGE FOR EACH STUDY

Authors	Treatment				Topic	Channel	Receivers	Opinion Measure	Opinion Change Measure
	Source		Message						
	Number	Credibility Manipulation	Number	Manipulation					
Hovland, Lumsdaine, and Sheffield	1	Low	1	—	Propaganda	Film	Army Recruits	Questions Agreement/Disagreement	Per Cent Respondents
Hovland & Weiss	8	{4 High 4 Low	4	—	Current Events	Written	College Students	Single Item per Topic Agreement/Disagreement	Net Per Cent Respondents
Weiss	—	—	1	Discounted	Smoking	Personal	High-School Students	Seven Items, Agreement/Disagreement	Mean Difference
Kelman & Hovland	3	{High Neutral Low	1	—	Juvenile Delinquency	Audio Tape	High-School Students	Twenty Items, Five Point Scale	Mean Difference
Whittaker & Meade	3	{High Low Peer	10	—	Current Events	Written	College Students	Ten Items Seven Point Scale	Mean Difference
Insko, Arkoff, & Insko	—	—	2	High Fear Low Fear	Smoking	Audio Tape Plus Slides	Junior High-School Students	1) 3 Item (Nine Point Scale) 2) 4 Item { Scale	Mean Difference
Stotland, Katz, & Patchen	—	—	4	Varying Self Insight Induction	Racial Prejudice	Written	College Students	Fifteen Items Agreement/Disagreement	Mean Difference
Stotland & Patchen	—	—	2	Similarity/Dis-similarity to Model	Racial Prejudice	Written	College Students	Fifteen Items Agreement/Disagreement	Mean Difference
Schulman & Worrall	2	{High Low	1	—	Drug Treatment	Written	College Students	Single Item Fifteen Point Scale	Mean Difference
Greenwald & Gillig	2	{High Low	7	Counterargument Defense/No Defense	Health Truisms	Videotape	College Students	Four Item Fifteen Point Scale	a) Mean Difference b) Per Cent Respondents

such results, the various designs would be better viewed as confounding the goal of inference.

In pointing out the design variations as sources of noise, we are suggesting that, although the evidence clearly disfavors a strong and general sleeper effect, it would not be capable of observation over such a wide range of conditions if it were subtle and weak. Thus, only under certain restrictive design conditions might it be observed. Further, there is evidence that such an effect may be found only with subsets of the population, divided on the basis of personality characteristics.

Stotland and Patchen's study provides evidence for this suggestion. Whereas their aggregated data did not support the sleeper effect hypothesis, when categorized along the dimensions of prejudice, defensiveness, and rationality, one subcell was significant in each case. Further support for this possibility may be adduced from the experiments of Greenwald and Gillig. While they were unable to isolate significant increments using the five-item scale measure as the dependent variable, the percentage respondent measure—which gives equal weight to each respondent—showed significant changes in three out of four cases. This suggests that a subset of individuals moved in a positive direction, although the amount of movement, when measured on a scale, was insufficient to offset negative movement and allow significance. Both studies, however, employed repetitive measures.

In this regard, it should be noted that increments were observed in Hovland, Lumsdaine, and Sheffield's and Hovland and Weiss' studies, which both employed percentage respondents as the dependent measure; but they were not observed in Weiss,' Whittaker and Meade's, or Schulman and Worrall's studies, which used scale measures.

Our final conclusion from the agglomeration of studies and evidence is as follows: there is no strong evidence for a generalized sleeper effect and we doubt that one exists. What evidence can be gleaned from the studies suggests that a sleeper effect phenomenon may be detected for certain subsets of the population. However, such evidence is, at best, weak; further studies, in which more attention is paid to design detail, are required.