

THE BOGUS PIPELINE:

A NEW PARADIGM FOR MEASURING AFFECT AND ATTITUDE¹

EDWARD E. JONES²

AND

HAROLD SIGALL

Duke University

University of Rochester

If subjects can be convinced that a physiological monitoring device is able to measure both the amplitude and direction of emotional response, their subsequent attempts to predict what the machine says about their attitudes should be uncontaminated by many of the biases that obscure paper-and-pencil measures. This assumes that subjects do not want to be second-guessed by a machine. Four experiments are described in which variations of this "bogus pipeline" paradigm were employed. In each case it proved quite possible to validate the device in the subject's eyes and to extract theoretically meaningful estimation measures of interpersonal attraction. It is not clear whether the differences between such estimation measures and the standard verbal measures of attitude are based simply on a lie detection principle, on a principle of error preference, or on the fact that the affective component of a complex attitude has been made salient. For whatever reason or reasons, subjects attached to the bogus device appear much more ready to express negative affect in experimental settings where one might normally expect the inhibition of such feelings, for instance, toward Negroes and the physically handicapped.

Many psychologists for many years, beset with the vexing difficulties associated with inferring true feelings from behavior, must have had fantasies about discovering a direct pipeline to the soul (or some nearby location). Wouldn't it be nice if people really did wear their hearts on their sleeves? Or even if we had access to reliable and valid physiological indicators of attraction and dislike, approach and avoidance, or belief and disbelief? But alas (and some would say, thank goodness), no one has yet achieved this last breakthrough on the path to invasion of privacy. The present study describes a paradigm that appears to bypass some of the major problems of inference from behavior and poses new problems for substantive research along the way. The paradigm is based on the simple premise that no one wants to be second-guessed by a machine. If a person could be convinced that we *do* have a machine that precisely measures attitudinal direction and intensity, we assume that he would be mo-

tivated to predict accurately what the machine is saying about him. The bogus pipeline paradigm involves, then, the use of a device or machine that purportedly measures one's true feelings about a person or an issue, some means of validating the machine in the subject's eye, and finally the dependent variable itself: a prediction by the subject of the machine's telltale reading.

Before proceeding with a detailed discussion of this paradigm and its possibilities, it may be useful to place the procedure in the broader context of measurement in social psychology. Throughout much of the 1930s it could be reasonably argued that the measurement of attitudes was the central preoccupation of social psychologists. There was little theory to speak of, only an occasional experiment to report, and the notion of theoretically relevant situational manipulations was only beginning to take hold in the aftermath of Lewin, Lippitt, and White's (1939) studies on autocratic and democratic leadership. With the development of this manipulative experimental tradition during the 1940s and 1950s, the interest in attitude measurement was pushed from center stage. Especially the Festingerians (cf. McGuire, 1969) lavished great care on the independent variables being manipulated, while the dependent vari-

¹The research reported in this paper was supported by National Science Foundation Grant G8857 to the first author and by National Institute of Mental Health Grant MH17180-01 to the second author.

²Requests for reprints should be sent to Edward E. Jones, Department of Psychology, Duke University, Durham, North Carolina 27706.

ables used were almost a casual afterthought. Instead of exploring the antecedent conditions that affected some standardized, reliable, and socially important dependent variable, experimenters in the Festinger tradition typically asked simple, direct, tailor-made questions to get at the effects of theoretically relevant situational variations. The consequences of elaborate experimental scenarios were often measured by a check on an *n*-point scale from "like" to "dislike" or a decision to wait alone versus wait with somebody else. Standardization, multimethod triangulation (cf. Campbell & Fiske, 1959), and cross-situational generality were typically sacrificed in favor of contextual relevance—measures introduced in as unobtrusive, "natural," and fitting a manner as possible.

Without in any way denying the importance of other research styles and strategies (e.g., McGuire's, 1969, Hovlanders with their search for interesting antecedents of standard dependent variables), we think the emphasis on contextual relevance has been generally well placed. It is hard to quarrel with the notion that the subject should respond to a reasonable question that fits into the cover story defining the purposes of the research, and that he not be dragged into a morass of complex and obtrusive psychometric measures. But there are obviously problems in assessing the answers that subjects give to any research question, and many of these problems are exacerbated by the one-shot direct probe. In the following section, we review some of these difficulties.

Potential Difficulties with Rating-Scale Measures of "True Feelings"

While many of the following difficulties apply to the general case of attitude measurement, we turn to the attraction area for our examples and consider ratings made by the subject of a target person.

1. *Generosity effect.* There is a common tendency for raters to give target persons the benefit of the doubt in recording judgments with obvious evaluative implications. On the kind of antonym scales often used in impression research, for example, it is rare to find subjects using the negative halves of scales like generous–stingy, wise–foolish, or

friendly–hostile. The same is true, though to a smaller extent, with Likert-type items involving statements with which the subject is asked to express degree of agreement or disagreement (e.g., "She is rather on the stingy side"). This general tendency to give positive ratings is probably an expression of the realization that it never pays to risk giving offense when it is easy not to. We assume that many subjects do not want to appear odd or extreme and choose a moderately positive rating as a "place to hide" on the scale.

2. *Evaluation apprehension.* This latter consideration merges into a more general interest of the subject to be evaluated by the experimenter as normal, mature, perceptive, and full of integrity. "Evaluation apprehension" has been highlighted by Rosenberg (1965) and by Sigall, Aronson, and Van Hoose (1970) as a potential source of artifactual error in measuring the dependent variable in an experiment. Similarly, Edwards (1957) and Crowne and Marlowe (1964) have noted that the social desirability of a response may affect whether or not it is emitted. Depending on the subject's interpretation of the situation, his desires to be well evaluated may lead him to play it safe, to show how mature he is by not responding with a negative rating to a critical attack, or to show how perceptive he is by letting his ratings reflect small nuances in the target person's behavior. In short, the subject's evaluation apprehension may have diverse and context-specific effects on his ratings. In general, the subject will try to respond as he thinks a mature and rational person would.

3. *Experimenter demand.* Closely related to the notion of evaluation apprehension is the tendency that Orne (1962) and Riecken (1962) have posited for subjects to try to please the experimenter or make a special effort to confirm what they perceive to be his hypothesis. To the extent that this tendency is operative, it may obviously affect the subject's ratings of the target person—with or without his conscious realization.

4. *Thoughtlessness.* A major difficulty with paper-and-pencil measures of all sorts is that it is too easy not to care, not to be concerned about the validity of one's answers. There is no clear penalty for casual sloppiness, for

checking only the extremes or only the neutral points, rather than taking the trouble to be discerning.

5. *Errors of "psychologic."* One of the classically noted rating errors is that of the halo effect—the tendency for ratings on different dimensions to be more highly correlated than warranted by objective data. Halo effects are often seen in the evaluative domain, so that a target person who is rated positively on one trait is likely to be rated positively on other traits. But halo effects may also reflect more subtle rating biases that map the general connotative relations between trait names or the implicit personality theories held by the rater. Thus, if the rater believes that people who are generous also tend to be warm, he will rate the target person similarly on both dimensions—in spite of the fact that the target person may have behaved in a warm and stingy manner. A different kind of psychologic error may be seen in the rater's tendency to restore balance in his ratings, to rate friends of friends and enemies of enemies more positively than they deserve. Both kinds of error may be more common with paper-and-pencil rating measures than with measures that commit the subject to action or reflect an involving social relationship. Once again, it is easier on such measures, perhaps, to rate people the way they should ("logically") be rated rather than in line with one's true inner feelings.

Difficulties with Behavioral and "Behavioroid" Measures of Attraction

Strictly speaking, all measurement in psychology involves measurement of behavior, but one can conceive of a continuum of consequentiality for the subject (Aronson & Carlsmith, 1969). At one end of the continuum would be those actions that commit the subject to costly, effortful, long-range, or for any other reason personally important consequences. At the other end might be the anonymous questionnaire. Since truly consequential behavior measures are typically difficult and impractical to arrange in a psychology experiment, investigators have often resorted to what Aronson and Carlsmith (1969) have called "behavioroid" measures—distinguished from true behavioral measures

in that the subject commits himself to consequential action without actually going through with it. From the subject's point of view, a behavioroid measure is the same as a behavioral measure at the time of commitment. If properly convinced, he truly expects to serve under the group member he votes for as leader (Schachter, 1951), or to escort visiting Negroes around campus (Marlowe, Frager, & Nuttall, 1965), or to have a large "drive carefully" sign displayed on his front lawn (Freedman & Fraser, 1966). The important consideration is that he expects to have to live with the very real consequences of his behavior, and therefore is hardly likely to take his response lightly or be much affected by evaluation apprehension or experimenter demand.

The most obvious advantage of behavioral and behavioroid measures of attraction is the likelihood that the subject will respond in a way more readily generalizable to the involving situations of real life. But there are disadvantages as well. If the measure is behavioroid, there is always the lingering suspicion that the subject does not really expect to follow through on his commitment. A more important problem is that behavior (and behavioral commitment) can be multiply or alternatively determined. In studies where the behavior itself is the sufficient focus of study (as in the effects of social pressure on volunteering behavior, or the role of frustration in determining the sociometric rejection of ethnic minorities), this problem is not of overriding importance. When behavior or effortful commitment is used as an index of attraction, however, the problem may be acute. A requirement of sociometric measures, according to Moreno (1934), is that subjects make their choices fully understanding that they can lead to restructuring the group. The original sociometric test was, then, a behavioroid if not a behavioral measure. But one can imagine many reasons that might underly a sociometric choice: The chooser feels that the group should be a mixture of good guys and bad guys, and since others are likely to choose nothing but good guys, it is up to him to choose a bad one; the chooser feels sorry for someone he expects others to reject (Bramel, 1969); the chooser may be

more fearful of the target of his choice than of others and be trying to appease him.

Similar difficulties arise with other behavioral measures of attraction. We may reward another, not because we like him, but because we feel sorry for him. We may shock the student in a teacher-student paradigm, not because we dislike him, but because we think he is intelligent enough to profit by clear feedback. We may agree to circulate petitions or make phone calls for another person because it bolsters our own self-esteem to serve a worthy cause, rather than because we like the requester. We may prefer to meet again with someone to convert him to our position, rather than because we find him especially attractive. As the consequences of behavior become more important, the person will presumably pay less and less attention to the picture he presents to the experimenter. Nevertheless, as Cook and Selltitz (1964) pointed out, even in a genuine behavioral setting in which choices or commitments are expected to have consequences, "an individual may be motivated to give a response that differs from his spontaneous private one, in order to present himself . . . as unprejudiced or to maintain his own image of himself as one who behaves in an unprejudiced way [p. 45]."

The division of measures into paper-and-pencil and behavioral hardly exhausts the various ways in which measures of affect and attitude can be measured. Both behavioral and questionnaire measures may be direct or indirect. A direct measure (to paraphrase Scott, 1969) is one in which the subject and the investigator are assumed to attach the same meaning to the question or stimulus and infer the same significance of the answer or response. There are many ways to disguise a measure of attitude or attraction. Some of these indirect measuring techniques are administered as part of a mild diversion, and some are projective and involve displacements of stimuli or responses. As an example of the former, students of interpersonal attraction often stress their interest in how first impressions are formed, but develop indexes of attraction from the first-impression ratings by adding scores on all the evaluative traits (e.g., Jones & deCharms, 1957; Jones, Stires,

Shaver, & Harris, 1968). This is not too different from the common procedure of measuring attitudes toward an object by measuring beliefs about it. Also in the attitude realm, similar but more clearly projective use may be made of what Campbell (1950) has called "disguised-structured" tests. An example here would be the test that purports to measure knowledge or information about a domain of attitudinally relevant facts, which is scored in terms of the prevalent direction of bias in the wrong answers given.

As an example of projective measures of attraction, Gerard (1965) used a Group Picture Impressions Test devised by Libo³ to obtain an indirect measure of subjects' evaluations of other subjects in a conformity situation. The test involves writing stories to four stylized pictures showing an individual in different group situations. These stories were then content-analyzed to determine approach or avoidance reactions to others by the main character in each of the pictures. The resulting algebraic sum of approach minus avoidance statements was used as a measure of each subject's attraction to those in his present situation, with results confirming Gerard's particular theoretical hypothesis.

The problem with indirect or projective measures of attraction is that they require careful validation, and the experimental situations in which they are eventually used may differ in important ways from the original validation settings. Even with tests of fairly high validity there is bound to be substantial error variance contributed by idiosyncratic interpretations of the items or pictures, by level of verbalization, and by various response styles.

Physiological Measures of Attitude and Attraction

It is hardly surprising that for many of the reasons implied by the above discussion, investigators have sought to discover physiological indicators of attitude. It is usually assumed that such indicators are less subject than verbal responses to voluntary distortions and defensive denial. Although they were well

³Libo, D. Measuring group cohesiveness. Ann Arbor, Mich.: Research Center for Group Dynamics, 1953.

aware of previous unsuccessful attempts to use the galvanic skin response (GSR) in the diagnosis of social attitudes, Rankin and Campbell (1955) obtained highly significant GSR-level differences between subjects responding to contact with white versus Negro experimenters. Although Porier and Lott (1967) were unable to replicate this finding in a better controlled study, they did find a low but significant correlation between GSR reactivity to Negro contact and ethnocentrism scale score. Westie and DeFleur (1959) recorded finger-pulse volume as well as GSR as subjects were shown pictures of Negroes and whites flashed on a screen. Prejudiced subjects (as measured by a social distance questionnaire) showed greater GSR reactions but lowered pulse rates to slides depicting Negroes. In a similar study, Cooper (1959) found that GSR magnitude to ethnic names like "Mexicans," "Swedes," and "Japanese" could be predicted from verbal reactions to the same labels. Hess (1965) reported that pupillary dilation seems to correlate with positive affect.

These and similar studies (cf. Leiderman & Shapiro, 1964; Shapiro & Crider, 1969) exploring autonomic indicators of affective arousal may shed some light on the structure and patterning of attitudes when the indicators are viewed in the context of specific experimental procedures and related to the pattern of conventional verbal responses to the same situation. As a direct measure of "true" attitude, however, physiological indicators are judged to be crude and responsive to too many variables. For one thing, there is increasing evidence that physiological variables are affected by such cognitive factors as instructional set, suggestion, and cognitive dissonance. Gerard (1964) measured finger-pulse amplitude changes during a period in which subjects were to decide which of two paintings they wished to have. By assuming that pulse rate should be high when the subject was experiencing dissonance, Gerard was able to provide a rough physiological charting of the decision process featuring dissonance elevation and reduction. Zimbardo, Cohen, Weisenberg, Dworkin, and Firestone (1969) also showed that predicted dissonance arousal and reduction effects can be tracked

by changes in GSR. In one sense these studies emphasize the considerable promise of using physiological indicators to reflect cognitive changes, but they also suggest that it is misleading to contrast voluntary control over verbal responses with involuntary physiological indicators.

A second and more obvious problem with the use of physiological measures of attitude and attraction is the difficulty of inferring affective *content* or *direction* from such measures. As Westie and DeFleur (1959) noted:

It cannot be assumed that conspicuous physiological changes manifested by an *S* [subject] upon being exposed to a stimulus object indicates simple favorability or unfavorability toward the attitude object, regardless of his scale-measured attitudes toward that object. It is easy to fall into thinking, especially when the stimuli are "race objects," that high response indicates a negative attitude [p. 346].

Shapiro and Crider (1969) raised a number of doubts concerning the use of physiological measures even in the detection of stress, arousal, tension, or activation. It is obviously hazardous in the extreme to make the still further assumptions that would be required to use GSR or pulse-rate measures as indicators of the direction and intensity of attitude.

For purposes of the bogus pipeline measure now to be described, however, the crucial consideration is the *plausibility* of having physiological measures of one's true attitudes and feelings. Here our task is greatly facilitated by public knowledge and stereotypes about the "lie detector" and its use in criminal investigations. According to Sternbach, Gustafson, and Colier (1962), lie detection by polygraph is still in the state of very primitive art, suffused by unfounded and often fraudulent claims concerning reliability and validity. But Lee (1953) emphasized the preventive utility of lie detection devices and operators because of the fact that many would-be larcenists and perjurers believe in the mystique and infallibility of lie detection. He retold the story of the prince who rounded up suspects and told them that each was to go into the next room blindfolded and pull the tail of the ass therein. The ass would bray when his tail was pulled by the guilty party. The prince examined the hands of each suspect after they had com-

pleted, in turn, their quiet visit to the next room. The one suspect whose hands were not black was accused of being the guilty party. The ass's tail had been dusted with black powder, and the prince assumed that the guilty party would only pretend to pull the telltale tail. The moral of this story for our purposes is that if people can be made to believe that there are devices reflecting their true inner attitudes or feelings, the measurement possibilities are almost limitless.

VERSIONS OF THE BOGUS PIPELINE

The idea of exploiting false physiological feedback in a social psychology experiment is not new. Unlike the present proposed use of the bogus pipeline concept, however, earlier uses have manipulated feedback as an independent experimental variable. Perhaps Gerard⁴ was the first to see the possibilities of such feedback in controlling certain assumptions that the subject was making about himself and his proclivities. Gerald attached electrodes to the forearms of subjects in a standard conformity setting requiring judgments of the larger of two multipointed stars. Each subject was led to believe that the electrodes were attached to highly sensitive electromyographic equipment that was capable of measuring "implicit muscle movements" in the forearm. Since the response task was to depress a button with either the left or the right index finger, this measure could plausibly reflect the subject's first impulse to choose one or the other star in the situation. By letting the subject see the first impulses of others as well as his own first impulse (all fictitious, of course) on meters in the subject's display panel, it was possible to convince the subject that he had a tendency to yield or be autonomous, regardless of his overt, conscious preferences in the situation.

Bramel (1962) convinced subjects that degree of homosexual arousal to pictures of men in various states of undress was being reflected on the meter in front of them. Brehm and Behar (1966) used a similar feedback technique to convince their subjects that

they were or were not sexually aroused to line drawings of a turtle, cup, flag, table, and rosebush. Valins (1966) played back amplified (fictitious) heart beats to subjects while they observed slides depicting nude women. Gerard and Rabbie (1961) and Behar (1967) were successful in making subjects think that they were more or less fearful in a shock-anticipation setting where they were allowed to observe their own (bogus) GSR readings. In each of these studies, the feedback information was a major part of the independent variable manipulation defining the various experimental conditions. The feedback was embedded in a reasonable cover story, and there was no apparent problem of suspicion.

We consider the bogus pipeline to be a dependent variable paradigm, rather than a specific device or machine for measuring attitudes and feelings. As indicated in the introduction, the paradigm involves convincing the subject that there is a machine that provides valid indication of his true feelings, hooking him up to such a machine, introducing theoretically interesting stimulus conditions to the subject while he is hooked up, and finally asking the subject to estimate the machine reading in terms of a scale presumably paralleling the summary readout of the machine. Thus far, we have conducted experiments with two different kinds of feedback machinery that illustrate the pipeline paradigm: the directional or "differentiated" GSR and the electromyograph. We present some procedural details and results for each of these devices to demonstrate their utility and certain special problems associated with their use.

Galvanic Skin Response "Breakthrough"

Jones, Bell, and Aronson (1971) conducted an experiment investigating the combined effects of opinion similarity and being liked on the reciprocation of attraction. In short, how do we feel about similar people who like versus dislike us? How about dissimilar likers and dislikers? Jones et al. predicted that subjects will be (a) more attracted to a dissimilar liker than a similar liker and (b) less attracted to a similar disliker than a dissimilar disliker. Their reason for predicting such a main effect for dissimilarity was the special

⁴ Gerard, H. B. Acts, attitudes and conformity. (Symposia Study Series No. 4) Washington, D. C.: National Institute of Social and Behavioral Science, 1960.

significance attached to expectations that are violated in a pleasant (being liked by a dissimilar person) versus an unpleasant (being disliked by a similar person) direction. We should be especially inclined to reciprocate attraction when we have unexpectedly gained the admiration of a dissimilar other, and we should be especially prone to dislike someone who dislikes us in spite of the opinions we share in common. When no information about being liked or disliked is provided, we should simply like the similar better than the dissimilar other.

The experiment was presented to the subjects as a study of the process of impression formation. Through a series of complicated but plausible procedures, each subject learned whether another subject was similar or dissimilar in attitudes and then learned whether the other subject liked or disliked her after watching her being interviewed by the experimenter. In some cases (control condition), information about being liked was not provided. As a final step in the experiment, the subject observed a videotape of the other subject being interviewed. The interview had been previously taped by an experimental confederate. At no point during the experiment was the confederate actually present.

Throughout her observation of this last videotape, the subject was hooked up to a GSR machine, the properties of which had been explained to her previously. The main dependent variable measure of attraction was obtained by asking the subject to try to predict the final meter reading of the machine as she thought about the other subject. By this time, of course, she knew whether the other subject was similar or not and whether the other subject liked her after observing the interview. The machine was originally presented to the subjects as a "new kind of galvanic skin response—you know, GSR—machine." The novel feature of the machine was allegedly the result of a "breakthrough," enabling it to measure not only the magnitude of emotional arousal but to distinguish between positive and negative emotional states, between like and dislike. The subject was then shown a standard GSR machine, and her attention was drawn to the summarizing dial that reflected an "integration" of the

TABLE 1
ESTIMATES OF "OWN GSR" ATTRACTION
READING (Experiment 1)

Attitude	Being liked			Total
	Yes	No	Uncertain	
Similar	1396	-642	1392	715
Dissimilar	1852	- 38	981	932
Total	1624	-340	1187	

Note.—The meter dial on which estimates were to be based ran from -3000 to 3000.

subject's feelings at a given point in time on a scale ranging from -3000 to 3000.

A number of steps were taken to validate the meter in the subject's eyes. She observed the accomplice (again via prerecorded tape) apparently hooked up to the GSR while thinking about her "best friend" and then while considering someone she disliked or was mad at, during which time there were appropriate movements of the visible meter toward the positive and then toward the negative end of the dial. She herself went through the same best friend, disliked-person procedure and, though not allowed to see her own meter movements, the experimenter assured her that the machine appeared to be in working order. Finally, in order to encourage subjects to do their best in estimating the meter readings, the experimenter stressed that the accomplice had been very accurate in estimating *her* own meter readings, and the experimenter was convinced and hopeful that the subject could be accurate too.

Although the meter estimation was the primary measure of the subject's attraction for the accomplice, each subject was also asked to fill out a first impression questionnaire, the major portion of which consisted of twelve 15-point trait rating scales, each expressed in antonym form (intelligent-unintelligent, friendly-unfriendly, etc.)

The results of this experiment are discussed in detail in the original publication, and for present purposes their methodological implications are stressed. The mean estimates of "own" GSR attraction reading are presented in Table 1. Although the predicted interaction is only of borderline significance ($p < .07$), the means are all in the proper order: If no in-

formation about being liked is reported, the similar accomplice is better liked than the dissimilar one; if information about being liked *or* disliked is provided, the dissimilar person is preferred. Although in all cases the questionnaire followed the GSR estimate, and might therefore be expected to be somewhat contaminated by it, the impression ratings did not confirm the prediction or even fall in the right direction. One cluster of items designed to measure "social attraction" showed a general overall (nonsignificant) preference for the similar accomplice. With both measures, not surprisingly, there was a highly significant preference for the liking over the disliking accomplice.

In a replication of this experiment (with a different accomplice made up to be less attractive), this very strong preference for the liking accomplice was maintained again for both measures. However, the effects of similarity-dissimilarity in the first experiment did not replicate. Once again, the meter-estimation measure and the social attraction cluster ratings yielded different results. This time, however, the subjects gave the similar accomplice much higher social attraction ratings than the dissimilar accomplice—regardless of her liking or disliking the subject. The meter-estimation data showed no effects of similarity whatsoever. In spite of this failure to replicate, in both experiments the meter-estimation data were more clearly in the direction of the hypothesis than were the conventional questionnaire data. As Jones et al. (1971) argued in their discussion of the entire pattern of results, the meter-estimation measure performs more like a measure of how the subject really feels about the accomplice, whereas the questionnaire data seem to be more a reflection of how the subject thinks she *should* feel in such a situation.

Such a conclusion is consistent with, but hardly forced by, the data of their two experiments. At a more conservative and more descriptive level, there is evidence that the meter-estimation measure and the attraction ratings are differentially responsive to experimental conditions, but have some proportion of variance in common. In the first experiment, for example, the within-cell correlations between the meter estimations and the attrac-

tion ratings averaged .66. In the second experiment, the average intercorrelation was .81. Thus it is clear that attractiveness ratings are moderately congruent with estimated affective reaction, though these correlations may indeed be inflated by the aforementioned confounding of measure and sequence of administration. On the other hand, the meter-estimation averages fall into a different pattern than the attraction ratings, as we have noted, and this suggests that the meter-estimation technique, while more cumbersome and deceptive than the questionnaire method, is tapping a different aspect (a truer aspect?) of attraction. This suggestion is strongly supported by the data from an experiment by Cooper (1971), to which we now turn.

Attraction as a Product of Dissonance Reduction

Cooper was interested in testing the theoretical hypothesis that for postdecisional dissonance to occur, any unfavorable consequences must have been foreseeable at the time of the decision. Cooper chose to test this hypothesis using female subjects in an alleged study of the effects of cooperation and competition on interpersonal attraction. He designed a two-person cooperative game in which each partner was to take turns attempting to answer problems from an aptitude test. Each player was also to announce whether she thought she was probably correct or incorrect in her answer to each problem. An incentive system was explained that provided maximum credit for answers that were correct and claimed to be correct by the respondent. The system, in effect, penalized the team if either partner was overly timid (failing to recognize her correct answers) or excessively confident (claiming incorrect answers to be correct). Some subjects were then led to expect a timid partner, some an overconfident one. Experimental subjects were given the option of exchanging the assigned partner for another randomly chosen one (none did so); control subjects were given no such choice. In the play of the game, subjects learned, through controlled feedback, that their partner was in fact very timid, moderately timid, moderately overconfident, or very overconfident. After playing the

game, each subject was provided the main dependent variable measure of attraction for her partner.

This main measure was again in the form of estimated readings of a GSR machine featuring directional as well as intensity information. Subjects appeared at each experimental session in groups of four to six, having volunteered for an experiment on "cooperative problem solving." It was explained that a new device would be used to measure how each person felt about her partner during the problem-solving game. The device was described as the differentiated GSR, whose properties were essentially the same as those attributed to the machine by Jones et al. (1971). As part of the overall strategy of validating the differentiated GSR in the subjects' eyes, they were ushered into a small adjacent control room and shown an imposing array of electrical and electronic equipment (actually consisting of the laboratory intercom patch board and a bank of preamplifiers). After they had returned to the experimental room, each subject was assigned to a desk, separated from her neighbor by a partition. On each desk was an electrode and earphones. The wire from the electrode ran through a wall conduit into the adjacent equipment room. When the electrodes were taped to the subject's hand, she would presumably be attached to the elaborate differentiated GSR equipment supposedly designed to provide simultaneous feedback measuring the affective responses of each subject.

The experimenter then asked each subject to listen to all further instructions through the earphones in front of her. Under the pretext of making minor adjustments in the differentiated GSR base line for individual subjects, the experimenter asked each subject in turn to consider her feelings about the statement, "Since it is known to be nonaddictive, the use of marijuana should be legalized." This statement had appeared as an item on an opinion questionnaire administered to all subjects earlier in the semester. With the results of each subject's answers to this questionnaire before him, the experimenter announced over the earphones that the first subject's meter reading was so and so, on a scale from zero (meaning extreme negative af-

fect or disagreement) to 300 (which indicated extreme positive affect). He then asked the subject, "Does this sound accurate to you?" In the vast majority of cases, the subject said yes and often expressed amazement at the accuracy of the meter. An occasional subject demurred and suggested that the meter needed a little adjustment. In such cases the experimenter purported to adjust the base line and proceeded to a second opinion item, again announcing the machine's reading and asking the subject if this was accurate. In all cases this was sufficient. The experimenter repeated this procedure with the second, third, fourth, etc., subject. It should be noted that this procedure uses each subject's acquiescent response as evidence to validate the discernment of the machine in the eyes of the others. The effect was very powerful, and the subjects were characteristically quite surprised to learn after the experiment that the machine's properties were fictitious.

After these validation procedures, the experimenter gave subjects some crucial information about their partners to create the requisite expectations of timidity and overconfidence, and the problem-solving game was then played. Each subject thought she was "Subject C," and the experimenter conveyed the same pattern of self- and partner feedback in each session. By indicating whether or not the partner gave the correct answer, and whether or not she certified the answer as probably correct, the experimenter could create the experience of playing with a partner who was either on the timid or the overconfident side.

At the end of 15 problems, subjects were asked to estimate and record "what you feel the differentiated GSR meter should read as you think of your partner." The subject's response at this point was the main dependent variable measure of the experiment. Subsequent to this, subjects were also asked a series of questions about their partners' intelligence, whether they would like to work with them again, etc.

For a detailed presentation of the results, the reader is referred to the original article. It is sufficient for our present purposes to indicate that the theoretical predictions were confirmed in detail with the differ-

entiated GSR measure. In the choice conditions (where the subject could have asked for another partner), (a) when timidity was expected, the extremely timid partner was liked better than a moderately timid partner; and (b) when overconfidence was expected, the extremely overconfident partner was liked better than the moderately overconfident one. In short, when the subject is personally responsible for choosing to play with a partner possessing known liabilities, she likes the partner more when these liabilities turn out to be fairly extreme. This is in spite of the fact, of course, that the extremely timid or the extremely overconfident partner causes the team to lose more money than their moderate counterparts. This pattern of results was reversed when the subject did not have any choice in the matter, or when the subject expected a timid partner and got an overconfident one, or expected an overconfident partner and got a timid one.

These results again provide a *prima facie* case for the utility of the bogus pipeline as a measure of attraction. Just as surely, however, they do not provide the crucial comparative information that such measures are better than a direct attraction question. Since Cooper was primarily interested in the theoretical question, the role of personal responsibility in dissonance reduction, he did not run control subjects whose feelings were assessed by standard rating techniques. The data he obtained from the subsequently administered questionnaire provided several indicators that would normally go along with an impression of attractiveness (e.g., "enjoy working with"). These verbal questionnaire responses did *not* confirm the dissonance prediction. Even in the choice conditions, subjects did not express any particular desire to work further with the extremely timid or overconfident partner, though the differentiated GSR showed their attraction to such a partner when her proclivities were expected. Since the verbal measures followed the meter-estimation measure, however, we cannot be sure that the former were not in some way contaminated by the latter. A more direct test of the difference between the bogus pipeline and standard rating procedures was of major interest in two studies by Sigall and Page, now to be reported.

Bogus Electromyograph and Reactions to Stigma

Sigall and Page⁵ were interested in directly comparing the bogus pipeline with a standard rating procedure. To this end they selected a situation in which subjects might be expected to distort their responses under normal rating conditions. As noted earlier, subjects are reluctant to be negatively evaluative. In addition, Goffman (1963) has nicely documented the awkward and affected nature of interaction with stigmatized others. Sigall and Page confronted female subjects with an obnoxious male stimulus person who was presented as either stigmatized or normal, and assessed subjects' reactions to the stimulus person under bogus pipeline or rating conditions.

The procedure went as follows: Each female subject, and male accomplice posing as a subject, reported to a waiting room in the psychology laboratory. The accomplice either wore massive leg braces or appeared to be physically normal. A posted sign directed subjects not to talk with one another. Moments later a telephone in the waiting room began to ring. When it became apparent that no one would answer it, the confederate picked up the phone and was extremely rude and gratuitously abusive to the caller, who evidently had dialed a wrong number. Shortly thereafter the experimenter appeared, seemingly unaware of the phone call. He ushered the subject and confederate into a small experimental room and asked them to respond to a short opinion questionnaire that consisted of five innocuous items concerning movie and music preferences. The experimenter then left the room, and subjects indicated agreement or disagreement with the items on a 7-point scale. While the subject answered her questionnaire, the confederate acted as if he were doing the same; actually he was copying the subject's responses. When the experimenter returned he "randomly" selected the subject to accompany him into a second experimental room. She was instructed to bring her questionnaire with her.

The subject was seated before a console.

⁵ Sigall, H., and Page, R. Attenuation of the expression of affect. Manuscript in preparation.

A steering wheel extending from the console turned a pointer along a 7-point scale ranging from -3 to 3. The drawn scale was mounted on the console, and there was a slot at each end of the scale in which a label could be placed. A small metal box sat next to the console. This box, labeled "EMG," contained a meter with a 7-point scale identical to the one on the console. The meter itself was marked "EMG Output." Two skin electrodes were connected to the box by cable, and cables also ran from the box to impressive-looking electrical junk.

The experimenter described the apparatus as an adapted electromyograph. It was explained that an electromyograph (EMG) was a device of long standing, used to measure implicit muscle potentials. The experimenter noted that standard questionnaires used in psychological research involved a variety of difficulties and that the newly developed adapted EMG was an important breakthrough that circumvented many of the problems. He explained that when the subject held the steering wheel, and electrodes were attached to her forearms, the EMG would pick up her tendency to turn the wheel to any point on the scale. She was told that she actually did not have to turn the wheel, that the EMG screened out *gross* muscle movements and recorded the "first, undistorted" reaction. The electrical junk was described as a small computer, which integrated the electrophysiological input, and the EMG output meter would reflect a close approximation of that integration. Similarities to a lie detector were noted, and it was pointed out that the EMG was an improvement over the lie detector in that it was sensitive to direction as well as intensity of responses.

The experimenter then told the subject that occasionally base-line responses differed among individuals, in which case an adjustment of the EMG would be required. To check on the subject's base line, they would go over the items on the questionnaire and look at the correspondence between it and the EMG readings. The wheel was locked in place, "agree" and "disagree" labels were inserted in the console, and the electrodes were attached to the subject. She was to listen to

the statements as they were read, to attend to the drawn scale, and to look at the EMG output readings if she wished. The experimenter read the first item and flipped a switch on the computer. The computer's lights flashed, and a mechanical noise was emitted. After a brief period the noise ceased, and the needle on the output meter swung to a number. Although these events appeared automatic to the subject, they were manipulated by the accomplice in the adjacent room. The experimenter continued by presenting the remaining items in similar fashion. On one item the experimenter suggested that the subject could, if she wished, try to fool the EMG either by exercising gross muscle movements or consciously concentrating in the direction opposite to her true position. The subject was reminded that since the EMG recorded implicit muscle movements, a reflection of her first, undistorted reaction, such strategy should not be successful. After all items had been completed, the subject was asked to produce the questionnaire, and the two sets of responses were compared. Of course, the EMG had been accurate in every case.

The experimenter then turned to the "purpose of the present experiment." He said that the study was concerned with social perception and that more information would be disclosed later. At this point he needed to get some information from the subject regarding her current perceptions about herself, the experimental situation, and the other subject. He indicated that while he would obtain this information directly from the EMG, he was also interested in discovering how sensitive people were to their own feelings—"whether they were in touch with how they felt." Thus, he introduced a shield to obstruct the subject's view of the meter, and told her to predict what she thought the meter would read. It was explained that predictive accuracy would be maximized if she relied on her undistorted reactions. He also told her that she could see how well she had predicted after all the items had been administered. The experimenter presented the items, recording the subject's prediction after each one, and pretending to note the EMG reading as well. The crucial items asked for the subject's evaluations of the accomplice. The subject

indicated her personal liking for the confederate, and her agreement or disagreement as to the applicability of various traits, such as "obnoxious," "pleasant," or "pushy" to him.

In the rating conditions the computer and EMG were in view, but turned off. The wheel was free to be turned, and the electrodes were hidden. The subject responded by turning the wheel and orally stating the number to which the pointer had been moved. It was explained that concentration and attention remained higher when a physical task was involved than when a paper-and-pencil questionnaire was used. The remaining apparatus was described as a small computer used for data analysis. As in the EMG conditions, subjects were asked to present first reactions, not to distort responses, and to be honest.

A detailed accounting of results is in preparation (see Footnote 5). Here we simply note some of the findings. All subjects were convinced that the EMG did in fact possess the powers attributed to it. As expected, subjects in the EMG conditions tended to be unfavorable in their evaluations of the confederate, while those in the rating conditions tended to be favorable. This was manifested in relative terms and also reflected by the fact that on a scale which ranged from 3 to -3, mean responses in the rating condition were different in algebraic sign from mean responses in the EMG condition. The EMG subjects said they did not like the stimulus person; rating subjects said they did. He was evaluated as pushy, obnoxious, and insensitive in the EMG condition, but not in the rating condition. In sum, EMG subjects seem to have responded reasonably to the staged obnoxiousness of the accomplice, whereas subjects in the rating conditions appear to have been overly generous.

Sigall and Page also expected a greater discrepancy in evaluation between measurement conditions when the confederate was stigmatized. They felt that most subjects would be hesitant to derogate someone with a stigma and that therefore especially high evaluations would result in the stigma-rating condition. Subjects in this condition did provide the most positive responses, and, although the form of the interaction was

realized, it failed to reach statistical significance. This leaves open the possibility that the EMG simply elicits more negative responses than usual rating procedures. The final experiment reported here provides some information on the viability of such an alternative.

New Look at Stereotypes

One of the findings recently reported by Karlins, Coffman, and Walters (1969) in their study of social stereotypes was that white, American subjects provided less favorable stereotypes of "Americans" and more favorable stereotypes of "Negroes" than was the case in earlier research. Without denying that those data may reflect a real change in attitudes over time, it is quite possible that such findings may be partly influenced by social desirability. Self-criticism by Americans and a tendency to express positive sentiments about Negroes seem to be active phenomena on many campuses. Such a climate easily could contribute to response distortion in a study of stereotypes.

A second experiment by Sigall and Page (1971) attempted to examine possible distortion by using the EMG. A 2×2 factorial design was employed in which half of 60 white male subjects indicated how characteristic each of a series of adjective traits was of "Americans"; the remaining subjects did the same in response to "Negroes." This variable was crosscut by assessment technique: As in the previous study, half the subjects responded by predicting their EMG readings, whereas the other half filled out standard rating forms.

The procedure used to "validate" the EMG was essentially identical to the one used in the experiment on stigma. However, since there was no need to expose the subject to a confederate, one variation was introduced. When the subject arrived, he filled out the small questionnaire. He was then escorted to the experimental room and seated facing the console, with his back to the door of the room. The completed questionnaire was placed on a table near the door, which was left slightly ajar so that an accomplice in the corridor secretly could copy the subject's

answers. The accomplice then went to the control room, able to manipulate the EMG readings during validation. As far as the subject could tell, no one had the opportunity to see his questionnaire.

After the preliminaries (including validation in the EMG conditions), subjects were asked to think about "a group," either "Americans" or "Negroes." The labels "characteristic (3) and "uncharacteristic" (-3) were inserted in the slots on the console, and the experimenter proceeded to read 22 traits, one at a time.

A detailed report is presented in Sigall and Page (1971). Here we summarize some of the more striking results. Some traits were unaffected by measurement technique and yielded only a Negro-American main effect. For example, Americans were described as more practical and more materialistic, while Negroes were described as more musical.

More interesting, especially for methodological considerations, were those traits that resulted in interactions. These traits tended to be highly affect-laden, and the form of the interaction was consistent over traits: "Americans" were described more favorably in the EMG than in the rating condition; "Negroes" were described more favorably in the rating than in the EMG condition. Two examples serve to illustrate. On a 7-point scale that extended from -3 to 3, "honest" was reported as characteristic (.60) of Americans in the EMG condition but as uncharacteristic (-.27) of Americans in the rating condition, while the same trait was viewed as uncharacteristic (-.33) of Negroes in the EMG condition but characteristic (.67) in the rating condition. This interaction was highly significant ($p < .001$). "Lazy" was reported as uncharacteristic (-.80) of Americans in the EMG condition and uncharacteristic (-.60), but less so, in the rating condition. "Lazy" was uncharacteristic (-.73) of Negroes in the rating condition, but characteristic (.60) in the EMG condition.

Regarding the question of whether the EMG simply elicits more negative responses, it should be noted that one of the groups, Americans, received more favorable evaluations in the EMG than in the rating condition.

Discussion and Conclusions

The examples reviewed in this report suggest a few of the various uses to which the bogus pipeline paradigm can be put. We wish to emphasize the rather unrestricted adaptability of the paradigm and to reiterate that particular procedures and rationales employed will vary with the investigator's interests and resources. As we penetrate more and more deeply into the layman's assumptions about physiological processes, it may be important to distinguish between bogus autonomic indicators and those alleged to reflect implicit muscle movements. For the moment, we are inclined to treat these as sufficiently equivalent to lump them together as one general paradigm.

While the above examples are encouraging and suggestive, they do not provide very precise information concerning the particular mensurational advantages of using bogus physiological indicators. This is especially true of the first two studies where there is no direct comparative information on the relation between meter estimates and straightforward verbal measures of attraction. The meter-estimate measure seems to give results that are more in line than the questionnaire measure with theoretical predictions of variations in underlying attraction, but there is no way of knowing whether this is because the meter-estimation measure is more penetrating or because it was administered first and therefore assigned greater importance by the subjects. The two Sigall and Page studies do provide direct comparative information, but it is far from entirely clear just what is being compared.

One of the strongest reasons for preferring a bogus physiological measure to a questionnaire rating is precisely that the rating is potentially vulnerable to so many extraneous and artifactual influences. It seems probable that different wordings and contexts of administration would produce rating results differing from each other and showing greater or lesser differences from meter-estimation measures. We know from the Sigall studies that electromyographic reading estimates differ systematically and understandably from paper-and-pencil ratings made on a scale with

the same range and verbal label. But we do not know if the observed differences would remain or be as strong if, instead of using a standard paper-and-pencil rating as our comparison standard, Sigall had taken his subjects aside and said, "Look I want you to be completely candid and tell me what you really think about that guy," or "Give me a rating that reflects your gut reaction" (cf. Aronson & Linder, 1965).

This discussion moves us toward the more fundamental question about *why* the observed differences between verbal ratings and bogus physiological measures occur. Here we face a number of reasonable alternatives. The first and most straightforward of these is that bogus pipeline measures *operate like lie detection devices in facilitating scrupulously truthful reporting*. To embrace this alternative is to accept the assumption that subjects do not typically tell us the truth when they report their attitudes and feelings about sensitive matters, or those where private feelings might be expected to differ from a socially respected norm. This is not an altogether unreasonable assumption—especially if we remember that all it takes is a few subjects who conceal the extremity of their feelings in the rating condition to produce a reliable difference between condition means.

A variant of this alternative is to posit a *preference for negative over positive errors*. Let us assume that subjects have no univocal or precise response when asked to assign a number to their feelings or attitudes on complex issues. If only they were allowed to do so, they would prefer to express themselves in terms of a range of acceptable positions rather than a particular rating point (cf. Sherif & Hovland, 1961). Part of this indecision may stem from the subject's awareness that his feelings about persons or his attitudes toward issues tend to vacillate over time and to vary somewhat with mood, circumstances, and so on. When asked to estimate a telltale meter, therefore, the subject can in effect tell the truth in many ways. He can tell a truth involving relatively more negative or more positive affect. All that is needed, then, is to make the one further assumption that the person would rather err on the side of being too harsh than too lenient.

To take the stereotype study as an example, it may be better to admit being a bigot and have it shown that you are really fairly liberal than to claim tolerance while being revealed as a bigot. The subject therefore hedges his bets, making it less likely for the latter, less preferred, error to occur. On the whole, this is an interesting alternative that is consistent with most of the data thus far collected. However, it does not account for the fact that subjects were more favorable in describing Americans when the meter was employed—unless one assumes that a favorable rating of Americans is an indication of negative attitudes toward others.

A third alternative, not incompatible with those already mentioned, is that the bogus physiological apparatus makes the *affective component of an attitude or evaluation salient over the more cognitive components*. If you ask a person, "What do you think?" he may respond differently than when you ask him, "How do you feel?" Similarly, to ask the subject to predict what his autonomic nervous system is telling the machine is to focus him on the more primitive, affective components of the impression or attitude structure. The resultant response may therefore differ from the verbal rating that is equally honest, but that reflects different dimensions of the attitude being measured. If this were true, it would cast the stereotype findings in quite a different light than if one were to accept the lie detection emphasis of the first alternative. It is easy to jump to the conclusion that the white subjects in the stereotype experiment are really less tolerant and more "racist" than they profess to be. Although this is certainly conceivable, it is equally conceivable that the subjects are trying to cope with two different tasks and responding honestly to both. Regardless of our own particular racial beliefs, we have all heard much about prejudice and probably have well-developed stereotypes about the actuarial likelihood that people have deep-seated negative feelings about other racial groups of which they themselves are unaware. When asked to estimate the affective component of their racial attitudes, therefore, it is not altogether surprising that they would accede to some extent to the implications of this stereotype concerning the nature of

prejudice. A similar argument could be developed concerning the actuarial likelihood of hidden negative affect in cases where attraction is being measured. People may have a kind of id-ego conception of their attitudes. They may assume that their ugly, negative feelings are buried in the autonomic nervous system and are expressed only in offguard moments; that portion of the attitude that controls behavior, on the other hand, is tied closely to conscious, rational processes in the intellect.

While it is impossible to choose among these alternatives with our present available knowledge, it is important not to embrace too hastily the proposition that the bogus pipeline is necessarily a better or truer measure of attitude than the standard devices and especially that physiological estimates are a better predictor of overt behavior than questionnaire measures of attitude and attraction. To return to the stereotype study, the results do not necessarily imply that someone who estimates a negative reading to the stimulus Negro will therefore be likely to discriminate against Negroes or to react with hostility to individual blacks. Undoubtedly, different attitudes, and attitudes held by different persons, are differently structured and forecast overt behavior with differential precision. It is quite possible that the more cognitive questionnaire response to an item about race reflects the way a person wants to see himself and the way in which he tries to behave. He may have some reservations about how his autonomic nervous system feels about Negroes, but nevertheless feel that he is in all the more important respects an unprejudiced man. It remains to be seen, certainly, whether the bogus pipeline is any more behavioroid than a questionnaire response, though in many contexts it should be freer of misleading extraneous determinants.

The different alternatives discussed above suggest different research programs. The first alternative, that the bogus pipeline provides a truer measure than other methods of how a person honestly feels, leads to experiments on the extraction of feelings in highly sensitive areas or after interpersonal experiences that arouse unpleasant or embarrassing emotions. The final alternative considers the bogus

pipeline as a method tapping a different aspect of attitude. To accept this seriously could lead in the direction of research on the structure of different kinds of attitudes and to an exploration of meaningful discrepancies between cognitive and affective dimensions as a function of issue and setting. It is our belief that research along both of these lines would prove interesting and constructive. By either avenue we shall begin to learn more about the multiple processes of self-attribution and the nebulous phenomenology of emotional experience.

REFERENCES

- ARONSON, E., & CARLSMITH, J. M. Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology*. (2nd ed.) Vol. II. Reading, Mass.: Addison-Wesley, 1969.
- ARONSON, E., & LINDER, D. E. Gain and loss of esteem as determinants of interpersonal attractiveness. *Journal of Experimental Social Psychology*, 1965, 1, 156-172.
- BEHAR, L. B. The effects of cognitive dissonance on inappropriate emotional reactions. *Journal of Personality*, 1967, 35, 505-519.
- BRAMEL, D. A. A dissonance theory approach to defensive projection. *Journal of Abnormal and Social Psychology*, 1962, 64, 121-129.
- BRAMEL, D. Interpersonal attraction, hostility and perception. In J. S. Mills (Ed.), *Experimental social psychology*. New York: Macmillan, 1969.
- BREHM, J. W., & BEHAR, L. B. Sexual arousal, defensiveness, and sex preferences in affiliation. *Journal of Experimental Research in Personality*, 1966, 1, 195-200.
- CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychological Bulletin*, 1950, 47, 15-38.
- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, 56, 81-105.
- COOK, S. W., & SELTZITZ, C. A multiple-indicator approach to attitude measurement. *Psychological Bulletin*, 1964, 62, 36-55.
- COOPER, J. Personal responsibility and dissonance. The role of foreseen consequences. *Journal of Personality and Social Psychology*, 1971, 18, 354-363.
- COOPER, J. B. Emotion in prejudice. *Science*, 1959, 130, 314-318.
- CROWNE, D. P., & MARLOWE, D. *The approval motive*. New York: Wiley, 1964.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- FREEDMAN, J., & FRASER, S. C. Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 1966, 4, 195-202.

- GERARD, H. B. Physiological measurement in social psychological research. In P. H. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior*. Stanford: Stanford University Press, 1964.
- GERARD, H. B. Deviation, conformity, and commitment. In I. D. Steiner & M. Fishbein (Eds.), *Current studies in social psychology*. New York: Holt, Rinehart & Winston, 1965.
- GERARD, H. B., & RABBIE, J. M. Fear and social comparison. *Journal of Abnormal and Social Psychology*, 1961, **62**, 586-592.
- GOFFMAN, E. *Stigma*. Englewood Cliffs, N. J.: Prentice-Hall, 1963.
- HESS, E. H. Attitude and pupil size. *Scientific American*, 1965, **212**, 46-54.
- JONES, E. E., BELL, L., & ARONSON, E. The reciprocation of attraction from similar and dissimilar others. In C. C. McClintock (Ed.), *Experimental social psychology*. New York: Holt, 1971, in press.
- JONES, E. E., & DECHARMS, R. Changes in social perception as a function of the personal relevance of behavior. *Sociometry*, 1957, **20**, 75-85.
- JONES, E. E., STIRES, L. K., SHAVER, K. G., & HARRIS, V. A. Evaluation of an ingratiation by target persons and bystanders. *Journal of Personality*, 1968, **36**, 349-385.
- KARLINS, M., COFFMAN, T. L., & WALTERS, G. On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, 1969, **13**, 1-16.
- LEE, C. D. *The instrumental detection of deception*. Springfield, Ill.: Charles C Thomas, 1953.
- LEIDERMAN, P. H., & SHAPIRO, D. (Eds.) *Psychobiological approaches to social behavior*. Stanford: Stanford University Press, 1964.
- LEWIN, K., LIPPITT, R., & WHITE, R. K. Patterns of aggressive behavior in experimentally created social climates. *Journal of Social Psychology*, 1939, **10**, 271-279.
- MARLOWE, D., FRAGER, R., & NUTTALL, R. L. Commitment to action taken as a consequence of cognitive dissonance. *Journal of Personality and Social Psychology*, 1965, **2**, 864-868.
- MCGUIRE, W. J. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology*. (2nd ed.) Vol. III. Reading, Mass.: Addison-Wesley, 1969.
- MORENO, J. L. *Who shall survive?* Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934.
- ORNE, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, **17**, 776-783.
- PORIER, G. W., & LOTT, A. J. Galvanic skin response and prejudice. *Journal of Personality and Social Psychology*, 1967, **5**, 253-259.
- RANKIN, R. E., & CAMPBELL, D. T. Galvanic skin response to Negro and White experimenters. *Journal of Abnormal and Social Psychology*, 1955, **51**, 30-33.
- RIECKEN, H. W. A program for research on experiments in social psychology. In N. F. Washburne (Ed.), *Decisions, values, and groups*. Vol. 2. New York: Pergamon Press, 1962.
- ROSENBERG, M. J. When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1965, **1**, 28-43.
- SCHACHTER, S. Deviation, rejection, and communication. *Journal of Abnormal and Social Psychology*, 1951, **46**, 190-207.
- SCOTT, W. A. Attitude measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology*. (2nd ed.) Vol. II. Reading, Mass.: Addison-Wesley, 1969.
- SHAPIRO, D., & CRIDER, A. Psychophysiological approaches in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology*. (2nd ed.) Vol. III. Reading, Mass.: Addison-Wesley, 1969.
- SHERIF, M., & HOVLAND, C. I. *Social judgment: Assimilation and contrast effects in communication and attitude change*. New Haven: Yale University Press, 1961.
- SIGALL, H., ARONSON, E., & VAN HOOSE, T. The cooperative subject: Myth or reality? *Journal of Experimental Social Psychology*, 1970, **6**, 1-10.
- SIGALL, H., & PAGE, R. Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, 1971, **18**, 247-255.
- STERNBACH, R. A., GUSTAFSON, L. A., & COLLIER, R. L. Don't trust the lie detector. *Harvard Business Review*, 1962, **40** (6), 127-134.
- VALINS, S. Cognitive effects of false heart-rate feedback. *Journal of Personality and Social Psychology*, 1966, **4**, 400-408.
- WESTIE, F. R., & DEFLEUR, M. L. Autonomic responses and their relationship to race attitudes. *Journal of Abnormal and Social Psychology*, 1959, **58**, 340-347.
- ZIMBARDO, P. G., COHEN, A. R., WEISENBERG, M., DWORKIN, L., & FIRESTONE, I. The control of experimental pain. In P. Zimbardo (Ed.), *The cognitive control of motivation*. Glenview, Ill.: Scott, Foresman, 1969.

(Received April 10, 1970)