

Twenty Years of Bogus Pipeline Research: A Critical Review and Meta-Analysis

Neal J. Roese and David W. Jamieson

The bogus pipeline (BPL), a procedure intended to improve the truthfulness of self-reports, was examined in terms of the validity of its effects, its optimal procedural format, and its appropriate domain of use. Social psychological research that has used the BPL is reviewed and meta-analyzed. Thirty-one studies were coded for effect size and relevant procedural characteristics. A significant mean BPL versus control condition effect was evident across these studies, indicating that the BPL engendered reliable effects consistent with a reduction in socially desirable responding. The BPL produced larger effects when task instructions required Ss to guess the BPL's output. These findings, coupled with previous indirect validation, provide reasonable documentation that the BPL shifts self-reports toward veracity. Past criticisms of the BPL are considered, and recommendations for its future use are made.

For words, like Nature, half reveal
And half conceal the Soul within.

—Alfred, Lord Tennyson, *In Memoriam*

Psychologists have long attempted to measure accurately individuals' attitudes and opinions. Ultimately, a direct pipeline to the soul has been desired, a method that somehow pierces through strategic facades and bypasses the concealing words of which Tennyson wrote. Such a direct pipeline is clearly not possible, but in their pioneering 1971 article, Edward Jones and Harold Sigall described a "bogus" pipeline (BPL) to the soul, a procedure ostensibly providing a far closer approximation of the contents of souls than was previously possible with traditional paper-and-pencil (PP) questionnaires. By convincing subjects that a physiological measurement apparatus was capable of recording their genuine attitudes and opinions, the BPL was thought to invoke a motivation in subjects to offer more veracious self-reports.

Neal J. Roese, Department of Psychology, University of Western Ontario, London, Ontario, Canada; David W. Jamieson, McMaster University, Hamilton, Ontario, Canada.

A preliminary version of this article was presented at the 99th Annual Convention of the American Psychological Association, San Francisco, 1991. Preparation of this article was supported in part by a doctoral fellowship awarded to Neal J. Roese and Research Grant 410-91-0908 to David W. Jamieson, both from the Social Sciences and Humanities Research Council of Canada.

We thank Joseph Roese for his translation of several articles from German into English. We are also grateful to John Adair, Victoria Esses, Karen Grabowski, James Olson, Donald Sharpe, Mark Zanna, and four anonymous reviewers for their insightful comments on drafts of this article.

Correspondence concerning this article should be addressed to Neal J. Roese, now at Department of Psychology, University of California, Santa Barbara, California 93106.

The procedure quickly became an accepted means of reducing social desirability biases in psychological experiments, and despite some questioning of its validity (e.g., Cherry, Byrne, & Mitchell, 1976; Ostrom, 1973), use of the technique spread to disparate fields of inquiry. A flurry of BPL studies were published in the late 1970s and early 1980s, but after this period, theoretical interest waned to the point that today in the early 1990s, the procedure has been largely abandoned as a laboratory technique in social psychology. This state of affairs is particularly intriguing because of the concomitant increase in the use of the BPL in assessments of drug-use behavior. A modification of the original BPL has enjoyed frequent service in improving the veracity of self-reported consumption of alcohol, tobacco, and other potentially health-threatening drugs (e.g., Murray & Perry, 1987). With the expansion of this literature have come methodological and ethical controversies similar to those that concerned social psychologists 15 years ago.

The waning of BPL use by social psychologists may perhaps be attributed to a variety of concerns. For example, researchers may perceive the BPL to be ineffective (i.e., it fails to produce the intended effect of greater veracity of response), to be impractical (i.e., its elaborate procedure is logistically cumbersome), to have questionable theoretical underpinnings (i.e., it is unclear whether veracity is achieved by reduction of social desirability biases or greater mindfulness), or to be unethical (i.e., its degree of deception has been argued by some to be at odds with contemporary ethical standards).

In the present article, we attempt to address these concerns through a critical review and meta-analysis of past BPL research. The first concern, if warranted, would clearly justify abandonment of the procedure. To date, however, a comprehensive appraisal of the BPL's effectiveness on the basis of the results of its myriad applications has not been available. The second concern might well be addressed through procedural revisions, but again, an assessment of the BPL's procedural de-

tails and their relation to intended BPL effects has not been undertaken. Because little information exists pertaining to the latter two of these possibilities, our review and analysis are necessarily pragmatic. However, if the first two concerns can be assuaged, the BPL could enjoy renewed interest among social psychologists, with more effective use by applied researchers. Indeed, increasing interest in such topics as prejudice and symbolic racism (see Olson & Zanna, 1993) could well stimulate renewed interest in techniques designed to reduce conscious dissimulation.

We begin by examining the BPL procedure itself and the basis for inferring its valid operation. We then review research that used the BPL and its variants, focusing on studies that illuminated issues of validity. Experiments comparing the BPL responses with those from a control condition are meta-analyzed to better understand the BPL's effects across studies. Criticisms of the BPL are then considered in the light of the review and meta-analysis, and recommendations for future use are made.

The Bogus Pipeline

Biases abound in social research. As Rosenzweig (1933) noted, psychology differs most basically from the physical sciences in that its subjects are both aware of and potentially reactive to the social situation of the experiment itself. A range of reactivity biases in self-reports have been documented by social psychologists, such as social desirability biases (e.g., Crowne & Marlowe, 1964), experimental demand (e.g., Orne, 1962), careless or mindless responding (e.g., Langer, 1989), positivity biases in interpersonal evaluations (e.g., Jones, Bell, & Aronson, 1972), and acquiescence biases (e.g., Bentler, Jackson, & Messick, 1971). Jones and Sigall (1971) introduced the BPL as a generic means of circumventing most or all of these biases, but most subsequent use has been directed at eliminating conscious strategies of overly positive self-presentation. The BPL was designed to convince subjects that a sophisticated new electronic device could accurately detect their "true" attitudes and opinions; it was then presumed that subjects would be motivated to respond truthfully to such questions, to avoid being revealed as a liar or as out of touch with themselves. In the process, the BPL was expected to inspire subjects to respond more conscientiously. The procedure as originally conceived comprised three elements.

First, subjects were shown an impressive physiological-monitoring device, which was purported to measure both the direction and the extremity of their attitudes toward some issue. In actuality, the machine detected nothing but was nevertheless convincing in its physical presence and apparent operation. Note that applications have varied in terms of the demand attached to the BPL's purported purpose. In some studies, subjects were directly told that the BPL was capable of lie detection, whereas other procedures described its purpose in more general terms, such as enhancing response accuracy.

Second, subjects were connected to the apparatus, and several rigged demonstrations convinced them of the machine's accuracy, constituting a "verisimilitude" phase. This was done to ensure that subjects believed the somewhat outlandish claims of the BPL's accuracy. Typically, the apparatus was at-

tached by means of electrodes to subjects' arms and validated in subjects' eyes by "calibrating" it, either by instructing subjects to answer as truthfully as possible or to give deliberately false answers to questions posed by the experimenter. Because subjects' responses to these questions had been surreptitiously copied during an earlier phase of the study, usually involving the completion of a supposedly unrelated attitude questionnaire, the machine could be easily manipulated by the experimenter to confirm subjects' previously recorded answers and to "detect" their false answers. Again, procedures have varied: Whereas many studies used this verisimilitude phase, other researchers thought it unnecessary, instead merely giving subjects the expectation of later objective verification.

Third, while subjects were still connected to the apparatus, their attitudes were assessed when they guessed the machine's readings in response to a series of Likert-format questions. It was assumed that subjects would respond truthfully so as to appear to be in tune with themselves. Hence, the BPL was predicated on the motivational assumption that a desire to avoid appearing to be a liar or to be self-unaware would supersede the typically assumed tendency to exaggerate possession of favorable traits (in other words, that self-protection would supplant self-enhancement). In some studies, however, the task instructions were not to predict the machine's output, but merely to respond as accurately as possible. The evaluation of whether the BPL is effective in reducing social desirability biases in social research and the specification of the minimal procedural characteristics necessary to ensure this effect are the foci of the remainder of this article.

Inferring Validity of the Bogus Pipeline

Throughout this review, the question of the BPL's validity is frequently referenced. By *validity*, we refer to whether the BPL motivates and achieves truth telling in the manner intended by its proponents. Contrary to what has been assumed in previous reports, this is by no means simple to demonstrate. Two criteria are necessary for inferring BPL validity.

The first criterion is that the BPL engender consistently measurable effects (i.e., a reliable difference in responding between the BPL and a control condition) that extend across a range of research settings thought to be tainted by reactivity biases.

The second criterion is that these effects be in the appropriate direction. By *appropriate direction*, we mean whether the BPL produces responses closer to, rather than further from, veracity. When the BPL assesses self-reports of factual matters that can, in principle, be verified (e.g., specific past behavior), the determination of direction is a simple matter of collecting objective, independent evidence about that behavior and comparing it with BPL responses. However, Jones and Sigall (1971) intended the BPL to be a proxy pipeline to the soul; that is, to improve the accuracy of reports of attitudes, opinion, and other intrapsychic variables (henceforth referred to simply as *opinions*). Because such hidden variables can never, in principle, be verified objectively, the validity of the BPL in uncovering dissimulation in reports of opinion can only be inferred indirectly. For the most part, the valid direction of the BPL effect has been assumed to be a shift toward more negative responses on such measures, because social desirability biases are usually

presumed to engender more positive self-reports (but see Kunda & Schwartz, 1983; Sigall & Page, 1972, for examples of theoretically predicted BPL shifts in the positive direction).

The interpretation of more generally negative responses under BPL than PP conditions as evidence for the second criterion is not unequivocal, however. An alternative explanation that might account for such a difference, quite apart from the BPL's putative efficacy in eliminating social desirability biases, is that the procedure may introduce its own unique, systematic bias (e.g., a demand for counternormative self-disclosure). Such a bias might produce responses that though more negative, are at best no closer to veracity than PP responses and at worst are even more biased than PP responses. Hence, an important corollary of the second criterion for inferring BPL validity is eliminating the possibility that the BPL introduces its own systematic biases.

Two conditions must therefore be met to infer BPL validity, assuming that reactivity biases are operative in the research setting: a significant BPL effect across many studies and evidence that this effect occurs in the appropriate direction. In the next section, we provide a qualitative review of the last 20 years of research using the BPL procedure, focusing in particular on efforts to demonstrate the valid operation of the BPL in assessing self-reports of opinion.

Review

The history of BPL research can be divided into three periods in time, or "waves," each distinct in the zeitgeist permeating its developments and controversies. In particular, these waves differ in their focus on whether social desirability was seen as a nuisance variable (e.g., obscuring attitudinal reports) or as a theoretical mechanism of interest in its own right (e.g., impression management theory). These three waves also differ in their focus on whether self-reports assessed matters of opinion or matters of fact.

The First Wave: 1970–1974

This first period of research was characterized by the initial use and assessment of the BPL, with applications restricted to examinations of racism and interpersonal attraction. Research in this wave focused exclusively on questions of opinion, such as racism and attraction, that have traditionally been assumed to be tainted by self-presentational biases. Three early studies (Cooper, 1971; Jones, et al., 1972; Jones & Gordon, 1972) made use of BPL procedures similar to those described by Jones and Sigall (1971) but did not include a control group in their designs, to allow an assessment of the BPL's impact. Sigall and Page (1971) published the first comparison between BPL and control ratings using a between-subjects design. They reexamined evidence indicating that negative stereotypes toward Blacks were becoming more favorable across successive generations of college students (Karlins, Coffman, & Walters, 1969). Given that overt prejudice of any sort was becoming increasingly unfashionable in the late 1960s and early 1970s, Sigall and Page (1971) reasoned that social desirability biases could have tainted the Karlins et al. data. As predicted, they found that White, male college students were more likely to attribute nega-

tive traits to Black Americans when assessed by BPL than by PP procedures. Similar findings were reported by Pavlos (1972, 1973). The tendency to use the BPL to detect negative self-disclosure continued in research on interpersonal attraction. For example, Sigall and Page (1972) reported that subjects gave more appropriately negative ratings of an obnoxious stimulus person under BPL conditions than under PP conditions (see also Jones & Wein, 1972).

It seemed clear to many at this point that the BPL had the power to unmask significant and interesting social phenomena previously obscured by subjects' self-presentational concerns. Nevertheless, Ostrom (1973) launched a particularly lucid attack on the BPL at this time, five main criticisms of which may be summarized. First, empirical evidence available at that time did not suggest that the BPL procedure was any more sensitive to experimental manipulations than were PP ratings. Second, the BPL did not appear to reveal substantially different functional relations between variables. Third, systematic differences in the mechanics of assessment between the BPL and PP ratings could confound any inferred methodological superiority (e.g., the BPL engendered a prominent lack of anonymity and a more time-consuming experimental session). Fourth, the procedure was logistically cumbersome, necessitating individual testing. Fifth, the procedure was vulnerable to ethical criticism.¹

In rebutting Ostrom's (1973) criticisms, Jones and Sigall (1973) argued that existing evidence indicated that the BPL engendered more veracious self-reports, was sensitive, and did not necessarily need to show different functional relations with independent variables. Acknowledging its labor-intensive nature, however, they stressed the need for selective application to domains in which reactivity biases were problematic and where important theoretical issues were at stake. Finally, Jones and Sigall noted that very few subjects were "even mildly upset by the bogus pipeline procedure" (p. 261; see also Gerdes, 1979).

Not long after this debate appeared in the pages of *Psychological Bulletin*, Brigham, Bloom, Gunn, and Torok (1974) further challenged the BPL's efficacy. In attitudes toward race and other issues, these researchers found few significant differences between BPL and PP ratings. One possible explanation for these findings, that the BPL was simply not credible to subjects, was refuted by postexperimental- questionnaire data. By and large, subjects did believe the BPL cover story and were convinced of the device's accuracy. Despite some evidence consistent with

¹ Although it is beyond the scope of the present article to consider ethical issues in detail, the BPL is generally criticized on two grounds. First, its multifaceted deceptions are seen by some as inappropriate and at odds with current ethical standards (see Baumrind, 1985; Christensen, 1988; Sharpe, Adair, & Roese, 1992, for relevant discussions of the ethics of deception). Second, the BPL may restrict the freedom of subjects' decisions regarding self-disclosure. Under PP conditions, subjects encountering a disconcerting question may skip over it or distort their response to protect their private or public self-image. In contrast, the BPL, in forcing subjects to respond truthfully, eliminates this freedom to refrain from honest self-disclosure. The only empirical evidence available, however, suggests that subjects perceive the BPL's brand of deception to be no worse than other experimental uses of deception (Gerdes, 1979).

BPL validity, namely that null BPL effects were somewhat more likely for expression of mundane attitudes (e.g., toward pay television) than for more intimate beliefs (e.g., sexual fantasies), this failure to replicate the findings of Sigall and Page (1971) questioned the practical utility of the BPL.

The first wave thus drew to a close mired in controversy, the paucity of methodological assessments fueling the charges of Ostrom (1973) and Brigham et al. (1974) that the BPL was of little practical significance. Both sides of the debate agreed, not surprisingly, that further empirical investigation was necessary before a resolution could be reached.

The Second Wave: 1975–1981

This period, extending through the late 1970s, was characterized by an expansion of the BPL into more varied areas of inquiry and by continuing controversy over its effectiveness. For the first time, the BPL was used to improve the accuracy of reports of facts as well as of opinions. In addition, impression management became an area of theoretical inquiry in its own right, the BPL playing an important role (as a manipulation of impression management effects) in its early development.

The finding that White students were more likely to admit racist attitudes under BPL than PP conditions was replicated (Allen, 1975; Mummendey, Bolten, & Isermann-Gerke, 1982) and extended (Carver, Glass, & Katz, 1978; Schlenker, Bonoma, Hutchinson, & Burns, 1976). Similar results were found in studies of sexism. For example, men but not women expressed less sympathy toward women's rights under BPL than PP conditions (Faranda, Kaminski, & Giza, 1979, cited in Myers, 1990), and a tendency for women to overstate their support for feminism was reduced by the BPL (Hough & Allen, 1975; see also Bowman & Auerbach, 1978; Ward, 1978). On self-reports of interpersonal attraction, Page and Moss (1975) found that subjects were more likely to negatively evaluate a dissimilar stranger when liking was assessed under BPL than under PP conditions. Also, subjects who received a disproportionately large monetary reward in relation to other subjects claimed to feel guilt and displeasure on PP ratings but admitted to secret pleasure when tested by the BPL (Rivera & Tedeschi, 1976). Thus, across several domains of opinion assessment, numerous studies showed the BPL to be effective in reducing biased responding.

It was during the second wave that the BPL became enmeshed in a heated controversy over interpretations of induced compliance effects: People induced to perform a behavior that they normally avoided (e.g., writing a counterattitudinal essay) would later moderate their attitudes in a direction consistent with this behavior (see Wicklund & Brehm, 1976). Cognitive dissonance theorists viewed such attitude change as real, ascribing it to a motivation to reduce the dissonance between thought and action. Impression management theorists (e.g., Tedeschi & Rosenfeld, 1981), on the other hand, interpreted such attitude shifts as feigned: Subjects were merely pretending to alter their attitudes to project favorable impressions of themselves as consistent social beings. Using the BPL to manipulate the opportunity to engage in strategic impression management, several studies showed that induced compliance effects were mitigated in the BPL condition (e.g., Gaes, Kalle, & Tedeschi, 1978;

Guild, Strickland, & Barefoot, 1976; Malkis, Kalle, & Tedeschi, 1982; Millham & Kellogg, 1980; but see Riordan & Tedeschi, 1981, for an exception).

Because the BPL was so crucial a methodology to impression management theorists, its operation was scrutinized closely by dissonance theorists. For example, Cialdini, Petty, and Cacioppo (1981) argued that the absence of induced compliance effects in the BPL condition was due to subjects misattributing their dissonance arousal to the foreboding BPL apparatus. If subjects were given time to become habituated, then the predicted attitude shift would occur (Stults, Messe, & Kerr, 1984; but see Paulhus, 1982, for conflicting evidence). Scheier and Carver (1980) suggested that any attenuation of attitude change under BPL conditions may have occurred because the very nature of the BPL redirected subjects' attention to their initial attitudes. With these attitudes clearly in mind, it was much more difficult to shift them than to alter interpretations of the counterattitudinal behavior.

Subsequent evidence supported the combined operation of both dissonance and impression management mechanisms in producing the induced compliance effect (see Baumeister & Tice, 1984; Paulhus, 1982; Tetlock & Manstead, 1985), providing something of a resolution. Nevertheless, as ardently as Tedeschi and his colleagues espoused their impression management theory throughout this controversy, so too did they voice their support for the utility of the BPL. Most of their studies reviewed the relevant criticisms leveled at the BPL and countered them with ingenious controls (e.g., Quigley-Fernandez & Tedeschi, 1978). Dissonance theorists chose to refute the impression management theorists' claims not by attacking the BPL's basic effectiveness, but by questioning the mechanisms underlying its effects and elaborating theoretical minutiae. Although this debate for the first time aroused an awareness of the complexity of interpreting BPL effects, many researchers had accepted the BPL as wholly valid and no longer gave much credence to the criticisms of Ostrom (1973).

Doubts persisted, however. Several researchers noted that subjects' self-presentational motives might be compounded by the BPL in a more complex manner than many had thus far realized (Arkin, 1981; Lerner, Peachey, & Meindl, 1979; Schlenker et al., 1976). Arkin argued that the BPL placed two self-presentational demands on subjects: the demand to give responses that were honest, but also the demand to reveal something undesirable, because the very presence of a lie-detector-like apparatus implied that the researchers expected and would respond approvingly to unfavorable self-disclosure. Subjects might then consciously or unwittingly generate more negative responses, resulting in a self-presentational bias that rather than resulting in overly positive self-reports, led instead to overly negative self-reports. Research from the first two waves, as Arkin (1981) pointed out, tended to find that "subjects responding via the bogus pipeline do present themselves in a highly negative light (e.g., as prejudiced, unfair, etc.)" (p. 322). Although the possibility of a counternormative BPL bias was a serious threat to inferences of BPL validity, this issue was not examined in subsequent research.²

² A minor controversy centered on Cherry, Byrne, and Mitchell's (1976) claim that the BPL was ill suited to attraction research because

Despite such doubts, the overall tendency toward significant and theoretically predicted BPL effects enhanced the procedure's perceived validity during the second wave (see Brackwede, 1980, and Mummendey, 1981, for brief reviews of this period). Additional converging evidence for the validity of the BPL as applied to opinion measurement was provided by comparing scores on social desirability scales completed under BPL and PP conditions (Howard, Millham, Slaten, & O'Donnell, 1981; Millham & Kellogg, 1980; Mummendey & Bolten, 1981). Endorsement of socially desirable statements was repeatedly found to be reduced by the BPL.

Quigley-Fernandez and Tedeschi (1978), however, interpreted studies such as these as indirect validation for the procedure. Their studies, on the other hand, were intended to be direct tests of BPL validity, in that subjects' self-reports were compared with an objective standard. Before subjects entered the laboratory, they were informed of the experiment's nature by a confederate. After completing an irrelevant task, subjects were asked whether they had received any prior information about the task. Frequency of confessions was significantly higher under BPL than under PP conditions. Because all subjects had been informed by a confederate, the inference that the BPL not only differed significantly from PP ratings, but that it differed in the right direction (i.e., toward and not away from veracity) could be made. Although an impressive demonstration of the BPL's effectiveness in detecting dissimulation of facts, such research may not necessarily confirm the validity of applications of the BPL to reports of opinion. Nevertheless, by the close of the second wave, these direct and indirect demonstrations had, on a very general level, supported the validity of the BPL in detecting dissimulation of both fact and opinion.

The Third Wave: 1982–1991

During this wave, the BPL continued its expansion into disparate theoretical realms, including examinations of self-serving attributional biases (Arkin, Appleman, & Burger, 1980; Riess, Rosenfeld, Melburg, & Tedeschi, 1981; Rosenfeld, 1990), intrinsic versus extrinsic moral motivation (Kunda & Schwartz, 1983), issue involvement as a determinant of attitude change (Schul & Knapp, 1984), personality scale completion (McGovern & Nevid, 1986), attributional accounts of excuses (Mehlman & Snyder, 1985; Mikulincer & Marshand, 1991), and the role of sympathy in prosocial behavior (Eisenberg et al., 1989). The majority (73%) of those that included control conditions reported significant BPL effects and theoretically meaningful interactions of the BPL with other manipulated variables.

of its susceptibility to such negative demand biases. They found that subjects scoring high on a social desirability scale shifted responses according to explicit experimenter demand when tested under BPL but not PP conditions. Low-social-desirability subjects demonstrated no such shift. Gaes, Quigley-Fernandez, and Tedeschi (1978), however, criticized the Cherry, Byrne, and Mitchell analysis on statistical grounds (see also Byrne & Cherry, 1978), and a later replication of the Cherry, Byrne, and Mitchell study failed to discern any heightened sensitivity to demand in the BPL condition, regardless of individual differences in social desirability (Arkin & Lake, 1983).

Permeating this wave was the assumption that the BPL was a completely valid methodological procedure, applicable both to examinations of psychological mechanisms in a form relatively undistorted by social desirability biases and to manipulations of impression management effects. Thus, few studies attempted to further probe the validity and functioning of the BPL as applied to opinions, despite the persistence of several important concerns highlighted in previous sections of this article. One exception was a direct validation study conducted by Jamieson and Zanna (1991), in which subjects were shown an inert polygraph device and were informed that their responses might later be subjected to lie-detector verification. This lie-detector-expectation procedure, a simplified variant of the BPL (see also Arkin & Lake, 1983; Riess, Kalle, & Tedeschi, 1981), eliminated attitude shifts known to be caused by social desirability concerns. Specifically, the BPL erased the moderation of attitudinal self-reports that typically accompanied an expectation of future counterattitudinal, persuasive communication (e.g., Cialdini, Levy, Herman, & Evenbeck, 1973).

The most notable aspect of the third wave was its striking shift in focus from research assessing reports of opinion to research assessing reports of fact. By the latter half of the 1980s, use of the BPL procedure by social psychologists for theoretical laboratory research had noticeably declined. In the period from 1986 through 1990, only 8 articles described the use of a BPL to examine theoretical issues of interest to social psychologists. In contrast, the period from 1986 to 1990 saw the publication of 20 articles detailing the application of a BPL to behavioral assessment. Specifically, the possibility of improving the accuracy of self-reports of drug use through a simplified BPL variant was explored, resulting in a research literature that had matured through the 1980s largely independent of social psychological research.

The BPL variant described in this literature differs substantially from the method used by social psychologists. First, subjects are typically tested for drug use in large groups rather than on an individual basis. Of course, a principal justification for BPL use in this setting is its ability to test samples that are too large for genuine biochemical verification to be economically feasible. Second, often only the expectation of later verification of self-report is manipulated. Third, independent factual evidence against which self-report information may be compared is potentially available. Moreover, this independent evidence can be provided by the very biochemical-verification techniques described to subjects. The term *bogus pipeline* is therefore something of a misnomer for the *genuine* procedure used in this literature.

The procedure used by Murray and Perry (1987) was typical of this methodology. Subjects were junior high school students tested in their schools for prevalence of smoking. They received verbal instructions and brief demonstrations of a biochemical-verification technique that was based on lung carbon monoxide levels. Sometime during the completion of PP self-reports of drug-use, subjects individually exhaled into a balloon attached to a genuine device that measured carbon monoxide levels. Subjects were not informed, however, of the outcome of this independent test. Disclosure of tobacco and marijuana use was higher when the BPL procedure was used. Similar findings were reported by other researchers (e.g., Bauman & Dent, 1982;

Evans, Hansen, & Mittlemark, 1977; Hill, Dill, & Davenport, 1988; Murray, O'Connell, Schmid, & Perry, 1987). For example, Lowe, Windsor, Adams, Morris, and Reese (1986) found that nearly twice as many pregnant women admitted consuming alcohol when they were informed that their self-reports would be verified through blood and urine analyses. Several studies used the available independent biochemical data to confirm that the BPL effect represented a shift of self-reports in the correct direction (e.g., Bauman & Dent, 1982; Cohen et al., 1988; Murray et al., 1987; Murray & Perry, 1987), thus further validating applications of this BPL variant to reports of fact.

Although it is beyond the scope of this article to consider this large literature in any detail, several failures to replicate the BPL effect (e.g., Campanelli, Dielman, & Shope, 1987; Martin & Newman, 1988) have implications for the present discussion of the BPL's validity and optimal functioning. Murray et al. (1987), in an attempt to account for the null BPL findings, noted the importance of ensuring sufficient impact of the BPL procedure. In many cases, simple verbal explanations were all that constituted a purported BPL manipulation. Jones and Sigall (1971) stressed that convincing subjects of the existence of independent physiological assessment, by making the apparatus impressive and credible, was crucial to the BPL's effectiveness. This would seem especially important when the form of physiological assessment was relatively foreign to subjects (as would be more true for biochemical than for lie-detector verification).

Summary

Social psychological reports that explicitly compared the BPL to a control condition on self-reports of opinion tended, overall, to find significant differences in responding between these two groups. This tendency was, however, not strong. Of the 38 published reports identified, only 65% reported a significant BPL effect, and 43% reported theoretically meaningful interactions between the BPL manipulation and another independent variable. Assertions that the BPL has no measurable effect (e.g., Brigham et al., 1974; Ostrom, 1973) cannot be dismissed lightly given this inconsistent record, though conclusions drawn from a qualitative summary such as this are necessarily imprecise.

Meta-Analysis

Overview

The inconsistent reports of significance regarding the BPL's effects prompted an examination of the BPL procedure by meta-analysis. The principal hypothesis to be tested was whether the BPL produced measurable effects (i.e., our first criterion for inferring BPL validity). Such a hypothesis would be supported by the finding that the mean BPL effect size differed significantly from zero. The interpretation of this effect size test was predicated on the assumption that the conceptual thread tying together the various BPL research areas was the operation of social desirability effects. Thus, to the extent that most research using the BPL sought to answer questions that implicated self-enhancing motivation, a meta-analysis of effect sizes could tell whether, across these studies, the BPL made a significant impact. Even if the proposition of a null BPL effect was rejected, however, the effect size test alone

could not comment further on why or under what conditions the BPL shifted responses on reports of opinion.³

Secondary analyses attempted to shed light on the BPL's mechanism and optimal procedural format. Specifically, possible predictors of effect size were identified a priori, then assessed for their predictive power by fitting both categorical and continuous models to the study effect sizes (Hedges & Olkin, 1985). The hypothesis that the BPL introduce its own unique bias was tested by coding for the demand associated with the BPL (i.e., whether the BPL was described as a lie detector). The effect of the type of instructions given to subjects (i.e., whether subjects were instructed to predict apparatus output or to respond accurately) could illuminate the issue of the BPL's mechanism. Other procedural characteristics—such as assessment format, use of a verisimilitude phase, and type of control group—were also coded. Identification of procedural characteristics that significantly predicted effect magnitude might suggest an optimal format that would ensure or enhance the intended effect of the BPL. Finally, the generality of a possible BPL effect was tested by coding year of publication, sex of subjects, and research area.

Sample

In keeping with the scope of this article, the meta-analysis examined only those research reports from the social psychology literature that used the BPL to assess reports of opinion. BPL studies were identified through literature searches in the *Psychological Abstracts*, PsycLIT CD-ROM, *Science Citation Index*, and *Social Sciences Citation Index*. We also searched the reference lists of all articles found. Studies were included in the sample if the BPL procedure was compared with a control condition using a between-subjects design and if sufficient statistical information were available for effect size calculation.⁴ Only published reports were included in the sample.⁵ Also, studies using

³ The goal of the typical meta-analysis is to determine whether a link between two theoretical constructs is supported across studies. Often, a series of experiments assessing a causal relation between an independent and dependent variable is quantitatively summarized. Although operationalizations may vary, all are assumed to approximate the constructs in question. In the present meta-analysis, only the independent variable construct (i.e., the BPL manipulation) was constant, with dependent measures embracing a variety of unrelated constructs. Linking these constructs together has been the threat of social desirability biases: To the extent that the BPL is viewed as an independent variable and social desirability reduction as a dependent variable, the present meta-analysis may be informative.

⁴ The few within-subjects tests of the BPL effect were not included in the meta-analysis, because the interpretation of them would have been clouded by the possibility of contamination by order effects. For example, all subjects in the Jones, Bell, and Aronson (1972) study made their responses first under BPL then under PP conditions.

⁵ The criticism of publication bias could be leveled at the present analysis. That only published reports were included could result in an overestimation of mean effect magnitude, given that published reports may tend to be accepted because of significant findings, whereas non-significant findings languish in "file drawers" (Bangert-Drowns, 1986; Rosenthal, 1979). We expected that a publication bias might be relatively less severe in the present meta-analysis because the focus of our analysis (i.e., BPL effects) was not the theoretical focus of most published reports. Rather, the BPL was typically manipulated in tandem with a more central theoretical variable, which formed the conceptual basis of the article. A publication bias might therefore apply more often to the central independent variable of each study rather than to the BPL manipulation, leaving our particular sample somewhat less vulnerable to publication bias.

substantially altered BPL variants (e.g., the voice analysis in Millham & Kellogg, 1980) were not included. Of the 38 studies from the social psychology literature identified by the initial literature search, 31 met these criteria and were included in the meta-analysis (these studies are tabulated in Table 1).⁶

Several articles reported multiple experiments and multiple dependent measures. To avoid biasing the results by overrepresenting these reports, we computed single effect sizes for each study and then a weighted mean for the report. Furthermore, when multiple dependent measures were reported within individual studies, all (with the exception of manipulation checks) were taken into account in calculating a mean study effect size. Several studies reported null effects of the BPL but included no statistical descriptions of these effects. In these cases, the effect size was coded as zero and was included in the meta-analysis, thus ensuring a more conservative test of the hypotheses.⁷

Computation of Effect Size

Effects sizes (g) were estimated as the difference between the mean rating by the BPL group and the control group, divided by the pooled standard deviation. For this analysis, calculation of g was based on F and t statistics for 56% of the studies, and on means and standard deviations or error terms for the other 44% of the studies. In many cases, the BPL manipulation was crossed factorially with another independent variable. If the interaction effect was not significant, levels of these other variables were collapsed together, and the experiment was summarized entirely in terms of the BPL factor. Where the interaction effect was significant, however, simply collapsing across the other factor could obscure a powerful BPL effect. For example, stereotype ratings made under BPL and PP conditions were crossed with rating target (Black vs. White Americans) in Sigall and Page (1971), and summary favorability scores evidenced a strong interaction. Relative to the control condition, ratings made in the BPL condition were more favorable for Whites yet less favorable for Blacks. The main effect magnitude for the BPL was small, yet a clear impact had been made by the BPL manipulation. In these cases, simple effects for the BPL versus PP within levels of the other independent variable were calculated then averaged. If no standard deviations were reported, effect size calculations were based on the mean square error of an analysis of variance either taken from the BPL factor or reconstituted from available sum of square information (Glass, McGaw, & Smith, 1981).

Variables Coded

The following variables were extracted from each study: (a) year of publication; (b) sex of sample (female, male, or both); (c) research area (attitudes/attributions, interpersonal attraction, prejudice [including racism and sexism], or other); (d) demand (subjects explicitly told that apparatus detected lies or subjects not told that apparatus detected lies); (e) task instruction (subjects were instructed to predict apparatus output or subjects were instructed to respond accurately); (f) assessment format (whether subjects' BPL responses were made in written form, on a mechanical device, orally, or as a combination of written and mechanical); (g) verisimilitude phase (whether BPL apparatus was demonstrated to subjects or not demonstrated to subjects); (h) control condition (PP only, BPL apparatus present but purpose not explained, BPL apparatus present but explained as unreliable). These variables were coded independently by both Neal J. Roesch and David W. Jamieson, with disagreements resolved by discussion.

Results and Discussion

Bogus Pipeline Effect

An unbiased estimator of effect size (d) was computed for each study by multiplying the raw effect size (g) by a correction for small-sample bias (Hedges & Olkin, 1985). A mean effect

size was then computed, with each study's effect size weighted by the reciprocal of its variance. The mean BPL effect size ($d_r = .41$) differed significantly from 0, $Z = 10.06$, $p < .001$ (95% confidence interval: .33 to .49). Rosenthal's (1979) method (with p set at .05) established that 35,910 null BPL studies languishing in file drawers would be necessary to invalidate this significant BPL effect. Taken together, BPL studies indicated that the procedure did in fact engender substantially different responses than control procedures. Criticisms that the BPL had no measurable effect (e.g., Brigham et al., 1974; Ostrom, 1973) could be safely discarded; this analysis clearly demonstrated that the assertion of a null BPL effect was untenable.

On the basis of Cohen's (1977) specification, the BPL effect size appeared to be somewhere between small and moderate. Recall, however, that this analysis was designed to be conservative. Ambiguity found in published reports was resolved in favor of smaller effect-size estimates, which reduced the chances of error artificially inflating the effect-size estimate. Therefore, the BPL effect was reliable and moderate in magnitude.

Predictors of Effect Size

An overall fit statistic calculated by Hedges and Olkin's (1985) procedure indicated that the hypothesis of homogeneity was rejected, $Q_T(30) = 66.69$, $p < .01$. Categorical models were then fitted to effect sizes using study characteristics. All within-class effect sizes, confidence intervals, and tests of homogeneity are displayed in Table 2. The type of task instruction used was found to significantly predict effect size, $Q_B(1) = 6.09$, $p < .05$.⁸ Instructing subjects to guess the BPL apparatus's output resulted in a larger BPL effect than merely instructing subjects to respond accurately. The mean effect size for the latter, however, still deviated significantly from zero, indicating that the BPL was still effective, albeit less so, when task instructions were less involving. This finding suggested that the BPL was more effective when subjects actively attempted to predict the device's output, perhaps because the motivation to avoid appearing self-unaware was primed. Homogeneity was indicated only within the respond-accurately group. Outliers were re-

⁶ Studies were included in the meta-analysis regardless of their methodological quality. This represents a potentially serious form of bias in meta-analysis, in that poorly controlled or conceptually flawed experiments can obscure or distort the conclusions drawn (see Bangert-Drowns, 1986; Glass, McGaw, & Smith, 1981 for discussions). In the absence of an objective means of discriminating quality among studies and given the small sample size to begin with, we elected to include all studies in the analysis; the strength of conclusions drawn from the meta-analysis were therefore limited by this fact.

⁷ An assumption made by BPL researchers is that their experimental domain is affected by self-presentational biases. But if such biases were in fact minimal, then the BPL, even if valid, would have produced a null effect. Thus, some of the small effect sizes in our sample may not reflect BPL invalidity, but valid BPL operation in domains free of these biases. Because these studies may nevertheless be represented in our sample (thereby shrinking the mean-effect-size estimate), the mean BPL effect obtained was an even more conservative estimate.

⁸ The small within-class sample size for the respond-accurately group limited the strength with which conclusions regarding homogeneity could be made.

Table 1
Summary of Meta-Analysis Study Characteristics

Reference	N	Sex	RA	Procedural details ^a	ME	IE	ES
First wave							
Sigall & Page (1971)	60	M	P	1/1/3/1/2	*	*	0.60
Jones & Wein (1972)	72	F	I	2/1/2/1/2	*	—	0.47
Pavlos (1972)	90	B	P	2/1/2/2/1	*	—	0.48
Sigall & Page (1972)	76	F	I	1/1/3/1/2	*	*	1.14
Pavlos (1973)	60	B	P	2/1/2/2/1	*	*	1.30
Brigham et al. (1974)	225	B	P	2/1/3/1/2	*	—	0.32
Second wave							
Allen (1975), Study 4	30	B	P	2/1/2/1/2	*	*	0.93
Hough & Allen (1975)	42	F	P	2/1/2/1/2	*	*	0.69
Page & Moss (1975)	40	M	I	1/1/3/1/2	—	*	0.00
Cherry et al. (1976)	192	B	I	1/1/2/1/1	—	*	0.15
Rivera & Tedeschi (1976)	60	F	O	1/1/2/1/1	*	*	0.88
Schlenker et al. (1976)	120	B	P	1/1/3/1/2	*	—	0.21
Guild et al. (1976)	48	B	A	1/1/2/1/2	—	—	0.38
Carver, Glass, & Katz (1978)	90	F	P	2/1/1/2/1	—	*	0.47
Gaes, Kalle, & Tedeschi (1978), Study 1	40	F	A	1/1/2/1/1	—	*	1.06
Ward (1978)	80	F	P	2/1/2/1/2	—	—	0.05
Arkin et al. (1980), Study 2	80	F	A	1/1/4/1/1	*	*	0.43
Riess, Kalle, & Tedeschi (1981)	20	B	A	1/2/1/2/3	*	—	0.34
Riess, Rosenfeld, Melburg, & Tedeschi (1981)	48	F	A	1/1/2/1/3	*	—	0.78
Riordan & Tedeschi (1981)	40	F	A	1/2/2/1/1	—	—	0.54
Third wave							
Malkis et al. (1982), Study 1	56	F	A	2/1/2/1/3	*	—	0.65
Mummendey et al. (1982)	108	B	P	1/2/3/1/1	—	*	0.24
Paulhus (1982)	61	B	A	1/1/3/1/1	*	*	1.30
Arkin & Lake (1983)	204	B	I	1/1/1/2/1	—	*	0.40
Kunda & Schwartz (1983)	60	B	A	1/1/4/1/1	*	*	0.56
Schul & Knapp (1984)	107	B	A	1/2/1/2/1	—	*	0.52
Stulms et al. (1984)	24	M	A	1/1/3/2/1	*	—	0.87
Mehlman & Snyder (1985)	96	M	A	2/2/1/1/2	—	*	0.19
McGovern & Nevid (1986)	120	M	O	1/2/2/1/2	—	—	0.00
Eisenberg et al. (1989)	74	B	O	2/2/1/1/2	—	—	0.00
Rosenfeld (1990)	117	F	A	1/1/2/1/1	—	—	0.00

Note. N = sample size used to calculate effect size; Sex = sex of sample (F = female, M = male, B = both female and male); RA = research area (A = attitudes/attributions, I = interpersonal attraction, P = prejudice [including racism and sexism], O = other); ME = main effect reported for bogus pipeline (BPL) (asterisk indicates significant effect); IE = interaction effect (asterisk indicates significant interaction reported between BPL and another independent variable); ES = corrected effect size (*d*).

^a The first variable was demand (1 = subjects were explicitly told that apparatus detected lies, 2 = subjects were not told that apparatus detected lies); the second variable was task instruction (1 = subjects were instructed to predict apparatus output, 2 = subjects were instructed to respond accurately); the third variable was assessment format (1 = subjects' BPL responses were made in written form, 2 = responses were made on mechanical device; 3 = responses were made orally, 4 = combination of written and mechanical); the fourth variable was use of verisimilitude phase (1 = BPL apparatus was demonstrated to subjects, 2 = BPL apparatus was not demonstrated); the fifth variable was type of control condition (1 = paper-and-pencil only, 2 = BPL apparatus was present, but its purpose was not explained, 3 = BPL apparatus was present, but it was explained as unreliable).

moved (Hedges & Olkin, 1985) from the guess-output group, but this did not result in homogeneity within this group.

The continuous variable of year of publication predicted effect size by means of a simple linear regression model ($R^2 = .13$), $Q_R(2) = 8.62$, $p < .05$. More recent studies tended to yield smaller effect sizes ($b = -.022$), but the test of model specification, $Q_E(29) = 58.07$, $p < .01$, indicated that the hypothesis of correct model specification was rejected. Although many factors can account for declining effect sizes over time, the significant correlation between task instruction and year ($r = .66$, $p < .01$) suggests that the task instruction effect previously noted may account for the year-of-publication effect. Alternatively, the progressive simplification of BPL instructions and procedure may account for both the declining effect size and the declining use of the BPL over time.

Several null findings were of interest. First, the lack of an effect for research area, $Q_B(3) = 4.25$, *ns*, or sex of subject, $Q_B(2) = 5.21$, *ns*, indicated that the BPL yielded roughly equivalent effects across research domains and across gender, suggesting some degree of generality of the procedure. Second, the lack of an effect for the verisimilitude phase, $Q_B(1) = 3.26$, *ns*, (in which subjects were convinced of the BPL's powers of detection) suggested that the BPL was just as effective with or without such validation.⁹ This offered some support for the claims of some researchers (e.g., Jamieson & Zanna, 1991; Riess, Kalle, & Tedeschi, 1981) who advocated the use of an expectation of later

⁹ The small number of studies that did not use a verisimilitude procedure limited the strength with which this inference could be made.

Table 2
Categorical Effects Tests for Study Characteristics

Study characteristic	<i>n</i>	Effect size (<i>d</i>)	95% confidence interval	Homogeneity within class ^a
Sex of sample				
Female	12	0.51	0.37 to 0.65	24.44*
Male	5	0.21	0.00 to 0.42	6.12
Both	14	0.40	0.29 to 0.51	30.91*
Research area				
Prejudice	10	0.41	0.28 to 0.54	17.97*
Attitudes	13	0.38	0.35 to 0.63	22.76*
Attraction	5	0.49	0.22 to 0.55	13.53*
Other	3	0.19	-0.06 to 0.44	8.18*
Demand				
Explicit	20	0.41	0.31 to 0.51	46.12*
Implicit	11	0.41	0.28 to 0.54	20.56*
Task instruction ^b				
Guess output	24	0.46	0.37 to 0.55	54.85*
Accurate	7	0.22	0.06 to 0.39	5.75
Assessment format				
Written	6	0.35	0.19 to 0.51	4.11
Mechanical	15	0.39	0.27 to 0.51	37.18*
Oral	8	0.46	0.31 to 0.61	23.95*
Written + mech	2	0.49	0.15 to 0.82	0.13
Verisimilitude				
Demonstrated	24	0.37	0.28 to 0.46	54.52*
Not demonstrated	7	0.54	0.37 to 0.70	8.91
Control group				
PP only	15	0.46	0.35 to 0.57	36.40*
BPL present	13	0.31	0.19 to 0.43	24.52*
Unreliable BPL	3	0.64	0.29 to 1.01	0.64

Note. Q_{wi} from Hedges and Olkin (1985); significance indicates rejection of hypothesis of homogeneity within that class of study characteristic. ^b Categorical effect for this study characteristic was significant at $p < .05$.

* $p < .05$.

verification by lie-detector in lieu of demonstrating the BPL apparatus, asserting that this procedure was similar in its effects to the standard BPL procedure. Third, the fact that demand (i.e., whether subjects were told that the BPL detected lies) was not a significant predictor, $Q_B(1) = .01$, *ns*, argued against the counternormative demand explanation for the BPL's mechanism.

General Discussion

What follows is a general assessment of the BPL procedure. We begin by tabulating and reconsidering, in light of the lessons of a 20-year-old research literature as well as the meta-analytic findings, some of the more serious criticisms leveled at the BPL. Conclusions regarding the BPL's validity are then drawn, and recommendations for its future use are made.

Summary of Criticisms of the Bogus Pipeline

Different functional relations. Ostrom (1973) argued that there was no evidence that the BPL related to other variables in a functionally distinct manner in relation to PP ratings. However, significant interactions between the BPL vs. PP manipulation and other independent variables were evident in half of the 31 studies that were meta-analyzed. Moreover, many of these were crossover interactions. This pattern strongly suggested

that BPL and PP ratings could differ in their functional relations to other manipulated variables—depending, of course, on the theoretical relation tested. Although few examples of crossover interactions were evident in BPL research at the time of Ostrom's publication, the BPL literature at this point clearly refuted this particular criticism.

Different assessment formats. Some have suggested that any BPL effect may be confounded by basic operational differences between BPL and PP assessment formats (e.g., Ostrom, 1973). In several early studies, BPL subjects were asked questions face-to-face or received more experimenter attention, whereas control subjects were left to respond on their own. Subsequent studies, however, ensured more equivalent assessments for all subjects. Moreover, an examination of survey items by Gaes, Quigley-Fernandez, and Tedeschi (1978) and Quigley-Fernandez and Tedeschi (1978) found no heterogeneity of variance across conditions of PP responses, face-to-face responses, and responses made on mechanical devices, and our meta-analysis indicated neither differences in effect sizes between these same conditions nor differences that were due to control-group format. The BPL effect could not therefore be attributed solely to confounds in assessment formats.

Motivation and the bogus pipeline. We previously noted that the BPL procedure was predicated on the assumption that it triggered subjects' desire to avoid appearing to be a liar or to avoid appearing to be out of touch with themselves (self-protec-

tion) and that this desire overwhelmed the typically assumed desire to exaggerate possession of positively valued traits (self-enhancement). It is one of the fascinating idiosyncrasies of this literature that this motivational shift, on which the entire BPL legacy is based, has never been tested directly. Researchers apparently assumed that if the BPL engendered effects as predicted, then its mechanism must operate according to these assumptions.

Some suggestive evidence, however, is consistent with this theoretical mechanism. First, the BPL does appear to evoke self-protection. Our meta-analysis showed that a task instruction to predict the BPL apparatus output produced a larger BPL effect. That is, when subjects were motivated to appear to be in touch with themselves, the BPL effect was enhanced. Second, the BPL also appears to decrease self-enhancement. Several studies showed that endorsement of socially desirable statements was attenuated by the BPL (Howard et al., 1981; Millham & Kellogg, 1980; Mummendey & Bolten, 1981). These findings combine to support a self-enhancement to self-protection motivational shift. In addition, a second possibility, namely a self-enhancement to counternormative responding shift, can be tentatively ruled out. Our meta-analysis indicated that experimental demand making BPL lie-detection salient did not enhance the BPL effect. Finally, two alternative BPL mechanisms, that the BPL increases focus on the affective component of attitudes and that the BPL reduces mindless or careless responding (see Jones & Sigall, 1971), have not been examined empirically.

Facts versus opinions. We have already noted the distinction between fact and opinion in self-report assessment and that validation of the BPL for one does not necessarily imply BPL validity for the other. There is at least one line of thinking to suggest that the mechanisms underlying self-reports of fact may sometimes differ dramatically from those underlying reports of opinion. Because they are not objectively justifiable, the latter may be more mutable. Consequently, constructs such as attitudes are more perturbable by assessment than are behavioral self-reports. Thus, the mere act of assessing an attitude may serve to crystallize or coalesce what was previously a murky collection of thoughts (e.g., Schuman & Presser, 1981). Social psychologists have assumed that attitudes, although fluid and malleable, are in most cases sturdy enough to withstand the relatively weak perturbation of standard assessment strategies (see Tourangeau & Rasinski, 1988, for a discussion). This state of affairs is different for the BPL, which may constitute a much larger perturbation: The impact of the situation is greater, the stresses involved more severe, and what in some cases are weakly held attitudes may well be greatly disturbed. This raises the possibility that the BPL may reactively shift attitudes or inspire their formation to a greater extent than PP ratings, even though subjects may give what they feel to be truthful self-reports. Although difficult to assess, this line of thinking suggests that use of the BPL in attitude assessment should proceed with caution if the attitudes are relatively weak or inaccessible.

Practicality. Beginning with Ostrom (1973), researchers have criticized the BPL for its impractical and cumbersome logistics. Subjects must be run individually, and long and careful debriefing sessions are required to adequately explain the complex deceptions. Accordingly, Jones and Sigall (1973) sug-

gested selectivity in applying the BPL, reserving it for cases in which reactivity biases may be particularly problematic.

Alternatively, a simplified variant of the BPL offers obvious advantages in its ease of implementation. Arkin and Lake (1983), Jamieson and Zanna (1991), and Riess, Kalle, & Tedeschi (1981) used the expectation of later assessment by a lie-detector apparatus to improve accuracy of attitude assessment. Subjects were told that their self-reports would be (or might be at random) verified by lie detector, and they were shown an inert polygraph device, but there was no deceptive and logistically challenging verisimilitude phase. This procedure offered partial solutions to criticisms of practicality and also to criticisms of the BPL's ethicality (if subjects were told that responses might be later verified, then less deception would be involved). Furthermore, our meta-analysis demonstrated that procedures eliminating the verisimilitude phase did not differ from the standard BPL in their effect magnitudes. Direct comparisons between standard and variant BPL procedures within the same study have not, to our knowledge, been attempted.

The use of simplified variants notwithstanding, the BPL's impracticality remains its chief drawback. Indeed, the primary advantage over the BPL of the randomized response technique (RRT), another strategy designed to reduce social desirability biases in self-reports, is its relative portability and ease of administration (e.g., Himmelfarb & Lickteig, 1982). The RRT does, nonetheless, represent more work for the researcher than a standard PP procedure (if only because of the generally larger sample size required by the RRT). Regardless of the method chosen, any reasonable effort to reduce social desirability biases necessarily entails a certain amount of additional effort.

Validity of the Bogus Pipeline: Conclusion

The meta-analysis clearly established that the BPL produces a measurable effect across studies. This was the first criterion we noted as necessary to infer the valid operation of the BPL. The second criterion was that the effect represent a shift in the appropriate direction (i.e., toward and not away from veracity). For BPL assessments of reports of fact, this was amply demonstrated using objective, independent data (e.g., Bauman & Dent, 1982; Murray & Perry, 1987; Quigley-Fernandez & Tedeschi, 1978). However, confirming that the BPL also shifts reports of opinion in the appropriate direction can only be inferred indirectly.

We conclude that the BPL effect reflects the valid operation of the procedure to reduce socially desirable responding. Indirect evidence from several sources converges to support such a conclusion. First, subjects were more likely to disclose negative information about themselves and others under BPL than PP conditions (e.g., Millham & Kellogg, 1980; Rivera & Tedeschi, 1976; Sigall & Page, 1972). Second, scores on a social desirability scale were correlated with attitudinal self-reports under PP but not BPL conditions (Paulhus, 1982). Third, whereas reactivity to experimental demands was predicted by social desirability status under PP conditions, individuals high and low in dispositional social desirability responded equivalently to such demands under BPL conditions (Arkin & Lake, 1983). Fourth, the meta-analytic results were consistent with the mechanism of social desirability reduction: The BPL effect was enhanced

when subjects were motivated to appear to be in touch with themselves. Fifth, little direct evidence exists for a BPL-induced bias for counternormative self-disclosure (Schlenker et al., 1976), and our own meta-analysis of the relation between procedural demand and BPL effect was not significant. Sixth, BPL self-reports were seemingly more stable by virtue of their closeness to truth: Self-esteem manipulations influenced expressed racial attitudes (Pavlos, 1973), and experimental demand manipulations shifted responses (Arkin & Lake, 1983) under PP but not BPL conditions. Seventh, subjects found the BPL cover story believable (e.g., Jones & Sigall, 1973), revealing little overt suspicion (e.g., Brigham et al., 1974). Thus, the findings from both qualitative and quantitative reviews of the relevant literature provided reasonable evidence that the BPL reduces social desirability biases in self-reports of opinion.

Recommended Use of the Bogus Pipeline

We turn finally to the question of how and where the BPL should be used. In the past, the BPL has been used in two ways: to examine psychological effects in a form relatively undistorted by social desirability biases and to manipulate (i.e., obviate) impression management effects. We believe that both uses are appropriate, but the former use requires additional attention. Specifically, before importing the BPL to a new research domain, researchers should first demonstrate that some form of bias affects reports in that domain. One obvious technique is to correlate reports with a social desirability scale. The BPL may also be used for this demonstration, but future researchers should avoid interpreting a null BPL effect as indicative of BPL ineffectiveness. Rather, such a null effect more likely reflects the weakness of social desirability biases within that domain. Clearly, if no initial evidence for bias exists in the domain, the BPL need not be used further. In any case, theoretical predictions of null BPL effects should be avoided to preclude criticisms that the BPL is insensitive or weakly manipulated. If BPL use is warranted, control conditions should always be included in further studies, if only to gather evidence refuting the possibility of BPL-induced counternormative demand biases.

As suggested by the meta-analysis, task instructions should explicitly require subjects to guess the apparatus output, so as to prime the motivation to avoid appearing self-unaware. Additional subjective or anecdotal evidence should also be collected, such as ratings of believability of the BPL manipulation, care in responding, and arousal. That the BPL engenders certain effects in addition to its intended role of reducing social desirability biases, such as enhancing subjects' focus on internal states (see Jones & Sigall, 1971), should also be assessed where possible. Although the meta-analysis was not consistent with a BPL-induced demand for negative self-disclosure, its indirectness rendered it equivocal. Therefore, instructions involving explicit demand should be avoided in experiments in which negative responding is predicted under BPL conditions, and the resulting data should be interpreted cautiously. Evidence for the operation of such additional mechanisms can, of course, profoundly alter the theoretical conclusions drawn.

Within the area of attitude research, the BPL is more cautiously recommended. We have noted that attitudes that are weak or inaccessible are susceptible to reactive distortion

wrought by the brute force of the BPL manipulation. The BPL can, therefore, be used with confidence when assessing attitudes that are relatively central and accessible, but its application must be more carefully considered when this is not the case. A simplified BPL variant, such as the lie-detector expectation procedure (Jamieson & Zanna, 1991), may be less likely to reactively distort weakly held attitudes and is therefore recommended for this domain in particular.

Conclusion

In the introduction to this article, we noted that the decline in use of the BPL by social psychologists could be attributed to any of several possibilities: the BPL's ineffectiveness, its impracticality, its uncertain theoretical grounding, and its questionable ethics. We argued, on the basis of a review and meta-analysis of 20 years of research using the BPL, that the first concern was untenable. Furthermore, the second and fourth concerns could be addressed through the use of simplified BPL variants that were perhaps equally effective. The third concern, however, warranted further investigation before more substantive conclusions could be drawn.

The near abandonment of this important methodology may also be explained in another way. The rise and fall of the BPL in social psychology is something of a textbook example of the faddishness bemoaned by Kenneth Ring in his 1967 essay. The three waves that we used to characterize the BPL's history eerily paralleled Ring's complaint that within social psychology "a new (or seemingly new) territory is discovered, explored for awhile, and then usually abandoned when the going gets rough or uninteresting" (p. 120). The epistemological and validation problems inherent in the application of the BPL have indeed become difficult, but these difficulties have always lurked behind apparently clean findings, regardless of whether researchers have in the past chosen to confront them. We hope that future researchers can, by addressing the concerns we have raised, take advantage of the relative merits of the BPL in research to come.

References

- Allen, B. P. (1975). Social distance and admiration reactions of "unprejudiced" Whites. *Journal of Personality*, 43, 709-726.
- Arkin, R. M. (1981). Self-presentational styles. In J. T. Tedeschi (Ed.), *Impression management theory and social psychological research* (pp. 311-333). San Diego, CA: Academic Press.
- Arkin, R. M., Appleman, A. J., & Burger, J. M. (1980). Social anxiety, self-presentation, and the self-serving bias in causal attribution. *Journal of Personality and Social Psychology*, 38, 23-35.
- Arkin, R. M., & Lake, E. A. (1983). Plumbing the depths of the bogus pipeline: A reprise. *Journal of Research in Personality*, 17, 81-88.
- Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic technique. *Psychological Bulletin*, 99, 388-399.
- Bauman, K. E., & Dent, C. W. (1982). Influence of an objective measure on self-reports of behavior. *Journal of Applied Psychology*, 67, 623-628.
- Baumeister, R. F., & Tice, D. M. (1984). Role of self-presentation and choice in cognitive dissonance under forced compliance: Necessary or sufficient causes? *Journal of Personality and Social Psychology*, 46, 5-13.

- Baumrind, D. (1985). Research using intentional deception: Ethical issues revisited. *American Psychologist*, 40, 165-174.
- Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. *Psychological Bulletin*, 76, 186-204.
- Bowman, P. C., & Auerbach, S. M. (1978). Measuring sex-role attitudes: The problem of the well-meaning liberal male. *Personality and Social Psychology Bulletin*, 4, 265-271.
- Brackwede, D. (1980). Das bogus-pipeline-paradigma: Eine Übersicht über bisherige experimentelle ergebnisse [The bogus pipeline paradigm: A survey of experimental results to date]. *Zeitschrift für Sozialpsychologie*, 11, 50-59.
- Brigham, J. C., Bloom, L. M., Gunn, S. P., & Torok, T. (1974). Attitude measurement via the bogus pipeline: A dry well? *Representative Research in Social Psychology*, 5, 97-114.
- Byrne, D., & Cherry, F. (1978). A plumber's friend in need is a plumber's friend indeed. *Journal of Research in Personality*, 12, 193-196.
- Campanelli, P. C., Dielman, T. E., & Shope, J. T. (1987). Validity of adolescents' self-reports of alcohol use and misuse using a bogus pipeline procedure. *Adolescence*, 22, 7-22.
- Carver, C. S., Glass, D. C., & Katz, I. (1978). Favorable evaluations of Blacks and the handicapped: Positive prejudice, unconscious denial, or social desirability? *Journal of Applied Social Psychology*, 8, 97-106.
- Cherry, F., Byrne, D., & Mitchell, H. E. (1976). Clogs in the bogus pipeline: Demand characteristics and social desirability. *Journal of Research in Personality*, 10, 69-75.
- Christensen, L. (1988). Deception in psychological research: When is its use justified? *Personality and Social Psychology Bulletin*, 14, 664-675.
- Cialdini, R. B., Levy, A., Herman, C. R., & Evenbeck, C. (1973). Attitudinal politics: The strategy of moderation. *Journal of Personality and Social Psychology*, 25, 100-108.
- Cialdini, R. B., Petty, R. E., & Cacioppo, J. T. (1981). Attitudes and attitude change. *Annual Review of Psychology*, 32, 357-404.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (2nd ed.). San Diego, CA: Academic Press.
- Cohen, S. J., Katz, B. P., Drook, C. A., Christen, A. G., McDonald, J. L., Olson, B. L., Cloys, L. A., & Stookey, G. K. (1988). Overreporting of smokeless tobacco use by adolescent males. *Journal of Behavioral Medicine*, 11, 383-393.
- Cooper, J. (1971). Personal responsibility and dissonance: The role of foreseen consequences. *Journal of Personality and Social Psychology*, 18, 354-363.
- Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York: Wiley.
- Eisenberg, N., Miller, P. A., Schaller, M., Fabes, R. A., Fultz, J., Shell, R., & Shea, C. L. (1989). The role of sympathy and altruistic personality traits in helping: A reexamination. *Journal of Personality*, 57, 41-67.
- Evans, R. I., Hansen, W. B., & Mittlemark, M. B. (1977). Increasing the validity of self-reports of smoking behavior in children. *Journal of Applied Psychology*, 62, 521-523.
- Gaes, G. G., Kalle, R. J., & Tedeschi, J. T. (1978). Impression management in the forced compliance situation: Two studies using the bogus pipeline. *Journal of Experimental Social Psychology*, 14, 493-510.
- Gaes, G. G., Quigley-Fernandez, B., & Tedeschi, J. T. (1978). Unclogging the bogus pipeline: A critical reanalysis of the Cherry, Byrne, and Mitchell study. *Journal of Research in Personality*, 12, 189-192.
- Gerdes, E. P. (1979). College students' reactions to social psychological experiments involving deception. *Journal of Social Psychology*, 107, 99-110.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Guild, P. D., Strickland, L. H., & Barefoot, J. C. (1976). Dissonance theory, self-perception and the bogus pipeline. *European Journal of Social Psychology*, 7, 465-476.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
- Hill, P. C., Dill, C. A., & Davenport, E. C., Jr. (1988). A reexamination of the bogus pipeline. *Educational and Psychological Measurement*, 48, 587-601.
- Himmelfarb, S., & Lickteig, C. (1982). Social desirability and the randomized response technique. *Journal of Personality and Social Psychology*, 43, 710-717.
- Hough, K. S., & Allen, B. P. (1975). Is the "women's movement" erasing the mark of oppression from the female psyche? *Journal of Psychology*, 89, 249-258.
- Howard, G. S., Millham, J., Slaten, S., & O'Donnell, L. (1981). Influence of subject response style effects on retrospective measures. *Applied Psychological Measurement*, 5, 89-100.
- Jamieson, D. W., & Zanna, M. P. (1991). *The lie detector expectation procedure: Ensuring veracious self-reports of attitude*. Unpublished manuscript.
- Jones, E. E., Bell, L., & Aronson, E. (1972). The reciprocation of attraction from similar and dissimilar others: A study in person perception and evaluation. In C. G. McClintock (Ed.), *Experimental social psychology* (pp. 142-179). New York: Holt, Rinehart & Winston.
- Jones, E. E., & Gordon, E. M. (1972). Timing of self-disclosure and its effects on interpersonal attraction. *Journal of Personality and Social Psychology*, 24, 358-365.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 76, 349-364.
- Jones, E. E., & Sigall, H. (1973). Where there is ignis, there may be fire. *Psychological Bulletin*, 79, 260-262.
- Jones, E. E., & Wein, G. A. (1972). Attitude similarity, expectancy violation, and attraction. *Journal of Experimental Social Psychology*, 8, 222-235.
- Karls, M., Coffman, T. L., & Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology*, 13, 1-16.
- Kunda, Z., & Schwartz, S. H. (1983). Undermining intrinsic moral motivation: External reward and self-presentation. *Journal of Personality and Social Psychology*, 45, 763-771.
- Langer, E. J. (1989). Minding matters: The consequences of mindlessness-mindfulness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 137-173). San Diego, CA: Academic Press.
- Lerner, M. J., Peachey, D. E., & Meindl, J. (1979, June). *Experimental efforts to understand the effects of "overpayment": Norms, role-playing, and the bogus pipeline*. Paper presented at the Canadian Psychological Association convention, Quebec City, Quebec, Canada.
- Lowe, J. B., Windsor, R. A., Adams, B., Morris, J., & Reese, Y. (1986). Use of a bogus pipeline method to increase accuracy of self-reported alcohol consumption among pregnant women. *Journal of Studies on Alcohol*, 47, 173-175.
- Malkis, F. S., Kalle, R. J., & Tedeschi, J. T. (1982). Attitudinal politics in the forced compliance situation. *Journal of Social Psychology*, 117, 79-91.
- Martin, G. L., & Newman, I. M. (1988). Assessing the validity of self-reported adolescent cigarette smoking. *Journal of Drug Education*, 18, 275-284.
- McGovern, F. J., & Nevid, J. S. (1986). Evaluation of apprehension on

- psychological inventories in a prison-based setting. *Journal of Consulting and Clinical Psychology*, 54, 576-578.
- Mehlmán, R. C., & Snyder, C. R. (1985). Excuse theory: A test of the self-protective role of attributions. *Journal of Personality and Social Psychology*, 49, 994-1001.
- Mikulincer, M., & Marshand, O. (1991). An excuse perspective of the learned helplessness paradigm: The self-protective role of causal attribution. *Journal of Social and Clinical Psychology*, 10, 134-151.
- Millham, J., & Kellogg, R. W. (1980). Need for social approval: Impression management or self-deception? *Journal of Research in Personality*, 14, 445-457.
- Mummendey, H. D. (1981). Methoden und probleme der kontrolle sozialer erwünschtheit [Methods and problems of social desirability control]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 2, 199-218.
- Mummendey, H. D., & Bolten, H. G. (1981). Die veränderung von social-desirability-antworten bei erwarteter wahrheitskontrolle (bogus-pipeline-paradigma) [Modification of social desirability responses when control of truthful responding is anticipated: The bogus pipeline paradigm]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 2, 151-156.
- Mummendey, H. D., Bolten, H. G., Isermann-Gerke, M. (1982). Experimentelle überprüfung des bogus-pipeline-paradigmas: Einstellungen gegenüber Türken, Deutschen und Holländern [Experimental replication of the bogus pipeline paradigm: Turkish, German, and Dutch views compared]. *Zeitschrift für Sozialpsychologie*, 13, 300-311.
- Murray, D. M., O'Connell, C. M., Schmid, L. A., & Perry, C. L. (1987). The validity of smoking self-reports by adolescents: A reexamination of the bogus pipeline procedure. *Addictive Behaviors*, 12, 7-15.
- Murray, D. M., & Perry, C. L. (1987). The measurement of substance use among adolescents: When is the bogus pipeline method needed? *Addictive Behaviors*, 12, 225-233.
- Myers, D. G. (1990). *Social psychology* (3rd ed.). New York: McGraw-Hill.
- Olson, J. M., & Zanna, M. P. (1993). Attitudes and attitude change. *Annual Review of Psychology*, 44, 117-154.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.
- Ostrom, T. M. (1973). The bogus pipeline: A new ignis fatuus? *Psychological Bulletin*, 79, 252-259.
- Page, R. A., & Moss, M. K. (1975). Attitude similarity and attraction: The effects of the bogus pipeline. *Bulletin of the Psychonomic Society*, 5, 63-65.
- Paulhus, D. (1982). Individual differences, self-presentation, and cognitive dissonance: Their concurrent operation in forced compliance. *Journal of Personality and Social Psychology*, 43, 838-852.
- Pavlos, A. J. (1972). Racial attitude and stereotype change with bogus pipeline paradigm. *Proceedings of the 80th Annual Convention of the American Psychological Association*, 7, 292-292.
- Pavlos, A. J. (1973). Acute self-esteem effects on racial attitudes measured by rating scale and bogus pipeline. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 8, 165-166.
- Quigley-Fernandez, B., & Tedeschi, J. T. (1978). The bogus pipeline as lie detector: Two validity studies. *Journal of Personality and Social Psychology*, 36, 247-256.
- Riess, M., Kalle, R. J., & Tedeschi, J. T. (1981). Bogus pipeline attitude assessment, impression management, and misattribution in induced compliance settings. *Journal of Social Psychology*, 115, 247-258.
- Riess, M., Rosenfeld, P., Melburg, V., & Tedeschi, J. T. (1981). Self-serving attributions: Biased private perceptions and distorted public descriptions. *Journal of Personality and Social Psychology*, 41, 224-231.
- Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology*, 3, 113-123.
- Riordan, C. A., & Tedeschi, J. T. (1981). Effort justification reflects genuine cognitive change. *Replications in Social Psychology*, 1, 41-44.
- Rivera, A. N., & Tedeschi, J. T. (1976). Public versus private reactions to positive inequity. *Journal of Personality and Social Psychology*, 34, 895-900.
- Rosenfeld, P. (1990). Self-esteem and impression management explanations for self-serving biases. *Journal of Social Psychology*, 130, 495-500.
- Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review*, 40, 337-354.
- Scheier, M. F., & Carver, C. S. (1980). Private and public self-attention, resistance to change, and dissonance reduction. *Journal of Personality and Social Psychology*, 39, 390-405.
- Schlenker, B. R., Bonoma, T. V., Hutchinson, D., & Burns, L. (1976). The bogus pipeline and stereotypes toward Blacks. *Journal of Psychology*, 93, 319-329.
- Schul, A., & Knapp, J. R. (1984). On attitude change: Issue-involvement and demand characteristics. *Psychological Reports*, 55, 547-553.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments in question form, wording, and context*. San Diego, CA: Academic Press.
- Sharpe, D., Adair, J. G., & Roese, N. J. (1992). Twenty years of deception research: A decline in subjects' trust? *Personality and Social Psychology Bulletin*, 18, 585-590.
- Sigall, H., & Page, R. A. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, 16, 252-258.
- Sigall, H., & Page, R. A. (1972). Reducing attenuation in the expression of interpersonal affect via the bogus pipeline. *Sociometry*, 35, 629-642.
- Stults, D. M., Messe, L. M., & Kerr, N. L. (1984). Belief discrepant behavior and the bogus pipeline: Impression management or arousal attribution. *Journal of Experimental Social Psychology*, 20, 47-54.
- Tedeschi, J. T., & Rosenfeld, P. (1981). Impression management theory and the forced compliance situation. In J. T. Tedeschi (Ed.), *Impression management theory and social psychological research* (pp. 147-177). San Diego, CA: Academic Press.
- Tetlock, P. E., & Manstead, A. S. R. (1985). Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review*, 92, 59-77.
- Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103, 299-314.
- Ward, C. (1978). Methodological problems in attitude measurement: Sex roles, social approval, and the bogus pipeline. *Representative Research in Social Psychology*, 9, 64-68.
- Wicklund, R. A., & Brehm, J. W. (1976). *Perspectives on cognitive dissonance*. Hillsdale, NJ: Erlbaum.

Received May 8, 1992

Revision received February 11, 1993

Accepted March 2, 1993 ■