European Journal of Social Psychology, Vol. 15, 263–280 (1985)

Methods of coping with social desirability bias: a review

ANTON J. NEDERHOF
LISPOR, University of Leyden,
Stationsplein 242, 2312 AR Leiden,
The Netherlands

Abstract

Social desirability is one of the most common sources of bias affecting the validity of experimental and survey research findings. From a self-presentational perspective, social desirability can be regarded as the resultant of two separate factors: self-deception and other-deception. Two main modes of coping with social desirability bias are distinguished. The first mode comprises two methods aimed at the detection and measurement of social desirability bias: the use of social desirability scales, and the rating of item desirability. A second category comprises seven methods to prevent or reduce social desirability bias, including the use of forced-choice items, the randomized response technique, the bogus pipeline, self-administration of the questionnaire, the selection of interviewers, and the use of proxy subjects. Not one method was found to excel completely and under all conditions in coping with both other-deceptive and self-deceptive social desirability bias. A combination of prevention and detection methods offers the best choice available.

INTRODUCTION

The validity of social science research data has been the subject of a deep and serious concern in the past decades (Nederhof and Zwier, 1983). A large number of artifacts have been identified which may impair the results of laboratory and survey research, such as social desirability, various response styles (e.g. Cronbach, 1950; Messick, 1968), the experimenter expectancy effect (Rosenthal, 1976), demand characteristics (Orne, 1973; Rosnow and Aiken, 1973), volunteer characteristics (Nederhof, 1985; Rosenthal and Rosnow, 1975), subject effects (Weber and Cook, 1972), evaluation apprehension (Rosenberg, 1969), response effects in surveys (Bradburn and Sudman, 1979; Sudman and Bradburn, 1974),

This paper was partly written while the author was a Fulbright Visiting Professor at the Departments of Rural Sociology and Sociology, Washington State University, Pullman Washington.

0046–2772/85/030263–18$01.80

Received 12 April 1984
Revised 5 November 1984

pre-test effects (Campbell and Stanley, 1966; Lana, 1969), non-response (Dillman, 1978; Nederhof, 1983), response sets (Berg, 1967; Crowne and Marlowe, 1960), and various others (Barber, 1976; Cook and Campbell, 1979; Nederhof, in press; Rosenthal and Rosnow, 1969).

Serious as the validity problems in data collection may be, for many of those mentioned above, techniques have been developed to mitigate the biasing effects. The present report will review methods of coping with social desirability bias. Previous reviews of the research to social desirability, although often excellent, have mainly been devoted to discussing the theoretical implications of studies on a related but more general concept, the need for social approval, and have not dealt with methods of coping with social desirability bias (e.g. Millham and Jacobson, 1978; Strickland, 1977).

A large number of studies have shown that social desirability may seriously bias data, both in the laboratory (Edwards, 1953; Crowne and Marlowe, 1964) and in surveys (e.g. Klassen, Hornstra and Anderson 1976; Phillips and Clancy, 1970, 1972). Psychologists as well as sociologists have developed methods for coping with social desirability bias. To assess the effectiveness of these various methods, evidence stemming from both the psychological and sociological literature will be integrated. Although social desirability is an interesting subject of investigation in its own right, and is studied as such in, for instance, personality research (*cf.* McGuire, 1969), here attention will be focused on social desirability as a nuisance factor and as a source of systematic bias in social research.

Social desirability reflects the tendency on behalf of the subjects to deny socially undesirable traits and to claim socially desirable ones, and the tendency to say things which place the speaker in a favourable light. Norms are important determinants of socially desirable behaviour, as they determine what constitutes a good impression in a given situation (*cf.* Atteslander and Kneubühler, 1975). An example concerns the admittance of violations of the law. According to the relevant norms of the situation, a person could admit more or fewer violations of the law than actually committed. A member of a youth gang, for instance, may admit to many more violations of the law than he or she committed in order to make an impression.

Social desirability will be regarded here as a distortion of responses in a socially desirably direction which is a resultant of two factors: 'self-deception' and 'other-deception' (*cf.* Sackeim and Gur, 1978; Paulhus, 1984). When the respondent actually believes a statement to be true of him or herself, even though it is inaccurate, 'self-deception' occurs (Millham and Kellogg, 1980). On the other hand, a person might purposely misrepresent the truth as a form of impression management motivated by a desire to avoid evaluation (Millham and Kellogg, 1980: 447; *cf.* also Goffman, 1959). The distinction is important, since it allows separation of situational and personal determinants of social desirability bias. In coping with social desirability bias, situational determinants mainly related with other-deception can be influenced and manipulated by the researcher, whereas the personal determinants associated mainly with self-deception are less easily controlled and can, in most cases, only be detected (*cf.* Baumeister, 1982).

An important early contribution to the investigation of social desirability was provided by Edwards (1953, 1957). Edwards attempted to measure the impact of social desirability in the Minnesota Multiphasic Personality Inventory. Judges'

ratings of the social desirability values of self-descriptive items were correlated 0.87 with their probability of endorsement by subjects, explaining more than 75 per cent of the total variance. Fortunately, the impact of social desirability upon outcomes is not always as large as in Edwards' case (Gove and Geerken, 1977). Klassen *et al.* (1976) found for instance a 'pervasive but small effect' of social desirability bias, with the highest correlation accounting for 10 per cent of the variance. Social desirability accounted for no more than 13.7 per cent of the variance in another study (Phillips and Clancy, 1970). Although the amounts differ considerably, as could be expected because of the importance of the environmental influence on social desirability, the accumulated evidence concerning the pervasiveness of social desirability strongly suggests the desirability of the application of appropriate methods to control the social desirability bias in many studies (Crowne and Marlowe, 1964; Edwards, 1953, 1970; Phillips, 1971, 1973; Rorer, 1965). A number of these methods will be discussed below.

Among the methods of coping with social desirability bias, two main categories will be distinguished. The first category comprises those methods which aim at detection and measurement of social desirability effects. In this context, the use of social desirability scales and the rating of item desirability will be discussed. The second category includes those methods which can be used to prevent or reduce social desirability bias. Seven of these methods are evaluated, including the use of forced-choice items, neutral items, the randomized response technique, the bogus pipeline, self-administration of the questionnaire, selection of interviewers, and the use of proxy subjects.

DETECTION AND MEASUREMENT OF SOCIAL DESIRABILITY BIAS

Social desirability scales

A number of social desirability scales have been developed (scales employed most often in a clinical context will not be discussed here). Among these, the Marlowe–Crowne Social Desirability Scale is easily the most popular scale, both in psychological and sociological studies[1].

Crowne and Marlowe constructed a social desirability scale consisting of items drawn from a domain of behaviours which are 'culturally sanctioned and approved, but which are improbable of occurrence', like 'Before voting, I thoroughly investigate the qualifications of all candidates' (Crowne and Marlowe, 1964). The Marlowe–Crowne Social Desirability Scale (or MC scale) has been translated into several languages (e.g. Grabitz-Gniech, 1971; Hermans, 1967), and various shortened versions are available (Greenwald and Satow, 1970; Nederhof 1981). The reliability of the MC scale is fairly good. Crowne and Marlowe (1964) reported a test–retest correlation of 0.88, and an internal consistency coefficient of 0.88. The authors contended that the MC scale was a measure of a 'need for social approval', a more general concept than social desirability.

Although the MC scale has exhibited behavioural correlates more clearly than other social desirability scales (Crowne and Marlowe, 1964; Strickland, 1977),

[1]For a discussion of several older, infrequently used scales see Wiggins (1968) and Paulhus (1984).

doubts have been raised with respect to the validity of the MC scale as a measure of the need for social approval. Various studies have tried to test the validity of the MC scale by behavioural assessments. Shulman and Silverman (1974) contended that subjects who scored low on the MC scale were behaviourally seeking approval and tried to avoid disapproval, as distinct from high scoring subjects, findings which are in contradiction with expectations based on the need for social approval theory of Crowne and Marlowe (1964)[2].

A recent investigation of the validity of the MC scale as a measure of social desirability bias was undertaken by Stocking (1979). She argued that the response bias interpretation of the MC scale precludes that scores on the MC scale be related to real life behaviour. Stocking found for instance that, in a face-to-face survey, older people scored higher on the MC scale, while actually engaging less in socially undesirably behaviour (drinking alcoholic beverages) than younger persons. It should be noted, that this could have been expected, since scores on the MC scale are related to conformity behaviour (e.g. Strickland, 1977). Stocking failed to show that high MC scorers did not distort their answers in a socially desirable direction. The MC scale seems not fit to predict the exact frequency of undesirable behaviour, but may be used to give an indication of the extent to which people distort their reports about a particular behaviour or attitude in a socially desirable direction.

Other validation studies investigated the relation between frequency and amount of cheating behaviour and MC scores (Jacobson *et al.*, 1970; Millham, 1974). When detection was viewed as unlikely, those who cheated had, as predicted, higher MC scores than noncheaters (Millham, 1974). With respect to amount of cheating, however, it turned out that high MC scorers cheated only enough to avoid dismal failure, but did not cheat when approval could be won. High MC scorers violated social norms when detection was viewed as unlikely, and violation was instrumental in avoiding disapproval (Millham, 1974). More evidence has accumulated that the MC scale predominantly measures a motive to avoid disapproval rather than a need for seeking social approval (e.g. Crandall, 1966; Jacobson and Ford, 1966).

Other evidence pertains to the internal validity of the MC scale. A number of investigations have attempted to show that the Marlowe–Crowne scale consists of two significantly related, but clearly distinguishable components; the tendency to attribute socially desirable, but improbable characteristics to oneself (the items keyed 'true'), and the tendency to deny undesirable but probable characteristics (the items keyed 'untrue') (Millham, 1974; Millham and Jacobson, 1978). However, after balancing for direction of keying, Ramanaiah and Martin (1980) showed that the attribution and denial subscales of the MC scale correlated at the same level as the individual scale reliabilities. They concluded that the attribution and denial components are equivalent measures of the same construct. In an

[2]However, several flaws disgraced this study. Shulman and Silverman (1974) had made subjects fail or succeed in solving all items of an intelligence test type of scale by halving the time and doubling the number of items in the 'failure' condition. It was noticed unobtrusively whether subjects headed for the secretary's office down the hall to make an appointment to discuss the results with the main investigator, or out of the laboratory. Before they reached either destination, subjects were called back by the experimenter. This design makes it impossible to assess how many subjects actually would have made an appointment and showed up. Both in the failure and success conditions, high and low MC-scorers did not differ in tendency to make an appointment (Schulman and Silverman reported a highly significant difference in the failure condition, but a check on the figures showed the χ^2 to be a nonsignificant 2.86 instead of 11.38).

interesting series of studies, Paulhus (1984) factor-analysed the MC scale, the Self- and Other-deception Questionnaires (Sackeim and Gur, 1978), the Social Desirability scale (Edwards, 1957), the MMPI lie-scale (Meehl and Hathaway, 1946) and the Wiggins Sd scale (Wiggins, 1959). Two factors were obtained, which were best marked by self-deception and other-deception (impression management). The MC scale loaded on both factors, although more strongly (0.68) on other-deception than on self-deception (0.36). Subsequently, Paulhus pitted the self-deception/other-deception model against the attribution/denial model in a confirmatory factor analysis. The self-deception/other-deception model provided a better fit to the data than either the attribution/denial model or a single-factor model (Paulhus, 1984, study 2). Using the same six scales as in his first study, Paulhus (1984, study 3) finally showed that administration under either anonymous or public conditions did not significantly change scores of scales loading on the self-deception factor, but affected scores of scales loading on the other-deception factor. Although the MC scale loaded on both components of socially desirable responding, the Self-Deception Questionnaire and the Other-Deception Questionnaire (Sackeim and Gur, 1978) marked both factors best.

By present, although not by contemporary standards, the procedures employed by Marlowe and Crowne in the construction of their scale cannot be considered as completely rigorous (Edwards, 1970; Millham and Jacobson, 1978). Millham and Jacobson (1978) list several methodological flaws concerning the development of the MC scale. First, the degree of social desirability was not controlled in item selection, as the test items were rated dichotomously. In addition, the final selection of items was dependent upon the responses of a small homogenous sample of 76 introductory psychology students. Therefore, the scale was not validated for use with other groups of subjects than students.

In sum, the MC scale has some validity problems. The scale seems best suited for measuring a motive to avoid negative evaluation or other-deception, but it is not a pure measure of both as it also indicates to some extent self-deception. The MC scale fails, however, as a measure of a desire to obtain positive evaluation.

New social desirability scales

Partly in response to the methodological problems of the MC scale, several new social desirability scales have been developed.

Larsen, Martin, Ettinger and Nelson (1976) constructed the 'Martin–Larsen Approval Motivation Scale'. Unfortunately, no information was offered with regard to the construct validity of this scale. This is also true of a recent revision of this scale (Martin and Greenstein, 1983).

Recently, Schuessler, Hittle, and Cardascia (1978) empirically developed a desirability scale based on attitude opinion items, intended to be particularly suited for use in survey research. The social desirability of the items was judged by a cross-section of adults in a large city. Items were screened in order that none would differ in social desirability across socially important sub-groups of the population, such as whites and blacks. Unfortunately, the reliability of the scale was rather low (0.64), and the mean inter-item correlation was only 0.10. Nevertheless, pending the necessary future validity studies, this scale might be a valuable alternative for

the MC scale in some survey contexts (*cf.* Schuessler, 1982). It is not clear, however, how the scale relates to both self-deception and other-deception.

Finally, the Other-Deception Questionnaire (ODQ) and the Self-Deception Questionnaire (SDQ), both developed by Sackeim and Gur (1978, 1979) should be mentioned. Both scales tap independent factors of socially desirable responding (*cf.* Paulhus, 1984). The alpha reliability of the SDQ was 0.73, and of the ODQ 0.74. Both scales were rationally developed. Although the performance of the ODQ and the SDQ was good in several studies (Sackeim and Gur, 1979; Gur and Sackeim, 1979; Paulhus, 1982, 1984), an important drawback for a wider use of these scales is that their construction has not been published in a refereed source. Also, the SDQ is confounded with acquiescence, whereas the ODQ confounds other-deception with nay-saying. Thus, as yet no alternative scale clearly supersedes the MC scale in all respects, although further research on the MC scale, and especially the ODQ and SDQ might result in one scale measuring both self-deception and other-deception components of socially desirably responding, but offering the possibility to assess both factors independently of each other.

Uses of scales in coping with social desirability bias

The user of a social desirability scale can employ three tactics to cope with social desirability bias: reject the data of high-scoring subjects; correct the data of high scorers; and, finally, merely register the impact of social desirability bias.

A questionable method, which nevertheless has been used sometimes, is to reject data of high-scoring subjects (*cf.* McGuire, 1969). McGuire (1969) even discerned a certain 'logic of desperation' in this tactic. Because there is no clear-cut distinction between contaminated and uncontaminated data, rejection of the data of high-scoring subjects is imbued with arbitrariness. Moreover, this approach impairs the representativeness of the sample.

A second approach aims at the correction rather than the rejection of contaminated data. An estimate is made of the degree of contamination by social desirability bias—for example, by using data of a control group or data from a neutral task. Then the contaminated scores are adjusted. Sometimes, data of high-scoring subjects are replaced by those of low- or medium-scoring subjects. However, as, for instance, research of Stocking (1979) shows, high-scoring subjects may differ not only with regard to their susceptibility to social desirability bias, but also in many other ways from lower scoring subjects. Therefore, simple correction may often lead to data which are not better than those originally at hand. Fortunately, more sophisticated methods are available.

The inclusion of a social desirability scale enables the investigator to examine whether data are contaminated at all. If not, or if not significantly so, one potential source of bias can be considered as eliminated, and more confidence can be put in the validity of the data (*cf.* Gove and Geerken, 1977). If the data are found to be considerably affected by social desirability, more attention should be paid in subsequent investigations to the application of prevention methods designed to reduce social desirability bias. A number of these methods are discussed in following sections. The biasing effect of social desirability can be viewed from several angles. Traditionally, social desirability is often seen as an issue of measurement. In personality assessment, for instance, it is often a problem whether a particular trait is measured validly, or something else, such as social desirability, is

tapped. Furthermore, social desirability can act as an intervening factor between dependent and independent variables. Three ways can be distinguished by which social desirable responding may bias results (*cf.* Ganster, Hennessey and Luthans, 1983). First, social desirability may be correlated with both dependent and independent variables, causing a spurious correlation between these variables (*cf.* Gove and Geerken, 1977). By partialling social desirability (i.e. both the self-deception and the other-deception component) from independent and dependent variables one controls for this spuriousness effect. To be on the safe side, before computing the partial correlation, one may 'correct' the zero-order correlations for attenuation if the reliability of the social desirability scale is considerably less than perfect. Second, social desirability can hide relationships by acting as a suppressor variable. Partialling social desirability from both independent and dependent variables would in this case result in an enhanced partial correlation. Finally, social desirability can act as a moderator variable, although it may (or may not) be uncorrelated with either the independent or dependent variables (e.g. Ganster *et al.*, 1983). In this case, there is an interaction effect between the independent variable and social desirable responding which can be tested for by using products terms in hierarchical multiple regression (Cohen and Cohen, 1975). In addition, it is possible to apply statistical correcton techniques such as factor-deletion, target-rotation, and covariance analysis (e.g. Paulhus, 1981). One of the main problems with these techniques is the possibility that together with social desirability bias also an intrinisc aspect of a construct is eliminated. For instance, the first factor of the MMPI, variously labelled 'anxiety', 'ego-resiliency' and 'neurosis', is intrinsically linked with social desirability (Block, 1965). Constructs like 'ego-resiliency' or 'perceived control' are actually defined in terms of biases in self-perception, and controlling for any self-deception bias in self-report which is honestly believed by the respondent should not be eliminated. However, in all of these and similar cases, the other-deception component of socially desirably responding should be controlled (*cf.* Paulhus, 1984).

When a scale is administered, an important question is whether all items should be presented sequentially or dispersed among the other items. If the effect of social desirability is constant over time, items should preferably be presented sequentially. If, however, the effect of social desirability varies over time, a dispersion of scale items is to be preferred. Empirical evidence for this last supposition was offered by Blake and Heslin (1971), who found that the effect of social desirability decayed over time during an experiment. Dickensberger, Holtz and Gniech (1978) showed that the MC scale does not merely measure stable personality traits, but is also dependent upon the social context. As the social context changes during the experiment of interview, so may the effect of social desirability, and, in particular the other-deception component. To allow for these changes in social desirability over time, items of a social desirability scale or other-deception scale should not be presented sequentially, but dispersed over the session.

Rating item desirability

The concept of item desirability, a quality of the measurement item, is measured by asking subjects to rate an item on a desirability scale. If 'being happy' is an item,

subjects are asked to indicate how desirable being happy is to them. Phillips and Clancy (1972) found item desirability to be generally unrelated to subjects' scores on the MC scale. The rating of item desirability is a cumbersome method when the number of items is large. The length of the questionnaire is doubled since one desirability rating is added to each item.

Alternatively, the social desirability of items can be assessed independently by judges. However, judges have frequently been shown to disagree considerably among themselves about the desirability of items (e.g. Scott, 1963). The rating of item desirability is a basic approach to the measurement of social desirability (*cf.* Edwards' (1970) Social Desirability Scale Values). The exact relationship of SDSV's to both other-deception and self-deception have not been clearly and unequivocally established. Presumably, ratings of item desirability are predominantly related to other-deception, and only in specific cases to self-deception. As yet, research on this topic has been lacking.

METHODS TO PREVENT OR REDUCE SOCIAL DESIRABILITY BIAS

After having discussed methods which aim primarily at detection and measurement of social desirability bias, attention will be focused upon methods intended to prevent or reduce the effects of social desirability distortion. Most of the methods in this category deal with the other-deceptive, situational determinants of social desirability bias. Three methods deal with the mode of presentation of items to subjects; one deals with the selection of interviewers; one with the choice of subjects; and two focus on the situation in which data collection takes place. Finally, the proxy subject method addresses the self-deceptive determinant of social desirability bias.

Forced-choice items

Several sources recommend the use of forced-choice items as a method of coping with social desirability (e.g. Edwards, 1957, 1970). Subjects are made to choose between two items relating to different topics, supposedly possessing an equal degree of social desirability. The rationale behind this procedure is, that if both items seem equally desirable or undesirable to subjects, their choice is not influenced by social desirability. Forced-choice items are among the earliest developed methods of coping with social desirability bias (e.g. Humm and Wadsworth, 1939). An example is to be seen in the Preference Record of Kuder (1939): 'Take a photograph of a champion swimmer', or 'take a photograph of a table you would like to make'.

Notwithstanding the creativity of this solution, it has been subjected to criticism (e.g. Block, 1965). Although the items may, on the average, be judged to be of equal social desirability, for each individual subject there may still be a difference in desirability of the items influencing the choice (*cf.* Scott, 1963). In addition, the choices are frequently unrealistic ones, which according to Scott (1968, p. 212), 'may antagonize some people and elicit responses that are highly unstable'. Missing data are also common with forced-choice items since respondents can become too frustrated to make a choice (e.g. Edwards and Diers, 1962). Finally, the consistent

pitting of the same two attributes against each other results in a perfect artificial negative correlation between them, establishing their status as 'opposites' by definition instead of empirical demonstration.

Neutral questions

A method comparable to forced-choice items is the posing of questions which are supposed to be 'neutral' with regard to social desirability. The same objection applies as made regarding forced-choice items. For individuals, some of the items may still be perceived as being socially desirable. However, as the item is judged consistently by individual members of different social groups as not being subjected to social desirability, the possible influence of dissenters is reduced, although not completely. In many cases, it will not be possible to use only neutral items (*cf.* Nunnally, 1978), the more so since items with a neutral social desirability scale value are comparatively rare (e.g. Edwards, 1970).

The randomized response technique

A particular kind of item is highly susceptible to social desirability bias. These items are referred to as 'threatening questions' (Sudman and Bradburn, 1974). Bradburn, Sudman, Blair and Stocking (1978) report that the more threatening a question is, the more often subjects choose not to respond, and the more frequently a socially desirable answer is given. Also, the reported frequency of a behaviour (e.g. drinking of liquors) declines (Locander, Sudman and Bradburn, 1976). questions concerning sex, drugs, and income appeared to be the most threatening. Under-reporting of the amount or frequency was reduced when long introductions and open response formats were used (Bradburn and Sudman, 1979).

A technique developed to reduce response distortion of threatening or personal items is the 'randomized response technique' (Himmelfarb and Lickteig, 1982; Warner, 1965). The randomized response technique allows subjects to answer to one of two randomly selected items with the interviewer not knowing what item was answered (for more details, see Bradburn and Sudman, 1979; Tracy and Fox, 1981). By knowing the probability of offering each item, the sample size, and the proportion of subjects who endorse the item to which they respond, the true proportion of subjects that have answered in a certain way can be estimated. An adequate estimation requires a substantial sample size. Bradburn and Sudman (1979) found the randomized response technique to produce less underreporting to very threatening items, although it did not eliminate underreporting altogether. In a study by Wise (1980), reduction in response bias was not related to score on the MC scale. Although the randomized response model seems to reduce response bias on very threatening items somewhat, especially when compared with direct questioning in face-to-face situations (e.g. Tracy and Fox, 1981), the evidence regarding the reduction of social desirability bias in particular is not convincing as yet (Miller, 1981). Additionally, the randomized response model poses some problems for more complex methods of analysis by containing a certain percentage (usually 30 per cent) of random 'noise'. Furthermore, in many cases, reports make mention of large percentages of subjects doubting the randomness of the choice categories offered (e.g. Soeken and Macready, 1982). For these subjects, the

technique may be ineffective. Finally, there is the problem of 'respondent jeopardy': respondents who do not have the sensitive attribute, but who are directed by the randomized response procedure to answer to the alternative item as if they did[3]. If the target item is sufficiently sensitive, a considerable percentage of subjects may refuse to answer truthfully to the insensitive alternative, thus undermining the effectiveness of the technique (Edgell, Himmelfarb and Duchan, 1982). Therefore, the randomized response technique seems to be of limited use for many types of study, and is certainly no panacea for coping with social desirability bias. It should be noted that the randomized response technique is heavily researched, so that solutions for these problems might yet be forthcoming.

Self-administered questionnaires

The use of self-administered questionnnaires seems a logical method to reduce social desirability bias in that it reduces the salience of social cues by isolating the subject. Several reports are available recommending the self-administration of questionnaires as a way to minimize the effects of social desirability (e.g. Sudman and Bradburn, 1974). Bradburn and Sudman (1979), however, found no difference in distortion of answers when varying the method of administration (face-to-face, telephone, self-administered, and the randomized response model). As Bradburn and Sudman applied the self-administration method as a part of a face-to-face survey, the effect of the self-administered method might have been reduced by previous interactions of subjects and interviewers, or by the fact that self-adminstration in the presence of the interviewer is not really anonymous. In mail surveys, where results cannot be biased by interviewer presence and anonymity seems more assured, data are generally found to be less influenced by social desirability than are results of telephone or face-to-face interviews (Nederhof, 1984; Wiseman, 1972).

Although self-administration may not always reduce social desirability bias, there are indications that mail surveys and anonymous mass self-administration give rise to less distortion than other methods of administration (*cf.* Baumeister, 1982; Paulhus, 1984). The absence of an interviewer limits use of self-administration to cases where the questionnaire is not complex and items are mainly of the closed-ended type.

The bogus pipeline

A relatively new development has been the introduction of a kind of pseudo-lie detector, called the 'bogus pipeline', in experimental laboratory research (Jones and Sigall, 1971; Sigall and Page, 1971). Subjects are induced to believe that an impressive looking physiological apparatus called the electromyograph (EMG) to which they are or will be attached by means of electrodes, can detect whether or not they speak the truth (Ries, Kalle and Tedeschi, 1981).

Subjects are convinced by several rigged demonstrations that the device can accurately measure attitudinal direction and intensity. Thus, they are placed in a

[3]Recently, Bourke (1984) has made some theoretical progress with regard to this point by means of his unrelated question design with symmetric response.

situation where they have to choose between revealing their true attitudes, appearing to be out of touch with their true feeling, or being perceived as liars (Quigley-Fernandez and Tedeschi, 1978).

The application of the bogus pipeline procedure is very circumstantial. The construction of the EMG is not an easy job, the equipment is hard to transfer, and only one subject can be managed at a time. Also, it takes a lot of time to give the bogus pipeline credibility in the eyes of the subjects. Even when these problems are solved, a liability of the bogus pipeline is that the procedure cannot be used again with the same subjects after debriefing. Public knowledge about the procedure would also limit its use, so that frequent or heavy use is prohibited (Ostrom, 1973). In fact, the procedure has hardly been used a score of times in the published literature since its introduction in 1971, although recently its popularity has been increasing somewhat.

Even though it is reasonably well established that the bogus pipeline reduces social desirability bias (Arkin and Lake, 1983; Quigley-Fernandez and Tedeschi, 1978), it is, as yet, unclear which psychological interpretation of the bogus pipeline effect is right: a pseudo-lie detector effect which reduces social desirability bias; an increase in accuracy pressure because of the evaluation situation by means of an apparatus (Brigham, Bloom, Gunn and Torok, 1974), or the increased salience of the affective rather than cognitive components of an attitude by the physiological character of the bogus pipeline. The study by Millham and Kellogg (1980) has shown that the bogus pipeline reduces the other-deceptive component of social desirability bias, but not the self-deceptive component (*cf.* Paulhus, 1982). Therefore, its effectiveness in coping with social desirability bias is limited, although useful. It should be added that all procedures which try to cope with social desirability bias by means of influencing the situation in which measurement takes place, are also unable to reduce self-deceptive social desirability bias.

It seems that, although the bogus pipeline is reasonably effective in reducing other-deceptive social desirability bias, as yet the procedure is not suited to use outside the laboratory. Inside the laboratory, it can be used in crucial experiments, but the procedure cannot be used frequently.

Selecting interviewers

When data are collected through interviewers, some efforts are usually made to establish an adequate rapport with subjects (Scott, 1968). Some texts refer to rapport as creating subject motivation; others view it mainly as generating free and frank answers, thereby obtaining more valid responses (*cf.* Goudy and Potter, 1975). As a number of studies have shown, the results of an interview are more likely to be biased in a socially desirable way when subjects are similar to interviewers with respect to social distance (Dohrenwend, Colombotos and Dohrenwend, 1968; Weiss, 1968, 1969; Williams, 1968, 1969; see also Dohrenwend, 1969). If subject and interviewer belong to similar social groups, the interview is likely to assume the character of a social exchange, where norms of polite interaction get the better of accurate reporting. The much recommended establishment of 'good rapport' between subject and interviewer may be quite detrimental in this way to the validity of responses (Weiss, 1970; Hyman, Cobb, Feldman, Hart and Stember, 1954). Therefore, when it seems not unlikely that

social desirability may contaminate the data, and use is made of interviewers, friendly, but task-oriented interviewers, or 'professional' interviewers (Weiss, 1970) have to be preferred above 'warm' or person-oriented interviewers (e.g. Rogers, 1976).

Proxy subjects

Another method by which social desirability distortions can be reduced is by the employment of proxy subjects. Instead of the 'target person', someone who knows him or her well is questioned about the behaviour of the target person.

According to evidence collected by Sudman and Bradburn (1974), results obtained by using proxy subjects are satisfactory when behaviour is measured. They doubt, however, whether the method is useful when attitudes are to be measured. Crandall (1976) asked respondents and two or three relevant others named by them to rate the respondents' satisfaction with aspects of their life. The largest convergence between self-ratings and ratings by proxy subjects was obtained on aspects of life such as house, job, and neighbourhood. Ratings of relatives correlated higher with respondents' self-reports than ratings of the average proxy subject. About 15 per cent of the variance of proxy subject's ratings may have accurately measured how the respondent feels, whereas 30 per cent of the variance could be attributed to bias and false opinions (Crandall, 1976, p. 397). In a study by Martin and Butcher (1982), agreement between target persons and proxies was larger than 84 per cent on topics like housing and employment, although variables requiring fine detail or attitudes yielded generally a much lower agreement rate. A recent review of the literature on agreement between self-reports and ratings by proxies of personality traits shows that the highest correlations between self-reports and ratings have ranged from 0.50 to 0.70, although generally the correlations were considerably lower ($\leqslant 0.30$) (McCrae, 1982; *cf.* also Kane and Lawler, 1978). The consistency with which an individual manifests a trait, as well as the degree of observability of the trait, positively influence rating/self-report agreement. When the proxy-respondent is a spouse, ratings may account for 20–50 per cent of the variance in self-report scales[4]. In these studies, it is not clear whether losses in accuracy are offset by gains in validity with regard to social desirability. Especially care should be taken to prevent target persons from interfering with proxy reporting.

By means of this method, both other- and self-deceptive components of social desirability can be dealt with. Other-deceptive social desirability might still, although probably to a lesser degree, influence the responses: particularly in cases when the proxy subject is loath to admit having frequent interactions with persons possessing some socially undesirable trait. Also, self-deceptive bias of the target person might be exchanged for bias due to the limited knowledge of the proxy subject about the target person.

[4]It should be noted that the high correlations obtained by McCrae (1982) and McCrae and Costa (1983) may well be contaminated by target person–proxy interactions since the data were collected among spouses in mail surveys. The inclusion of a single item stating 'have discussed some of these items with my wife/husband' is clearly insufficient as it depends entirely upon respondents' honesty to admit explicitly forbidden interactions.

CONCLUSIONS AND DISCUSSION

So far, a number of methods of coping with social desirability bias have been discussed separately. From that discussion, it was clear that not all methods were equally successful. Two methods were found to be of questionable use: the use of forced-choice items, and of neutral questions. Four methods of coping with social desirability contamination appeared to be effective only under certain conditions. The randomized response technique requires a larger number of subjects, and poses problems for more complex methods of data analysis. Also, the evidence with regard to the technique's effectiveness in reducing social desirability bias in general is not entirely convincing.

A second method which offers the subjects (some) anonymity is self-administration of the questionnaire by the subject. There is little evidence that self-administration effectively reduces the influence of social desirability when the interviewer or experimenter is present. However, in general, there is less social desirability bias in mail surveys than in either personal or telephone interviews. This might well be a reason to prefer mail surveys above other survey data collection methods, especially when sophisticated methods are employed to reduce the effect of other biases such as non-response (*cf.* Dillman, 1978; Nederhof, in press). Although solitary self-administration reduces social desirability bias somewhat, it does not completely solve the problem. The same is true of the non-anonymous bogus pipeline procedure. Although it is perhaps the most effective method available to reduce other-deceptive social desirability bias, it does not reduce self-deceptive bias. Self-deceptive social desirability bias can only be detected or measured, but not, as yet, be prevented or reduced other than by employing proxy subjects. Perhaps, self-deceptive bias could be reduced by applying self-awareness increasing methods (e.g. Wicklund, 1975). When self-awareness is provoked, the accuracy of self-knowledge may, at least temporarily, increase (Wicklund, 1975).

Third, when experimenters or interviewers have to be used, professional, work-oriented persons are to be preferred above warm, person-oriented ones. Even then, socially desirable responding is more likely to occur than when there is no direct person-to-person interaction. This approach might be combined with the employment of proxy subjects. Proxy subjects may yield reliable information about target persons where public behaviour is concerned, but the scarce evidence with regard to the validity of reports about attitudes of others is less favourable. In experimental research, the use of proxy subjects will not often be possible.

All methods of reducing or preventing social desirability bias are only partially effective. Therefore, supplementary measurements of social desirability to assess its impact are necessary. Two ways of measuring social desirability were distinguished. Asking subjects to rate the desirability of each item is a basic approach, which has the disadvantage of doubling the number of responses. Another possibility is the inclusion of a social desirability scale, of which several were reviewed. The most frequently used scale, the Marlowe–Crowne Social Desirability Scale, has some validity problems. In addition the MC scale is only a crude measure of both self-deception and other-deception. Taking into account that a satisfactory alternative for the MC scale has not yet been constructed, the development and validation of a (partly) new social desirability scale would be welcomed. This new

or revised scale should be able to index both deception components of social desirability.

In conclusion, not one method excels completely and under all conditions in coping with social desirability bias. Most methods were shown to be at best reasonable palliatives. A combination of one or several prevention methods and one of the detection methods seems the best choice available. It should be noted that the exact effectiveness with regard to both other-deception and self-deception of most of the prevention methods has yet to be determined empirically.

ACKNOWLEDGEMENTS

The author wishes to thank Don A. Dillman, Leo J. T. van der Kamp, and a reviewer of the *European Journal of Social Psychology* for their comments on an earlier draft of this paper.

REFERENCES

Arkin, R. M. and Lake, E. A. (1983). 'Plumbing the depths of the bogus pipeline: A reprise'. *Journal of Research in Personality*, **17**: 81–88.

Atteslander, P. and Kneubühler, H. V. (1975). *Verzerrungen im Interview*. Westdeutscher Verlag, Opladen, West Germany.

Barber, T. X. (1976). *Pitfalls in Human Research: Ten Pivotal Points*, Pergamon Press, New York.

Baumeister, R. F. (1982). A self-presentational view of social phenomena, *Psychological Bulletin*, **91**: 3–26.

Berg, I. A. (1967). *Response Set in Personality Assessment*, Aldine Press, Chicago, Illinois.

Blake, B. F. and Heslin, R. (1971). 'Evaluation apprehension and subject bias in experiments', *Journal of Experimental Research in Personality*, **5**: 57–63.

Block, J. M. (1965). *The Challenge of Response Set: Unfounding Meaning, Acquiescence, and Social Desirability in the MMPI*, Appleton-Century-Crofts, New York.

Bourke, P. D. (1984). 'Estimation of proportions using symmetric randomized response designs, *Psychological Bulletin*, **96**: 166–172.

Bradburn, N. M. and Sudman, S. (1979). *Improving Interview Methods and Questionnaire Design*, Jossey-Bass, San Francisco, California.

Bradburn, N. M. Sudman, S., Blair, E. and Stocking, C. (1978). 'Question threat and response bias', *Public Opinion Quarterly*, **42**: 221–234.

Brigham, J. C., Bloom, L. M., Gunn, S. P. and Torok, T. (1974). 'Attitude measurement via the bogus pipeline: A dry well?' *Representative Research in Social Psychology*, **5**: 97–114.

Campbell, D. T. and Stanley, J. C. (1966). *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago, Illinois.

Cohen, J. and Cohen, P. (1975). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, Wiley, New York.

Cook, T. D. and Campbell, D. T. (1979). *Quasi-Experimentation*, Rand McNally, Chicago, Illinois.

Crandall, R. (1976). 'Validation of self-report measures using ratings by others', *Sociological Methods and Research*, **4**: 380–400.

Crandall, V. (1966). 'Personality characteristics and social achievement behaviors associated with children's social desirability response tendencies', *Journal of Personality and Social Psychology*, **4**: 447–486.

Cronbach, L. J. (1950). 'Further evidence on response set and test design', *Educational and Psychological Measurement*, **10**: 3–31.

Crowne, D. P. and Marlowe, D. (1960). 'A new scale of social desirability independent of psychopathology', *Journal of Consulting Psychology*, **24**: 344–354.
Crowne, D. P. and Marlowe, D. (1964). *The Approval Motive: Studies in Evaluative Dependence*, Wiley, New York.
Dickensberger, D., Holtz, S. and Gniech, G. (1978). 'Bedürfnis nach sozialer anerkennung: validierung der Marlowe-Crowne Social Desirability Scale über ein konzept individuell relevanter gruppen', *Diagnostica*, **24**: 24–38.
Dillman, D. A. (1978). *Mail and Telephone Surveys: The Total Design Method*, Wiley, New York.
Dohrenwend, B. S. (1969). 'Interviewer biasing effects: towards a reconciliation of findings', *Public Opinion Quarterly*, **33**: 121–125.
Dohrenwend, B. S., Colombotos, J. and Dohrenwend, B. P., (1968). 'Social distance and interviewer effects', *Public Opinion Quarterly*, **32**: 410–422.
Edgell, S. E., Himmelfarb, S. and Duchan, K. L. (1982). 'Validity of responses in a randomized response model', *Sociological Methods and Research*, **11**, 89–100.
Edwards, A. L. (1953). 'The relationship between the judged desirability of a trait and the probability that the trait will be endorsed', *Journal of Applied Psychology*, **37**: 90–93.
Edwards, A. L. (1957). *The Social Desirability Variable in Personality and Assessment and Research*, Dryden, New York.
Edwards, A. L. (1970). *The Measurement of Personality Traits by Scales and Inventories*, Holt, Reinhard and Winston, New York.
Edwards, A. L. and Diers, C. J. (1962). 'Social desirability and the factorial interpretation of the MMPI', *Educational and Psychological Measurement*, **22**: 501–509.
Ganster, D. C., Hennessey, H. W. and Luthans, F. (1983). 'Social desirability response effects: Three alternative models', *Academy of Management Journal*, **26**: 321–331.
Goffman, E. (1959). *The Presentation of Self in Everyday Life*, Doubleday, New York.
Goudy, W. J. and Potter, H. R. (1975). 'Interview rapport: demise of a concept, *Public Opinion Quarterly*, **39**: 529–543.
Gove, W. R. and Geerken, M. R. (1977). 'Response bias in surveys of mental health: an empirical investigation', *American Journal of Sociology*, **82**: 1289–1317.
Grabitz-Gniech, G. (1971). 'Bericht über eine analyse von sieven persönlichkeits-fragebogen'. Bericht aus der Sonderforschungsbereich 24, Universität Mannheim.
Greenwald, H. J. and Satow, Y. (1970). 'A short social desirability scale, *Psychological Reports*, **27**: 131–135.
Gur, R. C. and Sackeim, M. A. (1979). 'Self-deception: A concept in search of a phenomenon', *Journal of Personality and Social Psychology*, **37**: 147–169.
Hermans, H. (1967). *Motivatie en Prestatie*, Swets and Zeitlinger, Amsterdam, The Netherlands.
Himmelfarb, S. and Lickteig, C. (1982). Social desirability and the randomized response technique. *Journal of Personality and Social Psychology*, **43**: 710–717.
Humm, D. G. and Wadsworth, G. W. Jr (1939). *The Humm-Wadsworth Temperament Scale*, Doncaster G Humm, Los Angeles, California.
Hyman, H. H., Cobb, W. J., Feldman, J. J., Hart, C. W. and Stember, C. H. (1954). *Interviewing in Social Research*, University of Chicago Press, Chicago, Illinois.
Jackson, D. N. (1967). *Personality Research Form Manual*, Research Psychologists Press, New York.
Jones, E. E. and Sigall, H. (1971). 'The bogus pipeline: A new paradigm for measuring affect and attitude', *Psychological Bulletin*, **76**: 349–364.
Kane, J. S. and Lawler, E. E. III (1978). 'Methods of peer assessment', *Psychological Bulletin*, **85**: 555–586.
Klassen, D., Hornstra, R. K. and Anderson, P. B. (1976). 'Influence of social desirability on symptom and mood reporting in a community survey', *Journal of Consulting and Clinical Psychology*, **43**, 448–452.
Kuder, G. F. (1939). *Preference Record*, Science Research Associates, Chicago, Illinois.
Lana, R. E. (1969). 'Pretest Sensitization'. In: Rosenthal, R. and Rosnow, R. L. (Eds) *Artifact in Behavioral Research*, Academic Press, New York.
Larsen, K. S., Martin, H. J., Ettinger, R. H. and Nelson, J. (1976). 'Approval seeking, social cost, and agression: a scale and some dynamics', *Journal of Psychology*, **94**: 3–11.

Locander, W., Sudman, S. and Bradburn, N. M. (1976). 'An investigation of interview methods, threat, and response distortion', *Journal of the American Statistical Association*, **71**, 269–275.

Martin, H. J. and Greenstein, T. N. (1983). 'Individual differences in status generalization: Effects of need for social approval, anticipated interpersonal contact, and instrumental task abilities', *Journal of Personality and Social Psychology*, **45**, 641–662.

Martin, J. and Butcher, B. (1982). 'The quality of proxy information. Some results from a large-scale study', *The Statistician*, **31**: 293–319.

McCrae, R. R. (1982). 'Consensual validation of personality traits: Evidence from self-reports and ratings', *Journal of Personality and Social Psychology*, **43**: 293–303.

McCrae, R. R. and Costa, P. T. (1983). 'Social desirability scales: More substance than style', *Journal of Consulting and Clinical Psychology*, **51**: 882–888.

McGuire, W. J. (1969). 'Suspiciousness of experimenter's intent'. In: Rosenthal, R. and Rosnow, R. L. (Eds). *Artifact in Behavioral Research*, Academic Press, New York.

Meehl, P. E. and Hathaway, S. R. (1946). 'The K-factor as a suppressor variable in the MMPI', *Journal of Applied Psychology*, **30**, 525–564.

Messick, S. (1968). 'Response sets'. In: *International Encyclopedia of the Social Sciences*, **13**: 492–496.

Miller, J. P. (1981). 'Complexities of the randomized response solution', *American Sociological Review*, **46**: 928–930.

Millham, J. (1974). 'Two components of need of approval score and their relationship to cheating following success and failure', *Journal of Research in Personality*, **8**, 378–392.

Millham, J. and Kellogg, R. W. (1980). 'Need for social approval: Impression management of self-deception?' *Journal of Research in Personality*, **14**: 445–457.

Millham, J. and Jacobson, L. I. (1978). 'The need for social approval'. In: London, H. and Exner, J. E. (Eds) *Dimensions of Personality*, Wiley, New York.

Nederhof, A. J. (1981). *Beter Onderzoek. Bestrijding van Foutenbronnen in Sociaal-Wetenschappelijk Onderzoek*, VUGA, The Hague, The Netherlands.

Nederhof, A. J. (1983). 'The effects of material incentives in mail surveys: Two studies', *Public Opinion Quarterly*, **47**: 103–111.

Nederhof, A. J. (1984). 'Visibility of response as mediating factor in equity research', *Journal of Social Psychology*, **122**: 211–215.

Nederhof, A. J. (1985). *Artifact in Social Science Research: Coping With Sources of Error in Data Collection*, Irvington, New York.

Nederhof, A. J. and Zwier, A. G. (1983). 'The "crisis" in social psychology, an empirical approach', *European Journal of Social Psychology*, **13**: 255–280.

Nunnally, J. C. (1978). *Psychometric Theory*, McGraw Hill, New York.

Orne, M. T. (1973). 'Communication by the total experimental situation'. In: Pliner, P., Krames, L. and Alloway, T. (Eds) *Communication and Affect: Language and Thought*, Academic Press, New York.

Ostrom, T. M. (1973). 'The bogus pipeline: A new *ignis fatuus*?' *Psychological Bulletin*, **79**: 776–783.

Paulhus, D. L. (1981). 'Statistical control of socially desirable responding in personality inventories: Principal-factor deletion', *Journal of Research in Personality*, **15**: 383–388.

Paulhus, D. L. (1982). 'Individual differences, impression management, and cognitive dissonance: Their concurrent operation in forced compliance', *Journal of Personality and Social Psychology*, **43**: 838–852.

Paulhus, D. L. (1984). 'Two-component models of socially desirable responding', *Journal of Personality and Social Psychology*, **46**: 598–609.

Phillips, D. L. (1971). *Knowledge from What?* Rand McNally, Chicago.

Phillips, D. L. (1973). *Abandoning Method*, Jossey-Bass, San Francisco, California.

Phillips, D. L. and Clancy, K. J. (1970). 'Response biases in field studies of mental illness', *American Sociological Review*, **35**: 503–515.

Phillips, D. L. and Clancy, K. J. (1972). 'Some effects of "social desirability" in survey studies', *American Journal of Sociology*, **77**: 921–940.

Quigley-Fernandez, B. and Tedeschi, J. T. (1978). 'The bogus pipeline as lie detector: Two validity studies', *Journal of Personality and Social Psychology*, **36**: 247–256.

Ramanaiah, N. V. and Martin, H. J. (1980). 'On the two-dimensional nature of the

Marlowe–Crowne social desirability scale', *Journal of Personality Assessment*, **44:** 507–514.

Ries, M., Kalle, R. J. and Tedeschi, J. T. (1981). 'The bogus pipeline and attitude moderation following induced compliance: Misattribution of dissonance arousal or impression management inhibition?' *Journal of Social Psychology*, **115:** 247–258.

Rogers, T. F. (1976). 'Interviews by telephone and in person: quality of responses and field performance', *Public Opinion Quarterly*, **40:** 51–65.

Rorer, L. G. (1965). 'The great response-style myth', *Psychological Bulletin*, **63:** 129–156.

Rosenberg, M. J. (1969). 'The conditions and consequences of evaluation apprehension'. In: Rosenthal, R. and Rosnow, R. L. (Eds) *Artifact in Behavioral Research*, Academic Press, New York.

Rosenthal, R. (1976). *Experimenter Effects in Behavioral Research*, Irvington, New York.

Rosenthal, R. and Rosnow, R. L. (Eds (1969). *Artifact in Behavioral Research, Academic* Press, New York.

Rosenthal, R. and Rosnow, R. L. (1975). *The Volunteer Subject*, Wiley, New York.

Rosnow, R. L. and Aiken, L. S. (1973). 'Mediation of artifacts in behavioral research', *Journal of Experimental Social Psychology*, **9:** 181–201.

Sackeim, H. A. and Gur, R. C. (1978). 'Self-deception, self-confrontation, and consciousness'. In: Schwartz, G. I. and Shapiro, R. D. (Eds) *Consciousness and Self-regulation. Advances in Research and Theory*, Vol. 2, Plenum, New York.

Sackeim, H. A. and Gur, R. C. (1979). 'Self-deception, other-deception, and self-reported psychopathology', *Journal of Consulting and Clinical Psychology*, **47:** 213–215.

Schuessler, K. F. (1982). *Measuring Social Life Feelings*, Jossey-Bass, San Francisco.

Schuessler, K. F., Hittle, D. and Cardascia, J. (1978). 'Measuring responding desirability with attitude-opinion items', *Social Psychology*, **41:** 224–235.

Scott, W. A. (1963). 'Social desirability and individual conceptions of the desirable', *Journal of Abnormal and Social Psychology*, **67:** 574–585.

Scott, W. A. (1968). 'Attitude measurement'. In: Lindzey, G. and Aronson, E. (Eds) *The Handbook of Social Psychology*, Volume II, Addison Wesley, Reading, Massachusetts.

Shulman, A. D. and Silverman, I. (1974). 'Social desirability and need approval: some paradoxial data and a conceptual re-evaluation', *British Journal of Social and Clinical Psychology*, **13:** 27–32.

Sigall, H. and Page, R. (1971). 'Current stereotypes: A little fading, a little faking', *Journal of Personality and Social Psychology*, **18:** 247–255.

Soeken, K. L. and Macready, G. B. (1982). 'Respondents' perceived protection when using randomized response', *Psychological Bulletin*, **92:** 487–489.

Stocking, C. (1979). 'Re-interpreting the Marlowe–Crowne Scale'. In: Bradburn, N. M., Sudman, S. and Associates (Eds) *Improving Interview Methods and Questionnaire Design*, Jossey-Bass, San Francisco.

Strickland, B. R. (1977). 'Approval motivation'. In: Blass, T. (Ed.) *Personality Variables in Social Behavior*, Erlbaum, Hillsdale, New Jersey.

Sudman, S. and Bradburn, N. M. (1974). *Response Effects in Surveys*, Aldine, Chicago, Illinois.

Tracy, P. E. and Fox, J. A. (1981). 'The validity of randomized responses for sensitive measurements', *American Sociological Review*, **46:** 187–200.

Warner, S. L. (1965). 'Randomized response: A survey technique for eliminating evasive answer bias', *Journal of the American Statistical Association*, **60:** 63–69.

Weber, S. J. and Cook, T. D. (1972). 'Subject effects in laboratory research: an examination of subject roles, demand characteristics, and valid inference.' *Psychological Bulletin*, **77:** 273–295.

Weiss, C. H. (1968). 'Validity of welfare mother's interview response', *Public Opinion Quarterly*, **32:** 622–633.

Weiss, C. H. (1969). 'Interviewer biasing effects: toward a reconciliation of findings', *Public Opinion Quarterly*, **33:** 127–129.

Weiss, C. H. (1970). 'Interaction in the research interview: the effects of rapport on response'. In: *Proceedings of the Social Statistics Section, American Statistical Association*, 17–20.

Wicklund, R. A. (1975). 'Objective self-awareness'. In: Berkowitz, L. (Ed.) *Advances in Experimental Social Psychology*, Vol. 8. Academic Press, New York.
Wiggins, J. S. (1959). 'Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions', *Journal of Consulting Psychology*, **23:** 419–427.
Williams, J. A. (1968). 'Interviewer role performance: a further note on bias in the information interview', *Public Opinion Quarterly*, **32:** 287–294.
Williams, J. A. (1969). 'Interviewer biasing effects: toward a reconciliation of findings', *Public Opinion Quarterly*, **33:** 125–127.
Wise, S. L. (1980). 'An approach to sensitive questions: the randomized response technique'. Paper presented at the annual meetings of the American Educational Research Association, Boston, Massachusetts, April.
Wiseman, F. (1972). 'Methodological bias in public opinion surveys', *Public Opinion Quarterly*, **36:** 105–108.

RÉSUMÉ

La désirabilité sociale est une des sources les plus communes de biais qui affecte la validité des résultats de recherches expérimentales et d'enquêtes. Du point de vue de la présentation de soi, la désirabilité sociale peut être considérée comme la résultante de deux facteurs séparés: l'auto-duperie et l'hétéro-duperie. On distingue deux façons de tenir tête au biais de la désirabilité sociale. La première façon comprend deux méthodes destinées à la détection et à la mesure du biais de désirabilité sociale: l'emploi d'échelles de désirabilité sociale, et la cotation des items pour la désirabilité. Une seconde façon comprend sept méthodes pour prévenir ou réduire le biais de désirabilité sociale, qui incluent l'emploi d'items à choix forcés, la technique des résponses au hasard, le 'bogus pipeline', l'auto-administration du questionnaire, la sélection des interviewers, et l'utilisation de sujets mandatés. Aucune méthode ne réussit complètement et dans toutes les conditions à tenir tête au biais de désirabilité sociale tant du point de vue de l'auto-duperie que de l'hétéro-duperie. Le meilleur choix disponible est celui d'une combinaison de méthodes de prévention et de détection.

ZUSAMMENFASSUNG

Die soziale Wünschbarkeit ist eine der bekanntesten Störquellen, welche die Validität der experimentellen Forschung und der Fragebogenmethode beeinträchtigt. Vom Gesichtspunkt der Selbstvorstellung aus kann die soziale Wünschbarkeit als Resultat zweier getrennter Faktoren betrachtet werden: die Selbsttäuschung und die Täuschung anderer. Man unterscheidet zwei Arten, die Verzerrungen auf Grund der sozialen Wünschbarkeit zu kontrollieren. Zur ersten Art gehören zwei Methoden deren Ziel es ist, die Wünschbarkeitstendenz aufzudecken und zu messen; es handelt sich um die Anwendung von Skalen der sozialen Wünschbarkeit und die Bewertung der Items bezüglich der Wünschbarkeit. Zur zweiten Art gehören sieben Methoden, deren Ziel es ist, die Verzerrungen, die durch die Wünschbarkeitstendenz entstehen zu verkleinern oder auszuschliessen. Dazu gehören der Gebrauch von Zwangswahl-Items, die Technik der Zufalls-Antwort, die 'Bogus-Pipeline', das anonyme Ausfüllen des Fragebogen durch die Vpn, die strenge Selektion der Interviewer, die Befragung von autorisierten, stellvertretenden Vpn. Keine dieser Methoden ist absolut zuverlässig bei der Vermeidung von Selbst- und Hetero-Täuschung. Die best Art, das Problem der sozialen Wünschbarkeit anzugehen, ist ein kombinierte Anwendung von Aufdeckungs- und Präventivstrategien.