



Social Measurement through Social Surveys

An Applied Approach



Edited by
Martin Bulmer, Julie Gibbs and Laura Hyman

- Paxton, P. (1999), 'Is Social Capital Declining in the United States? A Multiple Indicator Assessment', *American Journal of Sociology* 105: 88–127.
- Paxton, P. (2007), 'Association Memberships and Generalized Trust: A Model of Generalized Trust in Thirty-One Countries', *Social Forces* 86, 1: 47–76.
- Phelps, E. (2004), 'Young Citizens and Changing Electoral Turnout, 1964–2001', *Political Quarterly* 75, 3: 238–49.
- Portes, A. (1998), 'Social Capital: Its Origins and Applications in Modern Sociology', *Annual Review of Sociology* 24: 1–24.
- Putnam, R. (1993), *Making Democracy Work: Civic Traditions in Modern Italy* (Princeton, NJ: Princeton University Press).
- Putnam, R. (2000), *Bowling Alone: The Collapse and Revival of American Community* (New York: Simon & Schuster).
- Putnam, R. (ed.) (2002), *Democracies in Flux: The Evolution of Social Capital in Contemporary Society* (Oxford: Oxford University Press).
- Putnam, R. (2004), 'Commentary: "Health by Association": Some Comments', *International Journal of Epidemiology* 33, 4: 667–71.
- Putnam, R. (2007), 'E Pluribus Unum: Diversity and Community in the Twenty-first Century, The 2006 Johan Skytte Prize Lecture', *Scandinavian Political Studies* 33, 2: 137–74.
- Savage, M., Li, Y. and Tampubolon, G. (2006), 'Rethinking the Politics of Social Capital: Challenging Tocquevillian Perspectives', in R. Edwards, J. Franklin and J. Holland (eds), *Assessing Social Capital: Concepts, Policy and Practice* (Newcastle: Cambridge Scholars Press), 70–94.
- Sorensen, E. and Torring, J. (2003), 'Network Politics, Political Capital and Democracy', *International Journal of Public Administration* 26, 6: 609–34.

Chapter 11

Measuring Social Attitudes

Caroline Roberts

Introduction

The measurement of attitudes is common practice in most major academic surveys, and several important large-scale studies are devoted entirely to the exercise, including the British Social Attitudes (BSA) survey, fielded annually since 1983, the US General Social Survey (GSS), biennially, since 1972, and the German 'Allbus', also biennial, since 1980. In addition to these long-standing national studies, several large-scale cross-national studies, including the European Commission's Eurobarometer surveys, the International Social Survey Programme (ISSP) and, most recently, the European Social Survey (ESS), have regularly collected data about social attitudes Europe- and world-wide for many years. The aim of each of these studies is to describe the way people think and feel about the various features of the world around them, and to make comparisons across different groups in society. The data are used extensively by students and academics across all the social sciences. Outside of academic research, politicians, civil servants, journalists, market researchers and advertising executives frequently use surveys to gauge public reactions to social policies, electoral candidates, political issues, new products and innovations, and these measurements play a critical role in decision making both in governance and business development alike.

Public attitudes are as much a part of social reality as are a society's demographic characteristics or behaviour patterns. Unlike demographic characteristics and behaviour patterns, however, attitudes are not immediately accessible to the researcher interested in measuring them. Because they cannot be observed directly, survey researchers typically rely on respondents' self-reports, which can be notoriously unstable, fluctuating from one measurement attempt to the next. In this chapter, I describe the most common methods used to elicit self-reports of attitudes and discuss the main sources of error that contribute to the instability of measurements. Errors in attitude measurements can distort sample estimates of marginal distributions and expected relationships between attitudes and other variables, leading to inaccurate conclusions about the state of public opinion. Throughout the chapter, I consider how researchers can minimize this error, to enhance the precision of their measurements. To understand some of the challenges involved in measuring attitudes, it is important to have a shared understanding of what attitudes are, so in the next section I provide some definitions.

What Are Attitudes?

The term 'attitude' has become so ubiquitous in our society as to hardly merit a formal definition. Yet though widely used and understood in everyday discourse, the concept is relatively technical and specific to the discipline of social psychology from which it emerged. Like any theoretical construct, attitudes have been conceptualized and researched in many different ways throughout their history. Current thinking in the field, however, generally agrees on the idea that attitudes are cognitive representations of a person's positive or negative evaluations of different 'objects' (for example, Eagly and Chaiken, 1993). People make evaluations about all kinds of things, including actual physical objects, consumer products, other people or groups of people, events, behaviours, political issues and policies, and it is this positive or negative judgment that scholars generally refer to as an attitude, and which is the focus of most measurement attempts. This focus on the evaluative component of attitudes is relatively recent, however. Earlier contributors to the field (for example, Allport, 1935; Katz and Stotland, 1959) conceived of attitudes more broadly, incorporating affective (the emotional reactions a person has towards the attitude object), behavioural (how they act in relationship to it) and cognitive (what they think about it) components into the attitude construct. In more recent definitions of attitudes, however, scholars have sought to distinguish the evaluative judgment implicit in attitudinal response from these other components, which are perhaps best viewed as different types of information on which the attitude is based (Fabrigar, Krosnick and MacDougall, 2005). In fact their relation is more interactive, as each one may influence the formation and transformation of the attitude, which in turn may have a reciprocal influence on people's emotional, behavioural and cognitive responses to the attitude object (Albaracin et al., 2005).

As far as the measurement of attitudes is concerned, perhaps the most significant defining feature of attitudes is that as psychological constructs, they cannot be observed directly. We cannot know exactly what an attitude 'looks like' in the mind of the attitude holder, and our measurements can only ever approximate what we are trying to describe – they are not the same thing as the attitudes themselves (Krosnick, Judd and Wittenbrink, 2005). This characteristic of attitudes has led to disagreement about the cognitive processes that underpin the expression of attitudes. In particular, disagreement exists among scholars about whether attitudes are better described as 'memories' or 'judgments' (Albaracin et al., 2005: 4). While some prefer to think of attitudes as relatively stable structures stored in a person's long-term memory (for example, Fazio, 1990) and capable of being retrieved and reported on request, a substantial body of evidence suggests that this is not the case for all kinds of attitudes (for example, Zaller and Feldman, 1992). In many situations, people do not have a ready-made single evaluation about a specific object, and instead they must construct one on the spot that summarizes whatever considerations were brought to mind when they were asked to report their attitude (Krosnick, Judd and Wittenbrink, 2005; Zaller and Feldman, 1992). As we shall see, different

considerations may be brought to mind in different contexts (some negative, some positive), so the judgment made, and consequently the attitude reported, can vary substantially over time and across situations. To complicate matters, characteristics of the context in which the attitude must be reported (including methods used by researchers to measure the attitude) may influence whether positive or negative considerations are brought to mind. In the next section, I describe some of the methods commonly used for measuring attitudes.

Ways of Measuring Attitudes

Given that attitudes cannot be observed directly, investigators have to make inferences about how positively or negatively a person feels about an object, based on the different kinds of information that they can observe. Psychologists have used various sources of observable information to find out about people's attitudes, including their physiological reactions to the attitude object (for example, Mueller, 1977; Rankin and Campbell, 1955; Hess, 1965), their behaviour in relation to the attitude object (for example, Lapiere, 1934; Milgram, Mann and Harter, 1965) and their psychological reactions to it (for example, Dovidio and Fazio, 1992; Greenwald, McGhee and Schwartz, 1998). The problem with each of these sources of information about attitudes is that they are not very practical to work with. They can be difficult to gain access to, hard to interpret and their relation to a person's underlying evaluation is not always as clear cut as it might be presumed to be (Fabrigar, Krosnick and MacDougall, 2005). These methods are also less helpful when the aim is to describe the attitudes of society as a whole. For these reasons, by far the most popular source of information that researchers use to draw inferences about attitudes are people's own descriptions of how they think and feel about a particular object. Asking people directly to report their attitudes provides a relatively quick, easy and cost-efficient way of gathering information about attitudes from a large group of people. In the context of a sample survey, using direct attitude measures allows researchers not only to draw inferences about the nature of respondents' attitudes at the individual level, but also to describe the distribution of attitudes towards a particular object in the population as a whole. For the remainder of the chapter, therefore, I focus primarily on issues surrounding the use of this type of attitude measure.

Researchers face a number of choices when designing questionnaires to elicit respondents' own descriptions of their underlying attitudes. Many of these decisions are guided by resource considerations. For example, one preliminary decision includes whether to use 'open' or 'closed' questions. Compared with open-ended questions, in which respondents' answers to questions are recorded verbatim, closed questions, in which respondents are asked to select their answer from a set of predefined alternatives, are considerably more practical for researchers to work with and usually quicker and easier for respondents to answer (Lazarsfeld, 1944). Closed questions allow large quantities of data to be collected more efficiently

from large numbers of people and, as a result, they are used far more frequently by researchers nowadays. However, resource considerations are not the only guiding factor in the design of attitude measures. Concerns about costs must be weighed up against the need to collect data that provide the most accurate description possible of the underlying attitude. Research (for example, by Schuman and Presser, 1981) has shown that open questions tend to produce more diverse answers compared with closed questions and that this can affect the relationship observed between attitudes and respondents' background characteristics. If researchers want to be confident in their conclusions about the attitudes of particular groups in society, therefore, they must be careful to use the methods that will most accurately reflect the real differences present in the population.

Another decision that researchers face is whether to use one or more questions to measure a single attitude. Most direct closed-ended self-report methods used today are based on techniques developed by some of the earliest contributors to the field, including Thurstone's (1928) 'Equal Appearing Intervals' method, Likert's (1932) method of Summated Ratings, and Osgood, Suci and Tannenbaum's (1957) Semantic Differential (Krosnick, Judd and Wittenbrink, 2005). These approaches provide different ways of measuring attitudes using multiple items designed to work together to measure a specific attitude object. In Thurstone's approach, a set of statements is selected to represent the full range of possible attitudes to a given object (from extremely negative to extremely positive). These are then rated by a panel of judges using scores ranging (typically) from 0 to 11, and the two items best representing each point on the scale are then selected to construct the final measure to be administered to respondents. Respondents must indicate whether they agree or disagree with each statement in the final scale and their score on the scale is the mean of the scores that the judges gave to each statement that the respondent agreed with. In Likert's approach, positive and negative statements about the object are selected and respondents must report the extent of their agreement or disagreement with each one using five-point response scales. For each rating they obtain a score ranging from 1 (representing an extremely negative attitude toward the object) to 5 (representing an extremely positive attitude), so that a total score for all items can be computed (as the sum or mean of scores on each item). Correlations between specific items and the overall scale score are then examined, so that items with low correlations (perhaps because they are measuring something other than the attitude of interest) can be removed from the scale and the total score is recalculated. Both techniques have been proven to be effective at obtaining accurate descriptions of attitudes, but in general, researchers have found Likert's procedure to be less labour-intensive (Edwards and Kenney, 1946), which is why Likert scales are among the most widely used methods of attitude measurement today. Nevertheless, both procedures require considerable preparatory work, which is why the less laborious 'semantic differential' technique has proved to be a popular alternative (Osgood, Suci and Tannenbaum, 1957). This involves asking respondents to rate the object of interest on a relatively small number of seven-point response scales, anchored at either end by opposing

adjectives (such as good and bad, positive and negative), which Osgood and his colleagues' previous research showed capture the key features of the evaluative dimension of an object's meaning.

Given the high costs involved in collecting data on attitudes in social surveys, and pressures to keep questionnaire length to a minimum, it can be difficult to justify including large numbers of items to measure a single attitude. For this reason, most researchers nowadays tend to use just a few closed questions to assess attitudes in each domain of interest, and frequently just a single item. While this has advantages in terms of cost efficiency, it means that researchers have less information on which to base their descriptions of people's attitudes. This can be problematic because attitude measurements – like any other type of measurement – contain a certain amount of error. A person's score on a given measure consists not only of their 'true score' on the underlying construct of interest, but also of error associated with the method of measurement. Two kinds of error can affect the quality of the measurements obtained: random error, the effects of which are evenly distributed across the data collected; and systematic error, the effect of which is to bias results in a specific direction. Researchers use formal criteria to determine the precision of questions designed to measure attitudes directly, and the extent to which they contain different types of error. These criteria include whether the question measures what it is intended to measure, that is, the construct of interest, and not some source of systematic error (Krosnick, Judd and Wittenbrink, 2005); and whether the same measurement would be obtained using the same question on repeated occasions. The first of these criteria is referred to as the 'validity' of the measure and the second is referred to as its 'reliability' (see Bohrnstedt, 1977 for an introduction to measurement theory). To a certain extent, the greater the validity of a measure, the more likely it is to produce the same answers on repeated occasions. However, because attitudes cannot be observed directly, it is especially difficult for researchers to be sure that they are measuring the construct of interest, and this can lead them to obtain highly variable results from one measurement attempt to the next. To complicate matters, variability in responses to direct attitude measures may not always result from errors associated with the method of measurement, but also from the way in which people respond to direct attitude measures. These sources of variability in attitude measurements are the focus of the remainder of the chapter. I begin by looking at how the cognitive processes involved in responding to direct measures of attitudes can influence the results obtained.

Sources of Variability in Attitude Measurements

Cognitive Processes in Reporting Attitudes

Traditional approaches to the study of attitudes view them as relatively enduring, pre-existing constructs stored in long-term memory (see, for example, Fazio et

al., 1986; Fazio, 1990). However, a substantial body of research suggests that it is inappropriate to view attitudes in this way, in particular because measurements of attitudes have been shown to be highly unstable across different measurement attempts. One explanation for this suggests that shifts in responses to attitude measures are essentially random, and occur because people do not hold pre-existing attitudes at all on certain topics and lack the political sophistication required to link complex issues and ideas to form consistent attitudes (Converse, 1964). To avoid admitting that they are insufficiently informed about an issue to have formed an opinion about it, respondents pick one of the response alternatives seemingly at random or at least to concoct a response 'on the spot, based on little or no information' (Fabrigar, Krosnick and MacDougall, 2005: 25). Because these attitude reports do not reflect pre-existing attitudes towards an issue, and lack coherence (meaning they do not relate in expected ways with other variables measured in the same survey) they have been referred to as 'nonattitudes' (Converse, 1970).

Evidence for the existence of nonattitudes comes from studies that have tested the effect of offering versus omitting a 'Don't Know' response alternative in an attitude measure (for example, Krosnick et al., 2002). When the 'Don't Know' option is included, a significantly larger number of respondents will report that they do not know their attitudes than when the option is omitted. This finding suggests that when the 'Don't Know' is not legitimated as an explicit response category, people who are uncertain of their attitude prefer to select a substantive answer than admit to not knowing. This finding is supported by a number of studies that have deliberately asked respondents about *fictional* issues, to see whether they are still willing to avoid admitting that they do not know their attitude, even when they cannot hold a pre-existing view about the object in question (see, for example, Bishop et al., 1980; Schuman and Presser, 1980). These studies find that large proportions of respondents are willing to report attitudes on topics they know nothing about, so by extension it seems likely that they will report nonattitudes in other contexts too (see Smith, 1985 for a review).

Not only can attitude reports vary randomly across different surveys, but even over the course of a single survey interview can people's attitudes appear to change quite dramatically, depending on different features of the context in which they are measured (Tourangeau and Galesic, 2008). One feature that contributes to this measurement context includes the order in which questions about attitudes are presented in the questionnaire. 'Context effects' occur when respondents modify their responses to survey questions either to make them appear more consistent with responses given earlier in the questionnaire, as in assimilation or 'carryover' effects, or when they interpret later questions as requesting new information that may appear logically inconsistent with earlier reports, as in contrast or 'backfire' effects (Tourangeau and Rasinski, 1988). The presence of context effects in survey data further challenges the idea that respondents in attitude surveys simply report a pre-existing attitude retrieved from their long-term memory (Tourangeau and Galesic, 2008), though the nature of such effects suggests that they are not simply

picking answers randomly either. Rather, they show how sensitive attitude reports are to salient features of the measurement setting, lending support to the conclusion that attitudes are online judgments constructed on the basis of considerations brought to mind during questionnaire completion. It is the constructive process of producing an attitude report that explains variability observed in attitude reports.

Tourangeau, Rips and Rasinski (2000) have developed a model for understanding the processes involved in formulating responses to attitude questions in surveys, which is helpful for understanding how context effects and other types of errors arise in survey data. According to the model, answering a survey question involves four components of processing: the first component in the model is *comprehension*: in order to answer a survey question, a respondent must first understand what the question is asking; the second component is *retrieval*: having established the goal of the measurement attempt, the respondent must search his/her memory for any information relevant to the question that will help them to answer it. This process might involve retrieving an existing stored evaluation, or alternatively, if there is no pre-existing attitude, it might involve retrieving various stored 'considerations' associated with that object (including, for example, relevant beliefs, feelings and memories). The third component of the response process involves *judgment* – a process of generating a summary evaluation based on the considerations brought to mind during retrieval; and lastly, in the fourth component of processing, *response*, the respondent must decide how to report their answer by mapping the judgment onto one of the available response alternatives. Each component involves multiple cognitive processes, making formulating a response to a survey question a surprisingly demanding procedure. Furthermore, each component is prone to error, which can affect the accuracy of the reported attitude (Tourangeau and Rasinski, 1988; Tourangeau, Rips and Rasinski, 2000), although it is the retrieval and judgment phases in particular that provide the key to understanding inconsistencies often observed in attitude measurements (Tourangeau and Galesic, 2008).

People typically store in memory a large number of both positive and negative associations with different attitude objects. Even if they have never had to integrate those associations into a single summary evaluation of the object, they will still have a mix of different considerations that can be brought to mind on different occasions and from which they can construct an attitude (Zaller and Feldman, 1992). Because the pool of considerations from which respondents to attitude surveys can draw is typically mixed, the information retrieved at any one time can lead to very different judgments from one occasion to the next. According to Zaller and Feldman's model, 'individuals answer survey questions by averaging across the considerations that happen to be salient at the moment of response' (1992: 586). When confronted with an attitude measure, respondents draw a sample of considerations based on how accessible they are in memory and base the judgment they report on the composition of that sample. Samples containing more positive considerations will lead to more positive evaluations, whereas samples containing more negative considerations will lead to more negative evaluations. Thus, it is

the process of retrieval, by which a sample of accessible considerations is drawn (typically including salient features of the questionnaire and data collection context), that determines the valence of the reported judgment.

Response instability in attitude measurement has been found to be more common among respondents with little prior knowledge about or involvement in the attitude issue. This may be because people who have not thought much about the attitude object have a smaller pool of considerations from which to sample, and possibly only one consideration that is immediately accessible when the attitude must be reported (Zaller and Feldman, 1992). Another explanation is that the degree to which a person is involved in an attitude object influences the amount of effort they are willing and able to expend in generating answers to attitude measures (for example, Petty and Cacioppo, 1986). Given that random error in attitude reports can affect conclusions about the distribution of attitudes in the population, as well as the relationships between attitudes and other constructs, it is important to be able to distinguish people with more 'crystallized' attitudes from others. Crystallized attitudes have been shown to be more resistant to persuasion, and therefore less labile over time (Petty and Krosnick, 1995). This highlights the value of gathering information about the strength of reported attitudes at the same time as measuring the attitude itself (such as the certainty with which it is held, the importance and centrality of the attitude to respondents, how much knowledge respondents have about the object and how interested they are in it -- see Krosnick and Petty, 1995), so that researchers have additional data on which to judge the accuracy of their measurements (Krosnick, Judd and Wittenbrink, 2005).

The Design of the Questionnaire

To minimize the amount of error in attitude reports, it is important for researchers to be aware of how certain types of biases can arise as a result of decisions about how to design attitude questionnaires. Such decisions include how to word attitude measures; how long response scales should be; how to label the response scale; the order in which response categories should be presented; whether to include middle or 'Don't Know' alternatives in the response scale; as well as how to order items in the questionnaire. An accumulated body of research using split-ballot experiments, in which different versions of the same question are administered to a random sub-sample of the total survey sample (see, for example, Rugg and Cantril, 1944; Kalton, Collins and Brook, 1978; Bishop, 1987; Schuman and Presser, 1981) has shown that the outcome of these choices can have important implications for the quality of the data collected, because the answers people give to questions asking about their attitudes vary depending on how the questions are asked. It is beyond the scope of this chapter to extensively review findings from this research, and excellent summaries of what is currently known about best practice in questionnaire design are available elsewhere (see Krosnick, Judd and Wittenbrink, 2005; Fabrigar, Krosnick and MacDougall, 2005; Fabrigar and Krosnick, 1995; see also Krosnick and Fabrigar, forthcoming, for a detailed

treatment of this literature). However, given the impact that they can have on the accuracy of our conclusions about the nature of social attitudes, it is helpful to provide some illustrations of how decisions about the design of attitude measures can affect the quality of the data collected.

Media opinion polls often attempt to describe public attitudes using data from just a single item, sometimes in ways that can misrepresent the true distribution of attitudes in the population. The difficulty for analysts (and policy makers) is deciding what the true distribution of attitudes is, based on data collected from different attitude measures. For example, Rugg and Cantril (1944) presented data from studies in which they manipulated the wording of opinion poll questions about US 'interventionism' during World War II to see how the changes affected respondents' answers. They found wide variation in support for intervention depending on how the question was worded, ranging from just 8 per cent answering 'Yes' to the question 'Should the US go into the war now and send an army to Europe?', to 76 per cent saying 'Yes' when the question was worded as: 'Some people say that if the US goes on helping England, Germany may start a war against our country. Do you think we should continue to help England, even if we run this risk?' Kalton and his colleagues (1978) reported similar fluctuations in response depending on whether one or both sides of an issue were presented in a question. Furthermore, in both studies, not only was the comparability of marginal distributions affected across different forms of the same question, but so too were the correlations between related variables. For this reason, many scholars urge caution when drawing conclusions about the state of public opinion based on single-item attitude measures. Other pitfalls to avoid in the wording of attitude measures include unbalanced question formats (Schaeffer et al., 2005), questions containing persuasive arguments (Bishop, Oldendick and Tuchfarber, 1982) and ambiguities in wording, such as so-called 'double-barreled' statements, where it may be unclear which part of the question the respondent should answer (see Oppenheim, 1992: 128) for a list of similar 'rules' for writing attitude measures).

Deciding on the best wording represents just one half of the challenge involved in constructing effective attitude measures. Having written an appropriate question, researchers must decide how respondents should report their answer. Extensive research into questions such as how many points to include on a response scale (see, for example, Jacoby and Matell, 1971; Lehmann and Hulbert, 1972; Masters, 1974; Cox, 1980; Alwin, 1997; Krosnick and Fabrigar, 1997) and how to label scale points (for example, Hakeel, 1968; Chase, 1969; Bradburn and Miles, 1979) have revealed significant differences in reliability and validity resulting from the decisions researchers make. Social surveys still vary in terms of the types of response scales they use. For example, while the BSA, in common with other studies, includes many items with five-point, Likert-type response scales (fully labeled from 'Disagree strongly', 'Disagree', 'Neither agree nor disagree', 'Agree', and 'Agree strongly'), both the GSS and the American National Election Studies (ANES) still ask respondents to report liking and disliking of some objects (for example, different groups; electoral candidates) on so-called 'feeling thermometers'

(101-point scales). In contrast, experiments conducted in the context of the ESS have found that in a cross-national context, 11-point semantic differential scales produce higher reliability and validity coefficients than other response formats (for example, Saris, van der Veld and Gallhofer, 2004). Meanwhile, a consensus based on numerous empirical studies has emerged regarding best practice in national studies, agreeing that the optimal length for a response scale is between five- and seven-point depending on whether the construct being rated is unipolar or bipolar (Krosnick and Fabrigar, forthcoming). For bipolar constructs like attitudes, seven-point scales have been found to produce maximal reliability and validity and to be easier to administer (Krosnick and Fabrigar, 1997), particularly where a 'branching' method is used (Malhotra, Krosnick and Thomas, 2009), in which respondents are first asked to indicate the direction they lean in (negative, neutral or positive) and then asked in a follow-up question to indicate the intensity of their leaning (strongly, moderately or slightly negative or positive). Furthermore, providing verbal labels for all scale points 'that have relatively precise meanings for respondents that reflect equal intervals along the continuum of interest' has been shown to enhance both reliability and validity (Krosnick and Fabrigar, 1997: 152).

In the psychometric literature, it has long been recognized that the form of a question can produce particular patterns of answers irrespective of question content, referred to as 'response sets' (Cronbach, 1946). A classic example of a response set associated with questions offering a simple choice between a positive and negative response (such as yes/no, true/false, agree/disagree) is 'acquiescence': the tendency to over-prefer the positive option (see Cronbach, 1946 for other examples). Response sets have been distinguished in the literature from 'response styles' (Rorer, 1965; Fabrigar and Krosnick, 1995), patterns of responding attributed both to respondent personality traits and motivation to respond in biased ways (for example, deliberately guessing in a multiple choice task). As we shall see in the next section, however, this distinction is unhelpful given the complex mix of influences that can lead to the various 'response effects', or errors, often observed in attitude data.

Factors Affecting the Response Process

So far in this chapter I have described two principal sources of variability in attitude measurements: (1) the cognitive processes by which respondents generate their evaluations of objects; and (2) the design of attitude measures themselves. In this section, I consider a third source of error in attitude measurements: factors that affect the response behaviour of survey respondents. Completing each of the processes described in Tourangeau and his colleagues' (2000) model of the response process might be thought of as the 'optimal' way of constructing an attitude report in response to a survey question (Krosnick, 1991). Even with

¹ These conclusions are based on research findings from surveys conducted in the US.

measures developed using best practice principles of questionnaire design, there are no guarantees that the measurements that researchers obtain will be accurate descriptions of the respondents' underlying attitudes. For the response process to be executed in an optimal way, two conditions must be met. First, respondents need to be both willing and able to exert the necessary effort to generate an attitude report that describes their underlying evaluation in a meaningful way and that can be mapped onto the available response alternatives. Second, respondents must be both willing and able to report their attitudes to researchers truthfully. Meeting these two conditions depends not only on the design of the questions themselves, but also on characteristics of the respondents and characteristics of the measurement setting. When these conditions are not met, response effects can result, lowering the overall quality of the measurements obtained. In this section, I consider the types of response effects that can arise as a result of different factors influencing how respondents execute the survey response process.

Factors Affecting Processing Effort

The process of generating an answer to an attitude measure can place considerable cognitive demands on survey respondents. Even if the respondents' attitudes on a particular topic are relatively crystallized, simply comprehending the task of the measurement exercise, retrieving the relevant evaluation from memory and mapping it onto the most appropriate response category can present a challenge in itself. However, the demands of the exercise increase when the focus of the question is on an attitude object relatively unfamiliar to respondents, because they must first decide what their attitude is, and then decide how (and whether) to report it. To execute each of these processes systematically, respondents need to be both willing and able to devote the time and effort needed to do so. When motivation and ability are low, some respondents may be tempted to take shortcuts in the necessary processing to arrive at an, if not optimal, then adequate response to a question in a shorter amount of time. This shortcutting process has been called 'questionnaire satisficing' (Krosnick, 1991). Satisficing either takes the form of completing each stage of the response process less thoroughly than needed (called 'weak satisficing') or skipping stages altogether (called 'strong satisficing'). Krosnick (1991) classifies different kinds of response effects often observed in attitude measurements according to this typology. Examples of weak satisficing include acquiescence (for example, Knowles and Condon, 1999), in which respondents agree with assertions made by the interviewer (because this requires less cognitive effort than generating reasons to disagree with the statement) and response order effects (Krosnick and Alwin, 1987), in which respondents select the most accessible satisfactory response alternative from the list provided. Examples of strong satisficing include over-endorsement of the 'Don't Know' option, and 'non-differentiation', in which respondents rate a battery of attitude statements using the same point on the response scale. In both cases, answers are selected based on little or no thought.

The likelihood that a respondent will adopt a satisficing response strategy is determined by variables influencing motivation and ability to process and the difficulty of the response task. Motivation and ability to expend the effort to process both vary by individual. Some people enjoy thinking and problem-solving more than others (Cacioppo and Petty, 1982), and natural cognitive abilities, level of education and knowledge of the survey topic are all important determinants of the ability to systematically complete the different stages of the response process. Consistent with this, satisficing effects have been found to be more common among respondents with low levels of education (for example, Narayan and Krosnick, 1996). The previous section considered how the difficulty of the response task varies with the design of the attitude measure and this, in turn, can influence the cognitive demands of generating an answer. For example, ambiguous question wording, sub-optimal scale length, or vaguely labeled response alternatives can all affect the ease with which respondents comprehend what the question is asking and map their attitude judgment onto available response alternatives. However, other factors present in the measurement setting can also influence the difficulty of the response task, including the mode of data collection. Survey interviews conducted by telephone appear to place considerably greater burden on the respondent's processing capacity because they tend to be conducted at a faster pace, and the respondent must attempt to hold both the question and available response alternatives in working memory whilst thinking about their answer. By contrast, in face-to-face interviews and self-completion surveys, respondents can make use of visual cues to aid comprehension and take their time over formulating a response. Consistent with this, response effects associated with satisficing have been shown to be more common in telephone data (Holbrook, Green and Krosnick, 2003).

Factors Affecting Willingness to Report Attitudes Truthfully

One of the most difficult challenges associated with asking people directly to report their attitudes is that they may not always be willing to reveal their true preferences to the researcher. In these situations, respondents are assumed to adapt their answer to the attitude measure to avoid embarrassment (Tourangeau and Smith, 1998), to portray themselves in an otherwise more favourable light, or to give answers that they think the researcher will want to hear (Sudman and Bradburn, 1974; Bradburn et al., 1978). This tendency is referred to as 'social desirability bias' and it can manifest itself in different ways, such as the over-reporting of socially desirable behaviours and the under-reporting of socially undesirable ones. The problem is especially pernicious in the context of attitude measurement, however, because unlike with behavioural measures, which can sometimes be cross-validated against external records, it is particularly difficult to determine to what extent measures of attitudes are affected by the bias. Nevertheless, many different kinds of studies have documented the presence of the bias in attitudinal reports, including 'randomized response' studies, in which

interviewers are unaware which of the randomly assigned questions respondents are answering (Warner, 1965), and 'bogus pipeline' studies, where respondents are led to believe that the researcher can discover their true response using a device akin to a lie detector (see Roese and Jamieson, 1993 for a review). Perhaps the most researched and best understood example of how social desirability bias affects attitude measurement and, consequently, the conclusions that researchers draw from surveys about the nature of social attitudes, comes from research into white racial attitudes (for example, Schuman and Converse, 1971; Krysan, 1998). Both the GSS and the ANES have been measuring white attitudes towards different racial groups in the US for several decades. Using data from these surveys, analysts have concluded that white respondents' evaluations of African Americans have become increasingly tolerant over time. However, data from alternative measures of racial prejudice do not support this conclusion, which suggests that rather than the attitudes themselves, it is the social norms surrounding the public expression of those attitudes that have changed during this period. Whereas white respondents participating in these surveys 30 years ago seemed comfortable endorsing negative stereotypes about blacks, it has become increasingly unacceptable for them to do so in more recent years.

Researchers understand well the factors that influence the likelihood of respondents answering attitude measures truthfully, and methods of reducing the risk of social desirability bias are routinely used in social surveys. As with questionnaire satisficing, the propensity to modify attitude reports in line with social desirability concerns appears to be influenced by a mix of respondent characteristics, the design of survey questions, and characteristics of the reporting situation. Some personalities high in 'need for social approval' may be more inclined towards socially desirable responding than others and, indeed, scales have often been included in personality and clinical inventories to control for the bias (Crowne and Marlowe, 1960). Insofar as attitude topics vary by how sensitive they are, different attitude measures will vary in the extent to which they elicit social desirability bias, though there is some disagreement about whether this is a function of the social norms governing the acceptability of certain types of attitudes, or shared values among different groups about what is desirable (DeMaio, 1984). By far the most important factor influencing respondent willingness to report attitudes truthfully is the extent to which the data collection method provides adequate privacy. The more anonymous the respondent perceives their answers to be, the more likely they are to answer honestly (Tourangeau, Rips and Rasinski, 2000). This has been demonstrated in several studies comparing responses given by respondents assigned to different modes of data collection, where modes providing greater anonymity obtained more accurate reports of sensitive behaviours (Weisberg, 2005). Thus, self-administered modes including paper questionnaires, web surveys and computer-assisted self-interviewing, either with an audio component (ACASI) or without (CASI), have been shown to be more effective at reducing social desirability

bias than interviewer-administered modes (for example, Sudman and Bradburn, 1974; Tourangeau, Rips and Rasinski, 2000).

Enhancing the privacy of the reporting situation can help to minimize errors associated with direct attitude measures arising from respondents' social desirability concerns, but it may not eliminate it altogether (Tourangeau and Smith, 1998). To tackle the problem, scientists have developed a variety of alternative procedures that aim to measure attitudes indirectly, while respondents remain unaware of the true goal of the measurement exercise. Perhaps the most promising of these, and most amenable to being used in sample surveys, are methods that measure so-called 'implicit' attitudes (for example, Wittenbrink, Judd and Park, 1997). Implicit attitudes are the automatic evaluative responses people have to attitude objects, which they may not even be conscious of (Dovidio and Fazio, 1992). These methods involve presenting research participants with a stimulus and assessing the extent of attitude activation by measuring how long it takes participants to respond to the stimulus. One of the most widely tested examples of such a method is the Implicit Association Test (IAT) (Greenwald, McGhee and Schwartz, 1998), which is designed to test the relative strength of associations between pairs of objects (for example, Senator McCain and Senator Obama) and positive and negative evaluations (good, bad). Faster reaction times when pairing objects with evaluations are said to indicate stronger associations in memory, and hence the extent of attitude activation. Despite their increasing popularity, there has been some scepticism about the validity and reliability of such techniques (for example, Cunningham, Preacher and Banaji, 2001), not helped by the fact that until relatively recently they had only been tested on volunteer samples. However, thanks to advances in data collection methodology, it is now possible to incorporate implicit measures of attitudes in major sample surveys (for example, the 2008 ANES used both the IAT and the Affect Misattribution Procedure (Payne et al., 2005) to assess implicit racial attitudes and candidate liking during the 2008 presidential campaign), expanding the range of methods available to researchers seeking to minimize errors in attitude measurements caused by social desirability bias.

Summary

In this chapter I described the methods most commonly used to measure social attitudes. Because attitudes cannot be observed directly, survey researchers mostly rely on respondents' own descriptions of how they think and feel about different features of the world around them. The aim of measurement is to simplify the complexity of these thoughts and feelings and to describe the underlying attitudes precisely using rigorous methods (Fabrigar, Krosnick and MacDougall, 2005). Measurements of attitudes inevitably contain a certain amount of error affecting the extent to which they accurately capture the construct of interest and their replicability over time. To describe the distribution of attitudes in the population

as accurately as possible, it is essential that researchers attempt to minimize the amount of error in their measurements. Several sources of error were described here, including errors arising from the cognitive processes involved in generating attitude judgments in response to survey questions; errors arising from the design of the questionnaire; and errors arising from factors influencing response behaviour. Distinguishing between these different sources of error in this way suggests a somewhat artificial distinction between related influences on respondents' answers. In reality, the quality of attitude measurements, and the conclusions that researchers draw from them about the nature of social attitudes, depend on a complex interaction between characteristics of the respondent, the questionnaire and the data collection context. A wealth of evidence about the impact of these combined influences has led to the development of best practice guidelines about how to enhance the accuracy of attitude measurements. The responsibility now lies with researchers involved in designing attitude measures for social surveys not only to follow these guidelines, but also to exploit opportunities to advance and refine them.

References

- Albarracín, D., Johnson, B.T., Zanna, M.P. and Tarcán Kumkale, G. (2005), 'Attitudes: Introduction and Scope', in D. Albarracín, B.T. Johnson and M.P. Zanna (eds), *The Handbook of Attitudes* (Mahwah, NJ: Lawrence Erlbaum Associates), 3–19.
- Allport, G.W. (1935), 'Attitudes', in C. Murchison (ed.), *Handbook of Social Psychology* (Worcester, MA: Clark University Press), 798–884.
- Alwin, D.F. (1997), 'Feeling Thermometers Versus 7-point Scales: Which Are Better?' *Sociological Methods and Research* 24, 3: 318–40.
- Bishop, G.F. (1987), 'Experiments with the Middle Response Alternative in Survey Questions', *Public Opinion Quarterly* 51, 2: 220–32.
- Bishop, G.F., Oldendick, R.W. and Tuchfarber, A.J. (1982), 'Effects of Presenting One Versus Two Sides of an Issue in Survey Questions', *Public Opinion Quarterly* 46, 1: 69–85.
- Bishop, G.F., Oldendick, R.W., Tuchfarber, A.J. and Bennett, S.E. (1980), 'Pseudo-opinions on Public Affairs', *Public Opinion Quarterly* 44, 2: 198–209.
- Bohrstedt, G.W. (1977), 'Reliability and Validity Assessment in Attitude Measurement', in G.F. Summers (ed.), *Attitude Measurement* (London: Kershaw Publishing Company), 80–99.
- Bradburn, N.M. and Miles, C. (1979), 'Vague Quantifiers', *Public Opinion Quarterly* 43, 1: 92–101.
- Bradburn, N.M., Sudman, S., Blair, E. and Stocking, C. (1978), 'Question Threat and Response Bias', *Public Opinion Quarterly* 42, 2: 221–34.
- Cacioppo, J.T. and Petty, R.E. (1982), 'The Need for Cognition', *Journal of Personality and Social Psychology* 42, 1: 116–31.

- Chase, C.I. (1969), 'Often Is Where You Find It', *American Psychologist* 24: 1043.
- Converse, P. (1964), 'The Nature of Belief Systems in Mass Publics', in D.E. Apter (ed.), *Ideology and Discontent* (New York: Free Press).
- Converse, P. (1970), 'Attitudes and Non-attitudes: Continuation of a Dialogue', in E.R. Tufté (ed.), *The Quantitative Analysis of Social Problems* (Boston: Addison-Wesley).
- Cox II, E.P. (1980), 'The Optimal Number of Response Alternatives for a Scale: A Review', *Journal of Marketing Research* XVII: 407-22.
- Cronbach, L.J. (1946), 'Response Sets and Test Validity', *Educational and Psychological Measurement* 6: 475-94.
- Crowne, D.P. and Marlowe, D. (1960), 'A New Scale of Social Desirability Independent of Psychopathology', *Journal of Consulting Psychology* 24: 349-54.
- Cunningham, W.A., Preacher, K.J. and Banaji, M.R. (2001), 'Implicit Attitude Measures: Consistency, Stability and Convergent Validity', *Psychological Science* 12, 2: 163-70.
- DeMaio, T.J. (1984), 'Social Desirability and Survey Measurement: A Review', in C.F. Turner and E. Martin (eds), *Surveying Subjective Phenomena*, Vol. 2 (New York: Russell Sage Foundation), 257-82.
- Dovidio, J.F. and Fazio, R.H. (1992), 'New Technologies for the Direct and Indirect Assessment of Attitudes', in J. Tanur (ed.), *Questions About Questions: Inquiries into the Cognitive Bases of Surveys* (New York: Russell Sage Foundation), 204-37.
- Eagly, A.H. and Chaiken, S. (1993), *The Psychology of Attitudes* (Orlando, FL: Harcourt Brace Jovanovich).
- Edwards, A.L. and Kenney, K.C. (1946), 'A Comparison of the Thurstone and Likert Techniques of Attitude Scale Construction', *Journal of Applied Psychology* 30: 72-83.
- Fabrigar, L.R. and Krosnick, J.A. (1995), 'Attitude Measurement and Questionnaire Design', in A.S.R. Manstead and M. Hewstone (eds), *Blackwell Encyclopedia of Social Psychology* (Oxford: Blackwell Publishers).
- Fabrigar, L.R., Krosnick, J.A. and MacDougall, B.L. (2005), 'Attitude Measurement: Techniques for Measuring the Unobservable', in T.C. Brock and M.C. Green (eds), *Persuasion: Psychological Insights and Perspectives* (2nd edition) (Thousand Oaks, CA: Sage), 17-40.
- Fazio, R., Sanbonmatsu, D., Powell, M. and Kardes, F. (1986), 'On the Automatic Activation of Attitudes', *Journal of Personality and Social Psychology* 37: 229-38.
- Fazio, R.H. (1990), 'Multiple Processes By Which Attitudes Guide Behaviour: The MODE Model as an Integrative Framework', in M. Zanna (ed.), *Advances in Experimental Social Psychology*, Vol. 23 ed. (San Diego: Academic Press), 75-109.

- Fishbein, M. (1967), 'A Consideration of Beliefs, and Their Role in Attitude', in M. Fishbein (ed.), *Readings in Attitude Theory and Measurement* (New York: John Wiley & Sons), 257-66.
- Fishbein, M. and Ajzen, I. (1975), *Belief, Attitude, Intention and Behaviour: An Introduction to Theory and Research* (Reading, MA: Addison-Wesley).
- Greenwald, A.G., McGhee, D.E. and Schwartz, J.L.K. (1998), 'Measuring Individual Differences in Implicit Cognition: The Implicit Association Test', *Journal of Personality and Social Psychology* 74, 6: 1464-80.
- Hakel, M.D. (1968), 'How Often is Often?' *American Psychologist* 23: 533-4.
- Hess, E.H. (1965), 'Attitude and Pupil Size', *Scientific American* 212: 46-54.
- Holbrook, A.L., Green, M.C. and Krosnick, J.A. (2003), 'Telephone vs. Face-to-face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias', *Public Opinion Quarterly* 67: 79-125.
- Jacoby, J. and Matell, M.S. (1971), 'Three-point Likert Scales Are Good Enough', *Journal of Marketing Research* VIII: 495-500.
- Kalton, G., Collins, M. and Brook, L. (1978), 'Experiments in Wording Opinion Questions', *Applied Statistics* 27, 2: 23-50.
- Katz, D. and Stotland, E. (1959), 'A Preliminary Statement to a Theory of Attitude Structure and Change', in S. Koch (ed.), *Psychology: A Study of a Science* (New York: McGraw-Hill), 423-75.
- Knowles, E.S. and Condon, C.A. (1999), 'Why People Say "Yes": A Dual-process Theory of Acquiescence', *Journal of Personality and Applied Psychology* 77, 2: 379-86.
- Krosnick, J.A. (1991), 'Response Strategies for Coping with the Demands of Attitude Measures in Surveys', *Applied Cognitive Psychology* 5: 214-36.
- Krosnick, J.A. and Alwin, D.F. (1987), 'An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement', *Public Opinion Quarterly* 51: 201-19.
- Krosnick, J.A. and Fabrigar, L.R. (1997), 'Designing Rating Scales for Effective Measurement in Surveys', in L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin (eds), *Survey Measurement and Process Quality* (New York: John Wiley), 141-64.
- Krosnick, J.A. and Fabrigar, L.R. (Forthcoming), *The Handbook of Questionnaire Design* (New York: Oxford University Press).
- Krosnick, J.A., Holbrook, A.L., Berent, M.K., Carson, R.T., Hanemann, W.M., Kopp, R.J., Mitchell, R.C., Presser, S., Ruud, P.A., Smith, V.K., Moody, W.R., Green, M.C. and Conaway, M. (2002), 'The Impact of "No Opinion" Response Options on Data Quality: Non-attitude Reduction or an Invitation to Satisfice?' *Public Opinion Quarterly* 66: 371-403.
- Krosnick, J.A., Judd, C.M. and Wittenbrink, B. (2005), 'The Measurement of Attitudes', in D. Albarracín, B.T. Johnson and M.P. Zanna (eds), *The Handbook of Attitudes* (Mahwah, NJ: Lawrence Erlbaum Associates), 21-78.

- Krosnick, J.A. and Petty, R.E. (1995), 'Attitude Strength: An Overview', in R.E. Petty and J.A. Krosnick (eds), *Attitude Strength: Antecedents and Consequences* (Hillsdale, NJ: Lawrence Erlbaum Associates).
- Krysan, M. (1998), 'Privacy and the Expression of White Racial Attitudes: A Comparison Across Three Contexts', *Public Opinion Quarterly* 62, 4: 506-44.
- Lapiere, R.T. (1934), 'Attitudes vs. Actions', *Social Forces* 13: 230-37.
- Lazarsfeld, P.F. (1944), 'The Controversy Over Detailed Interviews - An Offer for Negotiation', *Public Opinion Quarterly* 8: 38-60.
- Lehmann, D.R. and Hulbert, J. (1972), 'Are Three-point Scales Always Good Enough?' *Journal of Marketing Research* IX: 444-6.
- Likert, R. (1932), 'A Technique for the Measurement of Attitudes', *Archives of Psychology* 140: 1-55.
- Maio, G.R. and Olson, J.M. (2000), *Why We Evaluate: Functions of Attitudes* (Mahwah, NJ: Lawrence Erlbaum Associates).
- Malhotra, N., Krosnick, J.A. and Thomas, R.K. (2009), 'Optimal Design of Branching Questions to Measure Bipolar Constructs', *Public Opinion Quarterly* 73: 303-24.
- Masters, J.R. (1974), 'The Relationship Between Number of Response Categories and Reliability in Likert-type Questionnaires', *Journal of Educational Measurement* 11, 1: 49-53.
- Milgram, S., Mann, L. and Harter, S. (1965), 'The Lost-letter Technique: A Tool for Social Research', *Public Opinion Quarterly* 29, 3: 437-8.
- Mueller, D.J. (1977), 'Physiological Techniques of Attitude Measurement', in G.F. Summers (ed.), *Attitude Measurement* (London: Kershaw), 534-52.
- Narayan, S. and Krosnick, J. (1996), 'Education Moderates Some Response Effects in Attitude Measurement', *Public Opinion Quarterly* 60: 58-88.
- Oppenheim, A.N. (1992), *Questionnaire Design, Interviewing and Attitude Measurement* (2nd edition) (London: Continuum International Publishing).
- Osgood, C.E., Suci, G.J. and Tannenbaum, P.H. (1957), *The Measurement of Meaning* (Urbana, IL: University of Illinois Press).
- Payne, B.K., Cheng, C.M., Govorun, O. and Stewart, B.D. (2005), 'An Inkblot for Attitudes: Affect Misattribution as Implicit Measurement', *Journal of Personality and Social Psychology* 89, 3: 277-93.
- Petty, R.E. and Cacioppo, J.T. (1986), *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* (New York: Springer-Verlag).
- Petty, R.E. and Krosnick, J.A. (1995), *Attitude Strength: Antecedents and Consequences* (Mahwah, NJ: Lawrence Erlbaum Associates).
- Rankin, R.E. and Campbell, D.T. (1955), 'Galvanic Skin Response to Negro and White Experimenters', *Journal of Abnormal and Social Psychology* 51: 30-33.
- Roese, N.J. and Jamieson, D.W. (1993), 'Twenty Years of Bogus Pipeline Research: A Critical Review and Meta-analysis', *Psychological Bulletin* 114, 2: 363-75.
- Rorer, L.G. (1965), 'The Great Response-style Myth', *Psychological Bulletin* 63, 3: 129-56.

- Rugg, D. and Cantril, H. (1944), 'The Wording of Questions', in H. Cantril (ed.), *Gauging Public Opinion* (Princeton, NJ: Princeton University Press).
- Saris, W.E., van der Veld, W. and Gallhofer, I. (2004), 'Development and Improvement of Questionnaires Using Predictions of Reliability and Validity', in S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin and E. Singer (eds), *Methods for Testing and Evaluating Survey Questionnaires* (Hoboken, NJ: Wiley-Interscience), 275-99.
- Schaeffer, E.M., Krosnick, J.A., Langer, G.E. and Merkle, D.M. (2005), 'Comparing the Quality of Data Obtained by Minimally Balanced and Fully Balanced Attitude Questions', *Public Opinion Quarterly* 69: 417-28.
- Schuman, H. and Converse, J.M. (1971), 'The Effects of Black and White Interviewers on Black Responses in 1968', *Public Opinion Quarterly* 35, 1: 44-68.
- Schuman, H. and Presser, S. (1980), 'Public Opinion and Public Ignorance: The Fine Line Between Attitudes and Nonattitudes', *American Journal of Sociology* 44, 2: 198-209.
- Schuman, H. and Presser, S. (1981), *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context* (New York: Academic Press).
- Smith, T.W. (1985), 'Nonattitudes: A Review and Evaluation', in C.F. Turner and E. Martin (eds), *Surveying Subjective Phenomena* (New York: Russell Sage).
- Sudman, S. and Bradburn, N.M. (1974), *Response Effects in Surveys: A Review and Synthesis* (Chicago: Aldine Publishing).
- Thurstone, L.L. (1928), 'Attitudes Can Be Measured', *The American Journal of Sociology* XXXIII, 4: 529-54.
- Tourangeau, R. and Galesic, M. (2008), 'Conceptions of Attitudes and Opinions', in W. Donsbach and M.W. Traugott (eds), *The Sage Handbook of Public Opinion Research* (London: Sage Publications).
- Tourangeau, R. and Rasinski, K.A. (1988), 'Cognitive Processes Underlying Context Effects in Attitude Measurement', *Psychological Bulletin* 103, 3: 299-314.
- Tourangeau, R., Rips, L.J. and Rasinski, K.A. (2000), *The Psychology of Survey Response* (Cambridge: Cambridge University Press).
- Tourangeau, R. and Smith, T.W. (1998), 'Collecting Sensitive Information with Different Modes of Data Collection', in M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls II and J.M. O'Reilly (eds), *Computer-assisted Survey Information Collection* (New York: John Wiley & Sons, Inc).
- Warner, S.L. (1965), 'Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias', *Journal of the American Statistical Association* 60, 309: 63-9.
- Weisberg, H.F. (2005), *The Total Survey Error Approach: A Guide to the New Science of Survey Research* (Chicago: University of Chicago Press).

- Wilson, T.D. and Hodges, S. (1992), 'Attitudes as Temporary Constructions', in L. Martin and A. Tesser (eds), *The Construction of Social Judgments* (New York: Springer-Verlag), 37-66.
- Wittenbrink, B., Judd, C.M. and Park, B. (1997), 'Evidence for Racial Prejudice at the Implicit Level and its Relationship with Questionnaire Measures', *Journal of Personality and Social Psychology* 74, 6: 1464-80.
- Zaller, J. and Feldman, S. (1992), 'A Simple Theory of the Survey Response: Answering Questions Versus Revealing Preferences', *American Journal of Political Science* 36, 3: 579-616.

Chapter 12 Challenges for Social Measurement

Martin Bulmer

Measurement

Measurement was defined in Chapter 1. One of the most influential twentieth-century statements of the classical approach to measurement was that of psychologist S.S. Stevens, who proposed four scales or degrees of measurement: nominal, ordinal, interval and ratio measurement (1946, 1975). Nominal and ordinal measurement are non-metric; interval and ratio measurement are metric. These theoretical standards are translated in measurement standards in the physical world through organizations such as the United States National Bureau of Standards (NBS). The NBS provides state, county and local officials with technical and operational guides that set out measurement specifications, standard tolerances and model laws designed to support the physical measurement system (Hunter, 1980: 869). The primary standards are those of the International System of Units (SI units) and are seven: length (meter, m), mass (kilogram, kg); time (second, s); electric current (ampere, A); temperature (kelvin, k); luminous intensity (candela, cd); and amount of substance (mole, mol) (Zebrowski, 1979).

This scientific paradigm of physical measurement provides a model that the social sciences, or some social scientists, seek to emulate. The quotation from scientist Lord Kelvin carved on Chicago's Social Science Research Building reflects that aspiration: 'When you cannot measure * your knowledge is * meager* and * unsatisfactory.' The poet e.e. cummings's scepticism reflects doubts as to whether the aspiration is worthwhile in the first place: 'who cares if some one-eyed son of a bitch / invents an instrument to measure spring with'. The place of measurement in social science research is a contentious issue; this tension runs through social science disciplines such as sociology and political science. It is reflected in the ambivalence with which many social scientists look upon research methods such as social survey research. The aim of this final chapter is to consider some of the hindrances to improved measurement of the social, and ways of circumventing these hindrances. There has been a notable failure to agree on standards for social measurement (as distinct from psychological or economic measurement), whether in terms of social indicators and conceptual unification, or at the practical level of operationalizing variables. A certain amount of this ambivalence is reflected in this book.