

## *Designing Questions to Gather Factual Data*

The focus of this chapter is on how to write questions to collect information about objectively verifiable facts and events. Some such questions ask for descriptions of people: their ages, genders, countries of origin, or marital status. Some such questions ask for reporting of what people have done or what has happened to them: obtaining service from doctors, being a victim of burglary, being laid off from a job, or being arrested for drunk driving. Still another class of topics is what people do or how they lead their lives: how much they exercise, what they have eaten or bought, or how they vote.

Although the range of topics is wide, the common element of all the questions to be discussed in this chapter is that, at least in theory, the information to be provided in the answers could be objectively verified. True, in many cases it would take an omniscient, omnipresent observer to keep track of how many soft drinks a person consumed in the last month or how many days a person spent all or part of the day in bed because of injury or illness. However, the fact that there is an objectively definable set of events or characteristics at issue makes a difference: there are right and wrong answers to these questions. The right answers are those that the omniscient, omnipresent observer would provide. This contrasts with the subject of the next chapter, the measurement of subjective states, for which, indeed, there are no right or wrong answers.

Among questions about objective facts, some are aimed at characterizing people whereas others are aimed at counting or describing events. Sometimes the same question can be used to do both.

For example, when a respondent is asked how many times he or she has been a patient in a hospital overnight or longer in the past year, two kinds of estimates could result. First, one could estimate the total number of hospitalizations experienced by respondents. Second, one could estimate the percentage of respondents who had at least one hospitalization experience during the past year. In the following pages, we are going to be discussing strategies for overcoming problems with

questions. Whether a question is aimed at counting events or characterizing people sometimes has a bearing on the optimal solution to a question design problem.

There are five challenges to writing a good question:

1. Defining objectives and specifying the kind of answers needed to meet the objectives of the question.
2. Ensuring that all respondents have a shared, common understanding of the meaning of the question. Specifically, all respondents should have the same understanding of the key terms of the question, and their understanding of those terms should be the same as that intended by the person writing the question.
3. Ensuring that people are asked questions to which they know the answers. Barriers to knowing the answers can take at least three forms:
  - a. never having the information needed to answer the question
  - b. having the information at some point, but being unable to recall the information accurately or in the detail required by the question
  - c. (for those questions that ask about events or experiences during some period of time) difficulty in accurately placing events in time
4. Asking questions that respondents are able to answer in the terms required by the question. It is possible to ask questions to which respondents literally know the answers but are unable to answer the way the investigators want because of a lack of fit between the desires of the investigator and the reality about which the respondent is reporting.
5. Asking questions respondents are willing to answer accurately.

All this must be accomplished with a question that can be administered consistently and has the same meaning to all the people who are going to answer the question so that answers can be aggregated to produce statistical data.

### QUESTION OBJECTIVES

One of the hardest tasks for methodologists is to induce researchers, people who want to collect data, to define their objectives. The difference between a question objective and the question itself is a critical distinction. The objective defines the kind of information that is needed. Designing the particular question or questions to achieve the objective is an entirely different step. In fact, this whole book is basically about

the process of going from a question objective to a set of words, a question, the answers to which will achieve that objective.

Sometimes the distance between the objective and the question is short:

*Objective:* Age

*Possible Question 2.1:* How old were you on your last birthday?

*Possible Question 2.1a:* On what date were you born?

The answers to either of these questions probably will meet this question objective most of the time. An ambiguity might be whether age is required to the exact year, or whether somehow broad categories, or a rounded number, would suffice. Question 2.1 produces more ages rounded to 0 or 5. Question 2.1a may be less sensitive to answer than Question 2.1 for some people, because it does not require explicitly stating an age. There also may be some difference between the questions in how likely people are to err in their answer, due to recall or miscalculations. However, the relationship between the objective and the information asked for in the questions is close, and the two questions yield similar results.

*Objective:* Income

*Possible Question 2.2:* How much money do you make per month on your current job?

*Possible Question 2.2a:* How much money did you make in the last twelve months from paid jobs?

*Possible Question 2.2b:* What was the total income for you, and all family members living with you in your home, from jobs and from other sources during the last calendar year?

First, it should be noted that there are imperfections in each of the three questions. However, the key point is that each of those questions is a possible approach to meeting the objective as stated, but the results will be very different. Obviously, current salary or wage rate might be the best measure of the quality or status of the job a person holds. However, if the purpose of measuring income is to find out about the

resources available to the person, income for the past year might be a more relevant and appropriate measure. Even more appropriate, because people tend to share in and benefit from income from other family members, the total family income from all people and all sources might have the most to do with how “well off” the person is.

A good question objective has to be more specific than simply “income.” More broadly, a question objective can be defined only within the context of an analysis plan, a clear view of how the information will be used to meet a set of overall research objectives. Measuring income is actually a way of measuring social status, resources, or quality of employment. It is necessary to be explicit about the question objective in order to choose a question.

In the course of trying to design and evaluate questions, researchers often are forced to be more specific about their research objectives, what they want to measure and why, than they had been before. Indeed, one of the most common complaints among methodologists who work on question evaluation is that researchers do not have a clear sense of their goals. Until researchers decide what their goals are, it is impossible to write the ideal question.

Another example:

*Objective:* Soft drink consumption

*Possible Question 2.3:* How many soft drinks did you drink yesterday?

*Possible Question 2.3a:* How many soft drinks did you drink in the last seven days?

Again, readers are warned that the above questions both have flaws. However, the issue is the relationship between the question objective and what the specific questions would achieve. One issue is whether the goal is to describe soft drink consumption, to estimate how many soft drinks are consumed by a sample of respondents, or to characterize the respondents in terms of their patterns of soft drink consumption.

For example, the first question will produce a more accurate count of soft drink consumption but for a very limited period. Because behavior on one day is not a very good way to characterize an individual, it would be a poor way to characterize individuals in terms of high, moderate, or low soft drink consumption.

The second question, although likely subject to more response error because it poses a more difficult reporting task, will do a better job than the first of characterizing the individual.

We need a picture of the role of the information in an analysis plan, what the purpose of having the information is, in order to refine the objective and choose a question. If this is a survey for soft drink manufacturers, and the goal is to make a good estimate of aggregate consumption, the questions should be aimed at getting good, precise estimates of total consumption (like Question 2.3). On the other hand, if this is a health survey, and the goal is to identify the extent to which soft drinks are part of people's diets, then characterizing individual patterns of consumption will be the goal (like Question 2.3a).

One more example:

*Objective:* Use of medical care

*Possible Question 2.4:* How many times have you seen or talked to a doctor about your health in the past two weeks?

*Possible Question 2.4a:* How many times have you received any kind of medical care in the last two weeks?

*Possible Question 2.4b:* How many times have you received any kind of medical care in the last 12 months?

There are many aspects of uncertainty generated by this question objective. Two are illustrated by variations in the questions. One issue is what is meant by medical care. Does it only mean visits to a medical doctor, or are there other kinds of experiences that should count? People receive medical care from nonphysicians such as chiropractors, nurses, physicians' assistants, or physical therapists. Should such care be included, or not? Another ambiguity might be whether services from M.D.s such as psychiatrists or ophthalmologists, that might not seem like "medical care," should be counted.

Another issue is, again, whether the goal is to count events, to get estimates of how much service is being used, or to characterize individuals: is this person a high or low user of medical care services? Collecting information for only a few weeks might be the best way to get a good count of visits to doctors, but it is a poor way to characterize the extent to which a particular individual has been using medical services.

The soundest advice any person beginning to design a survey instrument could receive is to produce a good, detailed list of question objectives and an analysis plan that outlines how the data will be used. An example of such a document is Figure 2.1. Although the level of detail can vary, the creation of a document similar to Figure 2.1 serves at least three critical functions. First, it is an outline for the question design process. It not only specifies the goals of each question; it also helps to identify questions that serve no purpose in a survey instrument. If a researcher cannot match a question with an objective and a role in the analysis plan, the question should not be asked.

Second, by relating proposed questions to an outline of objectives, weaknesses in the specified objectives can be identified.

Finally, by stating the objectives in advance, researchers are reminded that designing questions that people are able and willing to answer is a separate task, distinct from defining research objectives. Figure 2.1 does not specify any questions; it only begins to specify the kind of information the answers to some questions should provide. One of the main origins of terrible survey questions is that the researcher did not make the transition from a question objective to a question; the objective was simply put in question form. The hope was that the respondent would do the work for the researcher and produce information that would meet the objective. That seldom works.

Let us now turn to some of the specific challenges for designing questions that meet research objectives.

#### DEFINITION OF CONCEPTS AND TERMS

One basic part of having people accurately report factual or objective information is ensuring that all respondents have the same understanding of what is to be reported, so that the researcher is sure that the same definitions have been used across all respondents. This is one of the most difficult tasks for the designer of survey questions, and failure to do it properly is a major source of error in survey research.

For example, respondents were asked how many days in the past week they had any butter to eat. Many people use the terms butter and margarine interchangeably, so respondents were inconsistent in whether they included or excluded margarine when they answered the question. When the question was rewritten to explicitly exclude margarine, 20% fewer people said they had had any "butter" to eat at all in the past week than was the case when the term was left undefined (Fowler, 1992).

Purpose of Survey: Study correlates of use of medical care. We think medical care is likely to be a function of the following.

- Fiscal resources to afford medical care
- Need for medical care
- Access to medical care
- Perception of value of medical care

Within each of these categories, measurement objectives include:

Fiscal resources relevant to medical care

- Annual family income past year (all sources)
- Liquid assets (savings, bank accounts)
- Health insurance

Need for medical care

- Chronic health conditions that might require care
- Onset of acute illness
- Injuries
- Age/gender (to match with appropriate routine tests and exams)

Access to medical care

- Regular provider or not
- Perceived proximity of provider
- Perceived ease of access
- Perceived financial barriers

Perception of value of medical care

- When not ill (checkups, screening, etc.)
- For chronic conditions (not life-threatening)
- For acute conditions (self-limiting)

Use of medical care

- Visits to doctors
- Other medical services (not M.D.)
- Emergency room use
- Hospitalizations

**Figure 2.1.** Example of an Outline of Survey Content and Question Objectives

A similar example comes from efforts to measure exercise. The most common form of exercise for adults in the United States is walking. However, people are uncertain whether or not to include walking when they report the amount of exercise they do. Answers to survey questions about exercise are seriously affected by whether the wording explicitly includes walking, excludes walking, or leaves the matter undefined.

There are two basic approaches to ensuring consistent understanding of terms.

1. The researcher can provide complete definitions so that all or most of the ambiguities about what is called for are resolved.

2. The respondents can be asked to provide all the information needed in order for the researcher to properly classify events for respondents. In other words, rather than trying to communicate complex definitions to all respondents, if adequate information is reported by respondents, complex criteria for counting can be applied consistently during the coding or analysis phase of a project.

Certainly the most common way to write survey questions that are commonly understood is to build needed definitions into the questions.

*Example 2.5:* In the past week, how many days did you eat any butter?

*Problem:* There are two potential ambiguities in this question. First, it has already been noted that whether the term “butter” includes margarine or not is ambiguous. Second, sometimes it has been found that “the past week” is ambiguous. It could mean the seven days preceding the date of the interview. It also could mean the most recent period stretching from Monday through Sunday (or Sunday through Saturday).

*Possible Solution 2.5a:* In the past seven days, not counting any margarine you may have eaten, how many days did you eat any butter?

*Comment:* The reworded question reduces ambiguity both about whether to include or exclude margarine and about the period that is to be covered.

*Example 2.6:* How many times have you been hospitalized in the past year?

*Comment:* Possibly “hospitalized” is a complex term that everyone will not understand. Sometimes people receive services in hospital clinics, and people go to hospitals for day surgery. Do these services count? There also is the potential ambiguity, parallel to the last example, about the reference period. What is the period to which “the past year” refers?

*Possible Solution 2.6a:* In the past twelve months, since (DATE) a year ago, how many different times have you been admitted to a hospital as a patient overnight or longer?

*Comment:* The new question clarifies several possible ambiguities, including the fact that each new admission counts as a new hospitaliza-

tion event, that hospitalization requires that a person be a patient, and that the person must be in the hospital overnight or longer (which, e.g., excludes day surgery events). It also clarifies the reference period.

Sometimes, the definitional problems are too complicated to be solved by simply changing a few words or adding a parenthetical phrase.

*Example 2.7:* What is your income?

*Problem:* As discussed above, there are numerous issues about how to calculate income. Among them are whether income is current or for some period of time in the past, whether it is only income earned from salaries and wages or whether it includes income from other sources, and whether it is only the person's own income that is at issue or includes income of others in which the respondent might share.

*Example 2.7a:* Next we need to get an estimate of the total income for you and family members living with you during 1993. When you calculate income, we would like you to include what you and other family members living with you made from jobs and also any income that you or other family members may have had from other sources, such as rents, welfare payments, social security, pensions, or even interest from stocks, bonds, or savings. So, including income from all sources, for you and for family members living with you, how much was your total family income in 1993?

*Comment:* That is a very complicated definition. It is necessary because what the researcher wants to measure is a very complicated concept. Even this complex definition avoids, or fails to address, some important issues. For example, what does the respondent do if household composition at the time of the interview is different from during the reference period? The question also does not specify whether take-home pay or total income before deductions is to be reported.

*Example 2.8:* In the past year, how many times have you seen or talked with a medical doctor or a physician's assistant about your health?

*Problems:* This question is taken from the National Health Interview Survey and is frequently asked in health surveys. As noted previously, questions about medical care pose numerous problems regarding what should be reported because the definitions are so complicated.

When the rules for counting events are quite complex, providing a comprehensive, complex definition probably is not the right answer. At the extreme, respondents may end up more confused and the results may actually be worse than if definitions were not provided. A different approach is probably needed.

One approach is to add some extra questions to cover commonly omitted kinds of events. For example, in response to the general question about visits to doctors, it has been found that receiving advice over the telephone from a physician, seeing nurses or assistants who work for a physician, and receiving services from physicians who are not always thought of as "medical doctors" often are left out. One solution is to ask a general question, such as Example 2.5 above, and then ask some follow-up questions such as:

*Question 2.8a:* Other than the visits to doctors that you just mentioned, how many times in the past 12 months have you gotten medical advice from a physician over the telephone?

*Question 2.8b:* Other than what you've already mentioned, how many times in the past twelve months have you gotten medical services from a psychiatrist?

The same kind of thing can be done with respect to income:

*Example 2.9:* When you gave me the figure for your total family income, did you include any income you might have had from interest on stocks, bonds, or savings accounts?

*Example 2.9a:* When you gave me your income figure, did you include all the income that you had from rents?

*Example 2.9b:* Now, if you add in the kind of income you just mentioned that you did not include initially, what would be your estimate of your total family income in 1993?

Using multiple questions to cover all aspects of what is to be reported, rather than trying to pack everything into a single definition, often is an effective way to simplify the reporting tasks for respondents. It is one of the easiest ways to make sure that commonly omitted types of events are included in the total count that is obtained. However, this approach

can be pushed even further, in ways that may make for even better question design strategies.

In some cases, if definitions are very complex, it does not make any sense to try to communicate a shared definition to all respondents. Building on the examples above, instead of trying to communicate to respondents how the researcher wants to define total family income, respondents can be asked a series of questions about the kinds of income they and other family members have had over some period, and what they amounted to. Then the researcher can put together the reported components to fit a particular definition of income that is going to be used for a particular analysis.

There are three rather compelling advantages to such an approach. First, it makes the questions clearer; it is not necessary to communicate a complex, cumbersome definition consistently to all people. Second, it makes the reporting task simpler and more reasonable; the respondent does not have to add income from multiple sources. Third, it may enable the researcher to produce several different measures of income, which may serve different useful analytic purposes. For example, the income of the respondent from earnings might be a good measure of the quality of employment, but the total family income may be a better measure of available resources. Of course, asking multiple questions takes more interviewing time. However, that too may be an advantage. Taking more respondent time, by asking more questions, will improve respondent recall.

If a rough estimate of socioeconomic status is all that is required, a single general question, with all of its flaws, may be acceptable. However, the approach of using multiple questions is often a good alternative to trying to convey complex definitions to respondents.

*Example 2.10:* What kind of health insurance plan do you have: a staff model health maintenance organization, an IPA, PPO, or unrestricted fee-for-service health plan?

*Comment:* That may seem to be a ridiculous question; it is unreasonable to think that most people can make these distinctions among health insurance plans. The approach outlined above, of trying to communicate common definitions, would seem unlikely to succeed given the complexity of models of health insurance that exist in the United States. However, there are some questions that people can answer that probably would enable researchers to classify the kind of health insurance plan to which most people belong.

*Question 2.10a:* In your health plan, can you initially go to any doctor you want, or can you only go to certain doctors or places for your health care?

*Question 2.10b:* (If from a specific list or group) Do the physicians you see only see people who are part of your plan, or do they see other kinds of patients too?

*Question 2.10c:* When you receive medical services under your plan, do you yourself always pay the same amount, no matter that the service, or does the amount you pay depend upon the service you receive?

*Comment:* Maybe the answers to these questions would not enable researchers to make all the distinctions that they would want to make. Moreover, there is a possibility that some people might not be able to answer some of these questions. However, they are much more likely to be able to answer these questions accurately than to learn the definitions of IPA's and HMO's. The general idea of asking people a series of questions they can answer, then attempting to apply more complex definitional strategies to classify patients and their experiences, is a sound way to solve many definitional problems.

*Example 2.11:* In the past twelve months, were you the victim of a burglary?

*Example 2.12:* In the past twelve months, were you the victim of a robbery?

These again are examples of questions that have complex, technical definitions. Burglary is the crime of breaking and entering with intent to commit a felony. Robbery is the crime of taking something from someone by force or threat of force. If a person breaks into a home, and the residents are there and are confronted by the intruder, the would-be burglar becomes a robber. It makes no sense to try to communicate these definitions to respondents so that they can say whether they were burglary or robbery victims. Rather, what makes sense is to have people describe the relevant details of events they experienced, then code those events into the proper, detailed criminal categories.

Sometimes this can be done by asking a series of short, specific questions. For example, when the classification hinges on whether or

not the intruder was confronted by the residents, it is important to ask that specific question. In other cases, respondents may be allowed to respond in narrative fashion, describing their experiences, which then can be coded into categories using specific definitions and decision rules.

Proper question design means making certain that the researcher and all respondents are using the same definitions when people are classified or when events are counted. In general, researchers have tended to solve the problem by telling respondents what definitions the researchers want to use and then asking respondents to do the classification work. Although sometimes that may be the best way to solve the problem, good question design usually will make the task as simple as possible for respondents. It is a new extra step for most investigators to think about what information they need about people that would enable them to do the classification task. However, if investigators identify what simple, easy questions people can answer that will provide the basis for classification, on many occasions better measurement will occur.

### KNOWING AND REMEMBERING

Once a question has been designed so that all respondents understand what is wanted, the next issue is whether or not respondents have the information needed to answer the question. There are three possible sources of problems:

1. The respondent may not have the information needed to answer the question.
2. The respondent may once have known the information but have difficulty recalling it.
3. For questions that require reporting events that occurred in a specific time period, respondents may recall that the events occurred but have difficulty accurately placing them in the time frame called for in the question.

#### Do Respondents Know the Answers?

Often, the problem of asking people questions to which they do not know the answers is one of respondent selection rather than question design. Many surveys ask a specific member of a household to report information about other household members or about the household as

a whole. When such designs are chosen, a critical issue is whether or not the information required is usually known to other household members or to the person who will be doing the reporting.

There is a large literature comparing self-reporting with proxy reporting (Cannell, Marquis, & Laurent, 1977; Clarridge & Massagli, 1989; Moore, 1988; Rodgers & Herzog, 1989). There are occasions when it appears that people can report as well for others as they do for themselves. However, unless questions pertain to relatively public events or characteristics, others will not know the answers. Across all topics, usually self-respondents are better reporters than proxy respondents.

There is another dimension to the topic of knowledge that more directly affects question design. Sometimes respondents have experiences or information related to a question but do not have the information in the form that the researcher wants it. A good example is a medical diagnosis.

There is a literature that shows a lack of correspondence between what conditions patients say they have and what conditions are recorded in medical records (Cannell, Fisher, & Bakker, 1965; Jabine, 1987; Madow, 1967). At least part of this mismatch results from patients not being told how to name their conditions. For example, the patient thinks he has high blood pressure but says he does not have hypertension, because that is not a term he has been given. The patient knows she has growths but did not know the technical name was tumors. It is even easier to think that a physician would not bother to tell a patient that the name for "heart trouble" was "ischemic heart disease." Going back to an example discussed above, there is now a complex array of health plans. Health researchers would like to identify the kind of plans to which people belong, because they are potentially important covariates of the kind of medical care people receive. Respondents are likely not to know the technical terms for the kind of plan to which they belong, even though they have information about the way their plans work that could be used to classify them appropriately.

Having said that, it is common for surveys to ask respondents for information that they do not have. When insurance pays part of the bill, many respondents never know the total cost of the medical services they receive. Many people do not know the medical specialty of the physician that they see. Many people do not know how much their health insurance costs, particularly when a significant portion of it is contributed by their employer.

One critical part of the preliminary work before designing a survey instrument is to find out whether or not the survey includes questions

to which some respondents do not know the answers. The limit of survey research is what people are able and willing to report. If a researcher wants to find out something that is not commonly known by respondents, the researcher must find another way to get the information.

### Stimulating Recall

Memory researchers tell us that few things, once directly experienced, are forgotten completely. However, the readiness with which information and experiences can be retrieved follow some fairly well-developed principles.

Some memories may be painful and subject to repression. However, that is not the issue for the sorts of things measured in most surveys. Rather, the three principles that probably are most relevant include (Cannell, Marquis, & Laurent, 1977; Eisenhower, Mathiowetz, & Morganstein, 1991):

1. the more recent the event, the more likely it is to be recalled
2. the greater the impact or current salience of the event, the more likely it is to be recalled
3. the more consistent an event was with the way the respondent thinks about things, the more likely it is to be recalled.

How does one obtain accurate reporting in a survey? Obviously, one key issue is what one chooses to ask about. If the researcher wants information about very small events that had minimal impact, it follows that it is not reasonable to expect respondents to report for a very long period. For example, when researchers want reporting about dietary intake or soft drink consumption, it is found that even a 24-hour recall period can produce deterioration and reporting error resulting from recall. When people are asked to report their behavior over a week or two weeks, they resort to giving estimates of their average or typical behavior, rather than trying to remember (Blair & Burton, 1987). If one wants accurate information about consumption, reporting for a very short period, such as a day, or even keeping a diary are probably the only reasonable ways to get reasonably accurate answers (A. F. Smith, 1991).

This same kind of trade-off between accuracy of reporting and the length of time about which someone is reporting is a constant in survey design. The National Crime Survey, conducted by the Bureau of the Census for the Department of Justice, and the National Health Interview Survey both initially asked for one year reporting of crimes and hospi-

talizations respectively. However, there was such a drop-off in the accuracy of reporting of events that occurred more than six months before the interview that the surveys now use only events reported within six months of an interview as a basis for generating estimates of the quantity of those events. Indeed, the National Health Interview Survey reports the number of visits to doctors and the number of days people lose from work based on reporting for only the two weeks prior to the interview, because of concerns about inaccuracy of reporting for longer periods (Cannell, Marquis, & Laurent, 1977; Lehen & Skogan, 1981).

A defining characteristic of most interviews is that they are quick question-and-answer experiences. The level of motivation of respondents varies, but for the most part a survey is not an important event in respondents' lives. Hence, without special prodding, respondents are unlikely to invest a significant amount of effort in trying to reconstruct or recall the things that the survey asks them to report (Cannell, Marquis, & Laurent, 1977). For these reasons, researchers have explored strategies for improving the quality of the recall performance of respondents.

One of the simplest ways to stimulate recall and reporting is to ask a long, rather than a short, question. This does not mean making questions more complex or convoluted. However, adding some introductory material that prepares the respondent for the question has been shown to improve reporting (Cannell & Marquis, 1972). One reason may be simply that longer questions give respondents time to search their memories.

Two more direct strategies are used to improve recall. First, asking multiple questions improves the probability that an event will be recalled and reported (Cannell, Marquis, & Laurent, 1977; Sudman & Bradburn, 1982). Second, stimulating associations likely to be tied to what the respondent is supposed to report, activating the cognitive and intellectual network in which a memory is likely to be embedded, is likely to improve recall as well (Eisenhower et al., 1991). The two approaches are interrelated.

Asking multiple questions can be an effective way of improving recall for three different reasons. First, and most obviously, asking close to the same question more than once is a way of inducing the respondent to "try again." Every time a respondent dives into his or her memory bank, the chances of coming up with an answer are improved. Also, one of the effects of asking multiple questions may be to increase the level of motivation of the respondent, and thereby to increase the amount of dedication with which the respondent tries to perform the task of recall.

Second, a specific way of asking additional questions is to focus on the kinds of events that are particularly likely to be forgotten. For



example, one-day stays in the hospital are underreported at a much higher rate than other hospital admissions (Cannell & Fowler, 1965). Specifically asking respondents whether or not they have had a one-day stay (for example, in connection with a false labor) may trigger a slightly different approach to searching and lead to recall of events that were otherwise forgotten.

Third, additional questions can focus on some of the possible consequences of events to be reported, which in turn may trigger recall. For example, if one has been a victim of a crime, it is likely that the police were called or an insurance claim was filed. Asking about calling police or filing claims may trigger recall of a crime.

In a parallel way, recall of medical services received may be stimulated by asking about consequences of such medical services such as buying medications, filing insurance claims, missing work because of illness, or having to make child care arrangements.

There are limits to what people are able to recall. If a question calls for information that most people cannot recall easily, the data will almost certainly suffer. However, even when the recall task is comparatively simple for most people, if getting an accurate count is important, asking multiple questions and developing questions that trigger associations that may aid recall are both effective strategies for improving the quality of the data.

The above has been mainly focused on dealing with not recalling events that should have been reported. Equally important is the problem of overreporting. For example, suppose we ask people whether or not they voted in the last election. The most common response error in response to that question is overreporting, people reporting that they voted when in fact they did not (Sudman & Bradburn, 1982). Part of the reason (discussed in detail later in the chapter) is that voting is seen by some as a socially desirable behavior, so they are motivated to recall and report voting. In addition, however, getting people to remember *not* doing something is a particularly formidable challenge.

Psychologists have theorized that one way to improve the accuracy of reporting is to ask respondents to recreate an experience in their minds. For example, with respect to voting, it might be important to remind respondents who the candidates were and what other issues were on the ballot. Preliminary questions could ask respondents to report where they vote, whether or not they have to get off work to vote, how they are transported to the voting place, and the like. By taking respondents through a series of steps likely to be involved in doing something, the odds of triggering a key memory are increased, and the chances

become increasingly good that the respondent will be able to reproduce the experience more accurately.

### Placing Events in Time

Many of the issues discussed above could reflect an interrelationship between recalling the event at all and placing it in time. If a survey is to be used to estimate the annual number of hospitalizations for a particular sample, people are asked what essentially is a two-part question: Have you been in the hospital recently, and how many times were you in the hospital in exactly the last twelve months?

Studies of recall and reporting behavior show that many of the problems with survey data about such topics stem from difficulties in placing events properly in the time frame designated by the researchers. One of the reasons that hospitalizations that occurred ten to twelve months before an interview are particularly poorly reported is that, in addition to difficulty remembering whether or not the hospitalization occurred at all, respondents have difficulty remembering whether a hospitalization actually occurred before or after that arbitrary line of twelve months ago.

It does not matter a great deal whether the reference period is one week, one month, or one year. If a survey estimate depends critically on placing events in a time period, it invariably is a problem.

There are two approaches researchers use to try to improve how well respondents place events in time:

1. They stimulate recall activities on the part of respondents to help them place events in time;
2. They design data collection procedures that generate boundaries for reporting periods.

In order to improve the ability of respondents to place events in time, the simplest step is simply to show respondents a calendar with the reference period outlined. In addition, respondents can be asked to recall what was going on and what kind of things were happening in their lives at the time of the boundary of the reporting period. Filling in any life events, such as birthdays, can help to make the dates on the calendar more meaningful. If respondents are supposed to report events in the past year, respondents can be asked to think about what they were doing a year ago: where they were living, what was going on in the family, what they were doing at work. If they are able to conjure up

some events that they can associate with the date on or about a year before the interview, or that constitute a clearly definable point in time, it may make it easier for them to decide whether crimes or hospitalizations occurred before or after that point (e.g., Sudman, Finn, & Lannon, 1984).

A related strategy is to ask people to do some associating to improve their perception of the dates or times of year that certain events occurred. So, if respondents are being asked about crimes that occurred to them, they may be asked to think about what the weather was like, what they were wearing, or what else was going on in their lives, which may enable them to come closer to figuring out the approximate date when an event occurred.

These strategies are somewhat time consuming in an interview setting. They often require some individualized efforts on the part of interviewers that are not easy to standardize. As a result, relatively few surveys actually use these techniques. In addition, it probably is fair to say that although some of these techniques seem to improve reporting marginally, none seems to be a major breakthrough.

A very different approach to improving the reporting of events in a time period is to actually create a boundary for respondents by conducting two or more interviews (Neter & Waksberg, 1964). During an initial interview, respondents are told that they are going to be asked about events and situations that happen during the period prior to the next interview. The subsequent interview then asks people about what has happened between the time of the initial interview and the time of the second interview.

Such designs have three helpful characteristics. First, they do produce a clear time boundary. Although the initial interview may not be a big event in people's lives, it does have some cognitive significance for respondents. Second, in their first interview respondents usually are asked to report recent events of the sort to be counted. Researchers then are able to check events reported in the second interview against those reported in the initial interview. If there is double reporting, that is, telescoping events from before interview #1 into the period covered in interview #2, it can be identified. Third, the fact that respondents are alerted to the fact that they will be interviewed about certain sorts of events makes them more attentive and therefore better reporters.

Obviously such reinterview designs are much more expensive to implement than one-time surveys. However, when accurate reporting of events in time is very important, they provide a strategy that improves the quality of data.

Finally, giving respondents a diary to keep should be mentioned. There are special challenges to getting people to maintain diaries. However, to obtain detailed information, such as food consumption or small expenditures, for a short period of time, diaries are an option that should be considered (Sudman & Bradburn, 1982; Sudman & Ferber, 1971).

### THE FORM OF THE ANSWER

Most questions specify a form the answers are supposed to take. The form of the answer must fit the answer the respondent has to give.

*Example 2.13:* In the past 30 days, were you able to climb a flight of stairs with no difficulty, with some difficulty, or were you not able to climb stairs at all?

*Comment:* This question imposes an assumption: that the respondent's situation was stable for 30 days. For a study of patients with AIDS, we found that questions in this form did not fit the answers of respondents, because their symptoms (and ability to climb stairs) varied widely from day to day.

*Example 2.14:* On days when you drink any alcohol at all, how many drinks do you usually have?

*Comment:* Questions asking about "usual" behavior are common. However, they all impose the assumption of regularity on respondents. The question can accommodate some variability, but it is poorly suited to major variability. For example, if a respondent drinks much more on weekends than on weekdays, it is not clear at all how the question should be answered. Questions using the term "usual" need to be scrutinized closely to make sure the answers fit the reality to be described.

*Example 2.15:* How many miles are you from the nearest hospital?

*Comment:* It is easy to think that a respondent might know the exact location of the nearest hospital, yet have a poor notion of the number of miles. Moreover, although miles may be a good measure of distance in a rural or suburban area, time via the likely mode of transportation

might be a more appropriate metric for a city dweller and provide the units in which respondents could answer most accurately.

Asking people questions to which they know the answers is important. However, it is easy to overlook the next essential step—giving respondent an answer task they can perform and that fits the true answer to the question.

### REDUCING THE EFFECT OF SOCIAL DESIRABILITY ON ANSWERS

Studies of response accuracy suggest the tendency for respondents to distort answers in ways that will make them look better or will avoid making them look bad. Locander, Sudman, and Bradburn (1976) found that convictions for drunken driving and experience with bankruptcy were reported very poorly in surveys. Clearly, such events are significant enough that they are unlikely to have been forgotten; the explanation for poor reporting must be that people are reluctant to report such events about themselves. However, the effects of social desirability are much more pervasive than such extreme examples.

For example, when Cannell, Fisher, and Bakker (1965) coded the reasons for hospitalization by the likelihood that the condition leading to the hospitalization might be embarrassing or life-threatening, they found that the hospitalizations associated with the most threatening conditions were significantly less likely to be reported in a health survey. Record-check studies of health conditions, comparing survey reports with medical records, suggest that conditions that might be thought to be embarrassing or life threatening were less well reported in survey interviews (Cannell, Marquis, & Laurent 1977; Cannell & Fowler, 1965; Madow, 1967). Distortion can also produce overreporting. Anderson, Silver, and Abramson (1988) found notable overreporting of voting in elections.

Although social desirability has been used as a blanket term for these phenomena, there are probably several different forces operating to produce the response effects described above. First, there is no doubt some tendency for respondents to want to make themselves look good and to avoid looking bad. In addition, sometimes surveys ask questions the answers to which could actually pose a threat to respondents. When surveys ask about illegal drug use, about drinking alcohol to excess, or about the number of sexual partners that people have had, the answers,

if revealed, could expose respondents to divorce proceedings, loss of jobs, or even criminal prosecution. When the answer to a survey question poses such a risk for respondents, it is easy to understand why some respondents might prefer to distort the answer rather than to take a chance on giving an accurate answer, even if the risk of improper disclosure is deemed to be small.

Third, in a related but slightly different way, response distortion may come about because the literally accurate answer is not the way the respondent wants to think about him- or herself. When respondents distort answers about not drinking to excess or voting behavior, it may have as much to do with respondents managing their own self-images as managing the images that others have of them.

It is fundamental to understand that the problem is not “sensitive questions” but “sensitive answers.” Questions tend to be categorized as “sensitive” if a “yes” answer is likely to be judged by society as undesirable behavior. However, for those for whom the answer is “no,” questions about any particular behavior are not sensitive. When Sudman and Bradburn (1982) asked respondents to rate questions with respect to sensitivity, the question rated highest was how often people masturbated. Presumably its high rating stemmed from a combination of the facts that people felt that a positive answer was not consistent with the image they wanted to project and that it is a very prevalent behavior. Questions about drug use or drunk driving are not sensitive to people who do not use drugs or drive after drinking.

It also is important to remember that people vary in what they consider to be sensitive. For example, asking whether or not people have a library card apparently is a fairly sensitive question; some people interpret a “no” answer as indicating something negative about themselves (Parry & Crossley, 1950). Library card ownership is considerably overreported. Also recall that the event of going to a hospital, which normally is not a particularly sensitive question, can be a sensitive topic for respondents who are hospitalized for conditions that embarrass them or that they consider to be personal.

Thinking broadly about the reasons for distorting answers leads to the notion that the whole interview experience should be set up in a way to minimize the forces on respondents to distort answers. Some of the steps affect data collection procedures, rather than question design per se. The next part of this chapter will outline some of the data collection strategies that can help minimize those forces. This may seem a digression. However, the integration of data collection procedures and question design is critical to collecting good data about sensitive topics. The

balance of this section will be devoted to question design strategies that can reduce distortion.

### Data Collection Procedures

There are three general classes of steps a researcher can take to reduce response distortion:

1. assure confidentiality of responses and communicate effectively that protection is in place
2. communicate as clearly as possible the priority of response accuracy
3. reduce the role of an interviewer in the data collection process.

*Confidentiality.* Survey researchers routinely assure respondents that their answers will be confidential. Protecting confidentiality includes numerous steps such as:

1. minimizing the use of names or other easy identifiers
2. dissociating identifiers from survey responses
3. keeping survey forms in locked files
4. keeping nonstaff people away from completed survey answers
5. seeing to the proper disposal of survey instruments.

In addition, when survey researchers are collecting data that distinctively put people at risk, for example when they were asking about behaviors that violate laws, they can get legal protection from subpoena. Discussions of these issues is found in somewhat more detail in Fowler (1993) and in much greater detail in Sieber (1992).

The key threat to confidentiality is the ability to link an individual to the answers. One of the best ways to avoid that possibility is never to have information about which individual goes with which response. This can be done by using mail or self-administered procedures with no identifiers associated with returns. When there are identifiers, researchers can minimize risks by destroying the link between the respondents and their answers at the earliest possible moment.

However, it is only when respondents understand and believe that they are protected that such steps can result in reduced distortion of survey responses. If there are limits to the extent to which people are protected, ethical research requires that those limits be communicated to respondents as well. If researchers think that the limits to confidentiality they can promise would reasonably affect answers, they should

change the procedures to create a condition that is more conducive to accurate reporting.

*Emphasizing the Importance of Accuracy.* Sometimes, the goals of a survey interview are not clear. In particular, when there is an interviewer involved, there are rules that govern interactions between people that may interfere with the goal of getting accurate reports. Routinely, when we relate to people, we like to put a positive sheen on the way we present ourselves; we like to accentuate the positive; we like to please the other person; we like to minimize stressful topics. Forces such as these may undermine the accuracy of survey answers. To the extent that respondents are following such guidelines, rather than trying to answer as accurately as possible, they are likely to give distorted answers.

There are several steps that researchers can take to reduce the forces to distort interviews. One of the simplest is to have interviewers explicitly explain to respondents that giving accurate answers is the most important thing they can do (Cannell, Groves, Magilavy, Mathiowetz, & Miller, 1987; Cannell, Oksenberg, & Converse, 1977).

Routine interviewer training urges interviewers to minimize the personal side of their relationships with respondents. They are not to tell stories about themselves. They are not to express personal opinions. They are supposed to say there are no right or wrong answers. They are supposed to establish as professional a relationship as is feasible (see Fowler & Mangione, 1990).

In addition, Cannell has demonstrated that interviewer behavior can be systematically manipulated to improve reporting in ways that are probably relevant to response distortion as well. Three specific strategies that have been evaluated by Cannell and his associates (Cannell, Oksenberg, & Converse, 1977; Cannell et al., 1987):

1. Interviewers read a specific instruction emphasizing to respondents that providing accurate answers is what the interview is about and is the priority of the interview.
2. Respondents are asked to verbally or in writing make a commitment to give accurate answers during the interview.
3. Interviewers are trained to reinforce thoughtful answers, and not to reinforce behaviors that are inconsistent with giving complete and accurate answers.

Some of these behaviors are designed to encourage working for complete and accurate recall. They also serve the function of asserting the primacy of accuracy over other goals. Cannell, Miller, and Oksenberg

(1981) found that these procedures, for example, seem to reduce the number of books well-educated people reported reading during the past year, which they interpret as reducing social desirability bias in answers.

*Reducing the Role of Interviewer.* How interviewers affect reporting potentially sensitive information has been a matter of some debate. On the one hand, interviewers can help to motivate respondents, reassure them that they are protected, establish rapport, and thereby increase the likelihood that respondents will answer questions accurately. Alternatively, the fact of presenting oneself to another person through one's answers increases the forces to distort answers in a positive way. The data on this topic do not always support one view or the other; there probably is truth in both. Nonetheless, there is considerable evidence that having people answer questions in a self-administered form, rather than giving answers to an interviewer, may reduce the extent to which people distort answers in a socially desirable direction (Aquilino & Losciuto, 1990; Fowler, 1993; Mangione, Hingson, & Barret, 1982).

In addition to using procedures that do not involve an interviewer, such as a mail survey or a group-administered survey for which people fill out questionnaires and drop them in boxes, there are at least three ways that surveys using interviewers can be modified to reduce the effect of the interviewer on the data collection process.

First, a very well-developed strategy is to have a series of questions put in a self-administered form. The interviewer can hand a booklet of questions to the respondents, the respondents can fill out the questions without the interviewer seeing the answers, and they can put the answers in a sealed envelope. A recent study of drug use clearly demonstrated that the respondents were much more likely to report current or recent drug use in a self-administered form than in response to questions posed by an interviewer (Turner, Lessler, & Gfroerer, 1992).

Second, a modern variation on that is made possible with computer-assisted personal interviewing (CAPI). With such data collection procedures, questions appear on a screen and are answered by some data entry process. If there is a series of questions that the researcher wants to keep private, the computer can simply be given to respondents who can read questions on the screen and answer for themselves, without the interviewer participating. For studies of people who come to fixed locations, such as doctors' offices, schools, or work sites, computers can be set up and used in a similar way to collect data from respondents.

Third, an innovative technique has been introduced into the National Health Interview Survey for a study of teen health risk behavior. In

order to ensure the confidentiality of answers, and protect the teens from interviewer participation, sensitive questions are put on a tape player (such as a Walkman) that can be heard only through earphones. Respondents listen to questions read on the tape, then record their answers on an answer sheet. No observer, interviewer, or parent knows what questions are being answered.

All aspects of the data collection should be designed to reduce the forces on people to distort their answers, particularly when a survey covers material particularly likely to be sensitive. However, these are not substitutes for good question design. The next section deals with ways of designing survey questions to reduce response distortion.

### Question Design Options

There are four general strategies for designing questions to reduce response distortion:

1. Steps can be taken to increase the respondent's sense that a question is appropriate and necessary in order to achieve the research objectives.
2. Steps can be taken to reduce the extent to which respondents feel that answers will be used to put them in a negative light, or a light that is inappropriate or inaccurate.
3. The level of detail in which respondents are asked to answer can be adjusted to affect how respondents feel about giving information.
4. Respondents can be asked to perform a task by which their answer is given in a code that neither the researcher nor the interviewer can directly decipher.

*The Appropriateness of Questions.* Probably no single topic gives survey researchers more grief than asking about income. Interviewers frequently tell stories of respondents who willingly answer questions that seem quite personal, such as behaviors related to the risk of AIDS, only to object to answering a question about their family income.

In this society, certainly there is a sense that income is a private matter that is not to be generally shared. In addition, people's willingness to answer questions in surveys clearly is affected by the extent to which they can see a relationship between a particular question and the objectives of the research. When someone has agreed to participate in research related to health or to political opinions, it is not self-evident why researchers also need to know the respondents' incomes.

Of course, researchers would say that income is an indicator of the resources people have, as well as the kinds of problems they may be

facing. Information about income is helpful in describing and interpreting the meaning of other answers. When the purpose of a question is not obvious, it is only reasonable to explain the purpose to respondents. Although some respondents will answer any question, without worrying about its purpose, providing respondents with sensible explanations about why questions are included can only be helpful in getting them to give accurate information.

A variation on the same theme is that some questions seem inappropriate to certain subsets of the population. An excellent example comes from a recent series of studies aimed at trying to identify the extent to which people are at risk of contracting AIDS. Researchers wanted to find out about risky sexual behavior. One approach is to ask people whether or not they use condoms when they have sex.

Yet, in fact, condom use is relevant only for people who have high-risk partners or partners whose risk status is unknown. The majority of adults in American society have been monogamous or have had no sexual partners for some time. When people are confident that their sexual partners are not HIV positive, asking for details about their sex lives seems (and arguably is) intrusive, and it provides information irrelevant to risk of AIDS. However, for that subset of the population that has multiple sexual partners, asking about use of condoms makes perfectly good sense.

When we did our first survey study of behaviors related to risk of contracting HIV, our pretest instrument included questions about condom use and other sexual practices that increase the risk of transmission that were asked of all respondents. Interviewers and respondents had a great deal of difficulty with those questions; for many who were not sexually active or were monogamous, the questions were offensive and their purpose was hard to understand.

The interview schedule was then changed, so that respondents were first asked about the number of sexual partners they had had in the past year. In addition, respondents who were monogamous were asked whether their partners were in any way at risk. Once a group was identified that was either not monogamous or reported a high-risk partner, questions were asked about protection during sex. For that subgroup of the population, both interviewers and respondents could clearly see why the questions made sense, and the whole series went much more smoothly.

In one way, this is an example of thinking ahead about the purposes of the answers in a research analysis. Researchers should be asking people questions only when there is a clear role for the answers in

addressing the research questions. In addition, the willingness of people to answer questions accurately will be increased to the extent that they can see the role that accurate answers play in addressing the research questions being addressed.

*Managing the Meaning of Answers.* One of the key forces that leads people to distort their answers is a concern that they will be misclassified; that somehow the answers they give will be coded or judged in a way that they consider inappropriate. As a result, they will distort the answers in a way that they think will provide a more accurate picture.

Respondents look for clues to the way their answers will be interpreted. Researchers can reduce the distortion in answers by designing questions in such a way as to minimize respondents' perceptions about how their answers will be judged.

There are three general approaches:

1. The researcher can build in introductions or build a series of questions that minimize the sense that certain answers will be negatively valued.
2. The researcher can design a series of questions that enables the respondent to provide perspective on the meaning of answers.
3. The response task can be designed to structure the respondents' perceptions of how their answers will be judged.

One of the oldest techniques in question design is to provide questions with introductions that say both the answers, or all possible answers, are okay. For example, it was noted that people tend to overreport the extent to which they vote and the extent to which they own library cards (Parry & Crossley, 1950). One reason for this overreporting is that respondents are concerned that researchers will infer that nonvoters are not good citizens or that people without library cards are not literate or have no literary interests. Some people who feel that such a classification is inappropriate will distort their answers, to make them more socially desirable and, indeed, perhaps to make them communicate more accurately the kind of person they think they are.

*Example 2.16:* Did you vote in the presidential election last November?

*Example 2.16a:* Sometimes we know that people are not able to vote, because they are not interested in the election, because they can't get

off from work, because they have family pressures, or for many other reasons. Thinking about the presidential election last November, did you actually vote in that election or not?

*Comment:* The purpose of an introduction like this is to tell the respondent that there are various reasons why people do not vote, other than not being a good citizen. The hope is that respondents will feel more relaxed about giving a "no" response, knowing that the researcher knows some good reasons, some perfectly socially acceptable reasons, why someone might not vote.

*Additional Comment:* It should also be noted that both alternatives are presented in the question, with the perhaps rather feeble "or not." This particular question is not presented in a very balanced way. However, one clue that respondents sometimes use to investigate preferences is whether both options are given equal time, and thereby perhaps equal acceptability, when the question is framed.

*Example 2.17:* Do you own a library card?

*Comment:* When a question is phrased like this, there is a tendency for respondents to think that the researcher expects a "yes" answer. The negative option is not even presented. In fact, it turns out that people are more likely to say "yes" than "no" when asked this kind of question (which is a directive question) because both options are not presented.

*Possible Alternative 2.17a:* Many people get books from libraries. Others buy their books, subscribe to magazines, or get their reading material in some other way. Do you have a library card now, or not?

*Comment:* This question provides some legitimacy and some socially desirable reasons why the "no" response is acceptable. It tries to reassure the respondent that a "no" answer will not necessarily be interpreted as meaning the respondent is uninterested in reading.

There are studies of alternative question wordings that in some cases show little effect from these kinds of introductions. In other cases, they seem to make a difference (Sudman & Bradburn, 1982). When a researcher is concerned that one answer is more acceptable or socially valued than others, one step that may be helpful in reducing those forces is to include introductions to reassure respondents that both answers are considered by the researcher to be reasonable and neither answer will be interpreted as reflecting badly on the respondent.

*Example 2.18:* How many drinks did you have altogether yesterday?

*Comment:* Respondents tend not to like questions like this, standing by themselves, because of the possibility that their behavior yesterday was not what they consider to be typical. In particular, if yesterday was a day when the respondent had more to drink than average, he or she may be reluctant to give that answer, partly because it is seen as misleading.

*Possible alternative:*

*Example 2.18a:* On days when you have anything alcoholic to drink at all, how many drinks do you usually have?

*Example 2.18b:* Yesterday, would you say you had more to drink than average, less than average, or about the average amount to drink?

*Example 2.18c:* How many drinks did you have altogether yesterday?

With this series, the respondent has been allowed to tell the researcher what the usual pattern is and whether or not yesterday's behavior is representative and typical. Having provided that kind of context for the answer, it is easier for a respondent to give an accurate answer.

Loftus cites a similar example (Loftus, Smith, Klinger, & Fiedler, 1991).

*Example 2.19a:* Have you seen or talked with a doctor about your health in the last two weeks?

*Example 2.19b:* Have you seen or talked with a doctor about your health in the last month?

It turns out that different numbers of visits to doctors are reported, depending on the order in which these two questions are asked. When they are asked in the order indicated above, with the two-week question occurring first, there are more doctor visits reported in response to Question A. Moreover, Loftus has shown that the excess reporting from that order stems from overreporting. Apparently, when respondents have had a recent doctor visit, but not one within the last two weeks, there is a tendency to want to report it. In essence, they feel that accurate reporting really means that they are the kind of person who saw a doctor



recently, if not exactly and precisely within the last two weeks. However, by reversing the order of the questions, those who have seen a doctor within the preceding four weeks are given a chance to communicate that they are the sorts of people who have seen a doctor recently. Then, it is easier for them to be literal about the meaning of the two-week question, and their reporting is more accurate.

No doubt, there is both a cognitive and a motivational component to why the question order has the effect that it does. Possibly another effect of changing the question order is to highlight that the researcher really means two weeks, rather than the more generic "recently," when the question says "two weeks." Nonetheless, the tendency to want to be properly classified no doubt is a factor as well.

There is a more general principle to be noted about the relationship between the length of the reporting period and social desirability. It is less threatening to admit one has "ever" used marijuana than to say one has done so recently. Conversely, it is less threatening to say one did not vote in the last election than to say one never voted. A key issue in both cases is what the answer may mean about the kind of person the respondent is. A reformed "sinner" (especially if the sin was long ago) and a virtuous person who occasionally "slips" are images most people are willing to project. A goal is to permit respondents to present themselves in a positive way at the same time they provide the information needed. Attention to the interaction between the reporting period and the message that answers convey is a part of accomplishing this.

Sudman and his associates cite another example of how allowing respondents to provide a context to their answers can improve the quality of reporting (Sudman & Bradburn, 1982). Once again, their focus is the quantity of alcohol consumption.

*Example 2.20a:* In general, would you say that you drink more than your friends, less than your friends, or about the same amount as your friends?

*Example 2.20b:* Think about the friend you know who drinks the most? About how many drinks would you say that person usually has?

*Example 2.20c:* And how about you? On days when you have any alcoholic beverages, about how many drinks do you usually have?

Sudman and Bradburn found that by asking the first two questions, the answer to the third question is, on average, significantly higher. The

reason, they surmise, is that this series enables respondents to provide some context for their answers. Without questions such as A and B, the respondent is left to be concerned about how many drinks the researcher will think is "too many" drinks. There is pressure to be conservative, to downgrade the number of drinks reported, in order to reduce the chances of being judged negatively. However, given the first two questions, the respondent is able to provide some context about how much drinking "a lot" is in his or her social setting. The series almost guarantees that the respondent will be able to report at least one person who drinks more than he or she does. Given that anchor, it becomes much easier to report accurately on the respondent's usual behavior.

One other example from Sudman and Bradburn's research belongs here. Again, the question is about alcohol consumption.

*Example 2.21:* On days when you drink alcohol, how many drinks do you usually have—would you say one, two, or three or more?

*Comment:* In this question, the response categories themselves communicate something to some respondents about how their answers will be evaluated. Given these responses, one would be justified in concluding that three or more is a high category, the highest the researcher even cares about.

*Alternative Example 2.21a:* On days when you drink alcohol, how many drinks do you usually have—would you say one or two, three or four, five or six, or seven or more?

*Comment:* Readers will not be surprised to learn that many more people give an answer of "three or more" in response to this question than to the one that preceded it. In the first question, three drinks was an extreme response; in the second question, it is a much more moderate one. The response categories suggest the researcher thinks that some people drink seven or more drinks.

As a matter of fact, Sudman and Bradburn found that the very best way to ask this question is to have no categories at all; that is, to ask the question in an open-ended form and let people give the number. Although there are times when grouping responses may make the task easier and may be the right decision, researchers should be aware that the response categories provide information to respondents about what they think the range of answers is likely to be.



Finally, the perceived purpose of a question is affected by the subject matter of surrounding questions. For example, the meaning of questions about alcohol use may seem very different, depending on whether they follow questions about using cocaine and marijuana or questions about diet or steps people take to reduce their risk of heart attacks.

*Example 2.22:* Studies have shown that certain steps are associated with lower risks of heart attacks. We are interested in what people do that might affect their risks. For example, in the past 7 days, how many days have you:

- a. Taken any aspirin?
- b. Exercised for at least 20 minutes?
- c. Had at least one glass of wine, can of beer, or drink that contained liquor?

*Comment:* Clearly that question will appear to be very different from the same question in a survey about drug use or drunk driving.

The common theme of the several techniques presented is to minimize the extent to which respondents feel their answers will be negatively judged. Letting respondents provide some evaluative context and minimizing the extent to which the researcher's judgments appear in the questions are both likely to help respondents feel freer to give accurate answers to direct factual questions.

### Minimizing Detailed Answers

The above discussion included an example of when providing less structure to the response task seemed to have a beneficial effect on reporting. However, the opposite can be true as well. In some cases, it is less stressful to answer in broad categories than to give detailed answers.

*Example 2.23a:* To the nearest \$1,000, what is your annual salary rate?

*Example 2.23b:* Is your annual salary rate less than \$30,000, between \$30,000 and \$60,000, or over \$60,000?

*Comment:* Most readers will no doubt feel that the second question will be less subject to response distortion than the first question. Of course, the answers to the second question yield much less information.

However, it is the lack of information, the lack of detail, which also makes the question more acceptable and less stressful.

In telephone surveys, a variation on this approach is used routinely. Respondents are asked about their incomes in broad categories, such as those in Example 2.23b. One or, sometimes, two follow-up questions then are asked that further break down the broad categories. For example, for those saying "less than \$30,000," the follow-up might be:

*Example 2.23c:* Is it less than \$10,000, between \$10,000 and \$20,000, or over \$20,000? In this way, when respondents answer two three-response questions, they are actually being sorted into nine income categories.

Thinking about the level of detail in which answers need to be collected is an important part of the question design process. From an analysis point of view, it often is easier to collect information in greater detail, then combine answers at the analysis stage into bigger categories to yield useful data. However, that is putting the burden onto the respondent. In some cases, it may be best for the researcher to ask respondents to provide information in less detail. Such a strategy can pay off in higher rates of response and less response distortion.

*Giving Answers in Code.* In all of the above strategies, we have talked about ways of structuring the data collection task and the forms of the questions to increase the likelihood that respondents will give interviewers accurate answers. There is another class of strategies that absolutely prevents the researcher, the interviewer, or anyone else from knowing what the respondent's true answer is. Yet, the results can yield useful, analyzable data and estimates.

One example is currently being used in the National Health Interview Survey to estimate the rate at which respondents are at risk of contracting AIDS. The question reads like this:

*Example 2.24:* Is any of these statements true for you?

- a. You have hemophilia and have received clotting function concentrates since 1977.
- b. You are a native of Haiti or central East Africa who has entered the U.S. since 1977.

- c. You are a man who has had sex with another man at some time since 1977, even one time.
- d. You have taken illegal drugs by needle at any time since 1977.
- e. Since 1977, you have been the sex partner of any person who would answer "Yes" to any of the items above.
- f. You have had sex for money or drugs any time since 1977.

*Comment:* A "yes" answer does mean that a respondent has done at least one of the things on the list. However, it does not tell the interviewer or the researcher about any particular activity or risk factor. Only the respondent knows why he or she is at risk. Although it still may be socially undesirable to give a "yes" answer, it probably is easier to say "yes" to a question like that than it is to answer individual questions "yes" or "no."

There is, however, a much more elaborate scheme that has been used to enable researchers to make estimates of the rates of behaviors or events that are highly socially undesirable or illegal. This method is called the random response method and involves the use of the unrelated question (Droitcur, Caspar, Hubbard, et al., 1991; Fox & Tracy, 1986; Greenberg, Abdel-Latif, & Simmons, 1969).

*Example 2.25:*

- a. Have you used marijuana in the last month?
- b. Is your mother's birthday in June?

*Procedure:* The respondent is given two questions, such as those above. Then a procedure is designed to designate which question the respondent is going to answer. The procedure must be one such that only the respondent, and not the interviewer, will know which question is going to be answered.

Procedures that have been used include having the respondent flip a coin or having some system of colored balls that can be seen only by the respondent that designates which question is to be answered.

Let us take flipping the coin, because that is the easiest one to explain. Suppose we have the respondent flip a coin so that only the respondent can see the result. We tell the respondent that if the coin comes up heads, he is to answer Question A; if the coin comes up tails, he is to answer Question B. The answer must be in the form of "yes" or "no."

In this way, although the interviewer will know whether the answer is "yes" or "no," the interviewer will not know whether the respondent is answering Question A or Question B.

**Table 2.1**  
Using Random Response to Make Estimates

	<i>Response to All Questions</i>	<i>Estimated* Response to Unrelated Questions</i>	<i>Inferred Responses to Target Questions</i>	<i>Repercentagized Responses to Target Question</i>
Yes	20%	4%	16%	32%
No	80%	46%	34%	68%
	100%	50%	50%	100%

\*Unrelated question was whether or not your mother was born in June. One half of sample was asked this question. The other half was asked a target question, such as, "Have you used marijuana in the past month?"

How is this useful to researchers? Look at Table 2.1.

Twenty percent of all respondents answered "yes." We know that half the respondents answered Question B, and the true rate for all those answering the question giving a "yes" answer is a little over 8% (i.e., one out of twelve). Hence, about 4 of the 20% yeses can be attributed to respondents whose mothers were born in June. That means the other 16% of the sample answered "yes" because they had used marijuana in the past month. Moreover, because only half the sample answered Question A, our estimate is that about 32% of this population used marijuana in the past month.

There are variations using the same idea that have been tried.

*Example 2.26a:* I want you to perform the following addition. Take the number of days in the past week in which you have used any marijuana at all and add to that the number of working television sets you have in your home now. What is that sum?

*Example 2.26b:* How many working television sets do you have in your home now?

*Procedure:* In some defined percentage of interviews, the interviewer asks Question A; in the rest, Question B is asked. From the sample asked Question B, the researcher is able to estimate the distribution of working television sets; hence, the difference between the mean answer to Question A and the mean answer to Question B constitutes the mean number of times the people answering Question A are reporting that they have used marijuana in the past week.

These techniques clearly have some drawbacks. First, they are time-consuming in an interview. Interviewers have to explain how they work to respondents and convince respondents that in fact no one can figure out what the answer means for a given individual. Second, to be credible, the choice of unrelated questions must be carefully thought out, so that people do not feel exposed either because the "yes" answer is such a rare event for the unrelated question or they think someone could guess what the answer is to the unrelated question. For example, in the second example regarding marijuana use, someone who smoked marijuana every day might be reluctant to perform that task because it would be unlikely that there would be seven working television sets in the home.

Third, the strategy for communicating which question is answered has to be one by which respondents feel confident that interviewers cannot easily guess which question they are answering.

An additional downside of these approaches is that individual level analyses are not possible. It also should be noted that the standard errors of these estimates are based on the number of people who answer the target question, not the number of people who are in the whole sample. Hence, standard sampling errors are larger when this technique is used than they would be if the same information were collected by direct questions asked of everyone. Nonetheless, it is possible to estimate rates for various definable subgroups from these techniques, as well as for populations as a whole.

These problems account for the fact that random response techniques, and their variations, are not very commonly used in survey research. In addition, perhaps because of their complexity, these techniques certainly do not eliminate reporting error. Nonetheless, on some occasions, they have been shown to produce estimates that look more accurate than what researchers have been able to generate from direct questions (Greenberg et al., 1969). Moreover, they absolutely do protect the respondents, because there is no way whatsoever to link a respondent specifically to a reported behavior.

### Conclusion

Many strategies for reducing the forces to avoid socially undesirable answers have been discussed in this chapter. Some, such as the random response techniques, would be used only for a few key measurements that were thought to be extraordinarily sensitive. For example, if one wanted estimates of the rate at which people had done something illegal,

and those estimates were central to the purposes of the research, it might be worth investing five or ten minutes of interviewing time to get two answers using random response. However, a well-designed self-administered strategy for data collection might be just as effective for improving the reporting of socially undesirable material.

For most research, the key messages are: ensure and communicate to respondents the confidentiality of answers; make clear to respondents that being accurate is more important than self-image or rapport with the interviewer; and design questions to minimize the likelihood that respondents will feel their answers will be put in negatively valued categories.

These steps are likely to improve the quality of reporting in every area of a survey, not just those deemed to be particularly sensitive. Researchers never know when a question may cause the respondent some embarrassment or unease. A survey instrument should be designed to minimize the extent to which such feelings will affect answers to any question that is asked.

### CONCLUSION

There are many suggestions for designing good questions embedded in this chapter. The fundamental guidelines are to ask questions that respondents can understand and that they are able and willing to answer. To translate those principles into practice:

1. Avoid ambiguous words; define the key terms in questions.
2. Minimize the difficulty of the recall and reporting tasks given to respondents.
3. For objectives that pose special definitional or are recall challenges, use multiple questions.
4. Give respondents help with recall and placing events in time by encouraging the use of association and other memory aids.
5. Make sure the form of the answer to be given fits the reality to be described.
6. Design all aspects of the data collection to minimize the possibility that any respondent will feel his or her interests will be best served by giving an inaccurate answer to a question.