

center stage in discussions of frequency judgments within the psychological literature. Investigations of rounding in survey reporting (e.g., Huttenlocher, Hedges, & Bradburn, 1990) have shed further light on how respondents compensate for vague or incomplete memories. Since the time of Bartlett's classic work (1932), psychologists have acknowledged that memory involves both retrieval and reconstruction; the survey-based studies on generic memory, estimation, and rounding have added considerable detail to this picture.

Another area where the survey literature has much to offer cognitive psychology involves what might be called *proxy memory* – memories for events experienced by other people. This topic has been almost completely neglected within the mainstream memory literature (see Larsen, 1988, for an exception) but has been a lively area within the movement to apply cognitive theories to issues of survey methods (e.g., Blair, Menon, & Bickart, 1991).

In addition to its implications for the study of memory, the recent efforts to apply cognitive theories to survey issues have much to contribute to the study of attitudes. As we've remarked, investigations of attitudes in social psychology often seem to assume that respondents have a preexisting answer to most attitude questions and need only to read out this answer. The results from the survey literature present quite a different picture: Responses to survey questions can become unreliable over time (Converse, 1964, 1970) and show fluctuations as a consequence of seemingly minor changes in question wording (Schuman & Presser, 1981). In fact, simply changing the order of the questions can produce large swings in the answers (Tourangeau & Rasinski, 1988). If answers to attitude questions are simply readouts of stored judgments, it is not clear why question order should make such a difference. The study of order effects on responses to attitude questions has been a particularly fruitful area for the application of cognitive methods to a long-standing survey problem (Schwarz & Sudman, 1992; Tourangeau & Rasinski, 1988). These models are aimed at explaining survey results, but they have greatly expanded our understanding of assimilation and contrast effects in judgment more generally.

In several areas, then, the effort to apply concepts and methods drawn from psychology to problems in surveys has yielded benefits to both fields. Still, the sailing hasn't always been smooth. In Chapter 11, we consider some of the barriers to further progress.

Respondents' Understanding of Survey Questions

Survey designers don't need to be reminded that the wording of the questions has an important impact on the results. Respondents can misinterpret even well-formulated questions, and when that happens, the question the respondent answers may not be the one the researcher intended to ask. Because of this obvious danger, the questions on national surveys are often subjected to empirical pretests. For example, the questionnaire designers may conduct cognitive interviews or focus groups in which they probe respondents' understanding of the questions and invite them to describe how they go about answering them (see Willis, DeMaio, & Harris-Kojetin, 1999, for a survey of these methods; we present a briefer discussion of them in Chapter 11). This practice is useful in bringing to light problems the designers may have overlooked.

This chapter looks at those aspects of survey questions that make them difficult for respondents to understand. These aspects are of many different sorts, ranging from features of grammar and word meaning to the broader situation in which the respondent and interviewer find themselves. Grammar can come into play either because the sentence is structurally ambiguous or because it includes complex clauses that respondents cannot parse. As an example of structural ambiguity, Item (1) asks respondents whether they agree or disagree with this statement:

- (1) Given the world situation, the government protects too many documents by classifying them as SECRET and TOP SECRET. [GSS]

As Fillmore (1999) points out, this sentence has two readings: According to one reading, the government, motivated by the world situation, protects too many documents; according to the other, the government protects more documents than can be justified by the world situation. The ambiguity relates to syntactic structure (it depends on what part of the

sentence the initial clause modifies), and it may affect how a respondent answers the question. A recent study by Stinson (1997) centered on a question that illustrates the sort of grammatical complexity that can lead to trouble:

- (2) Living where you do now and meeting the expenses you consider necessary, what would be the smallest income (before any deductions) you and your family would need to make ends meet EACH MONTH?

Although syntax can present hurdles for respondents, most studies of question wording have focused on semantic problems – problems of meaning – especially those involving the meaning of individual words. Many words in natural language are *ambiguous* (have more than one meaning) or are *vague* (have imprecise ranges of application). In addition, survey questions may include obscure or technical terms that are *unfamiliar* to respondents. Opinion surveys, for example, may ask about newly emerging issues that are unfamiliar to many of the respondents.

Vagueness and ambiguity can lead respondents to interpret questions in variable ways. For example, when Belson (1981) probed respondents in a follow-up interview about the meaning of Question (3), he found differences among respondents in the age range they attributed to *children*:

- (3) Do you think that children suffer any ill effects from watching programmes with violence in them, other than ordinary Westerns?¹

The respondents' difficulties might have been due to the ambiguity of the term *children*, which can refer either to sons and daughters of any age (as in *How many children do you have?*) or to youngsters in particular. Vagueness may be a more likely source, however, because the meaning of *youngster* is not crisply bounded (the division between youngster and adult is not well defined). And, of course, the term *ill effects* is deliberately vague. Similarly, Belson found that respondents gave a wide range of interpretations to the adverbial quantifier *usually* in (4):

- (4) For how many hours do you *usually* watch television on a weekday?

¹ Questions (3) and (4) are not from actual surveys. Belson (1981) composed them for research purposes in order to embody the types of question-wording problems most often found in items "provided by research organizations whose representatives made available questionnaires which they had used over the past two years" (p. 23).

Question (4) also illustrates a further problem that affects comprehension. Many questions *presuppose* that certain characteristics apply to the respondent and then *focus* on an associated aspect. In (4), for example, *usually* presupposes that there is some usual pattern in the respondent's weekday television viewing, and the question focuses on how many hours per day make up that pattern. Presupposition and focus are normal components of a sentence's meaning, but they lead to difficulties in surveys when the presupposition fails to apply. If there is no regular pattern to the respondents' TV watching, then they must either opt out of the question (e.g., by responding *don't know*) or reinterpret the question in ways that apply to them. Difficulty with presuppositions may also occur in (5), another item from the General Social Survey, in which respondents have to rate their agreement or disagreement:

- (5) Family life often suffers because men concentrate too much on their work. [GSS]

This item presupposes that men concentrate too much on their work and focuses on its effect on family life. A respondent who agrees with the subordinate clause (men concentrate too much) and disagrees with the main clause (family life suffers) should have no special difficulty answering (5). But a respondent who disagrees with the subordinate clause may feel that the question doesn't properly apply to him or her (Fillmore, 1999).

This item raises another difficulty: What sort of position would someone be advocating by making the statement in (5)? Is the intent to convey the feminist view that men should take on a fairer share of the household chores and child-rearing responsibilities? Or is the intent to convey a more fundamentalist position that family life should take priority over outside activities? Depending on which reading the respondents give to the item, they may embrace or reject its implied sentiment.

These examples illustrate the major classes of interpretive difficulty that survey designers encounter. The question's grammatical structure (its *syntax*) may be ambiguous or too complicated for respondents to take in. Lengthy or complex questions can exceed respondents' capacity to process them, resulting in misinterpretations (e.g., Just & Carpenter, 1992). The question's meaning (or *semantics*) may elude respondents if they misunderstand vague, unfamiliar, or ambiguous terms or if they are misled by inapplicable presuppositions. Finally, the intended use of the question (its *pragmatics*) may create difficulties, as in (5). To begin exploring these comprehension difficulties more systematically, we begin

by looking at the nature of questions and the processes involved in understanding them. The remaining sections then examine the contributions of grammar, meaning, and use in respondents' approach to questions.

2.1 What Is a Question?

The comprehension difficulties that respondents face usually involve understanding questions, and we will focus on questions here. Obviously, respondents also have to comprehend sentences of other sorts, especially at the beginning of the survey interview, in explanatory passages, and during transitions between parts of the survey instrument (e.g., *Now I'd like to ask you some questions about your children*). With self-administered questionnaires, comprehension of various kinds of instructions, especially those about the route respondents are supposed to take through the questionnaire, can create problems as well (Jenkins & Dillman, 1997). But many of the aspects of comprehension that we will discuss in connection with questions carry over to other sentences as well.

One immediate difficulty in thinking about questions, however, is that we can view them at different levels of analysis. Questions are associated with certain surface forms, generally give rise to a particular class of meanings, and are usually intended to perform a specific kind of action. But although these levels of form, meaning, and action are correlated with each other, the correlation is far from perfect. We cannot concentrate on one level at the expense of the others.

Like other complex linguistic objects, questions display a characteristic grammatical and phonological structure. For example, questions often have inverted word order (*Where was Catherine working last week?* rather than *Catherine was working where last week?*) and a rising intonation contour.² Thus, the term *question* often refers to a class of lin-

² But not always. *Echo questions* can preserve the order of a preceding statement in conversation (e.g., *I fed your headband to the gerbil. You fed my headband to the gerbil?*). *Subject questions*, such as *Who fed your headband to the gerbil?*, also have normal word order (cf. *I fed your headband to the gerbil*; see Radford, 1997, Section 7.7). Rising intonation likewise appears at the ends of some questions but not all. Bolinger (1957, p. 1) cites the following example from Raymond Chandler, noting that the final question probably doesn't rise:

Mr. Hady is on nights and Mr. Flack on days. It's day now so it would be Mr. Flack would be on.
Where can I find him.

guistic objects – interrogative sentences – that we typically use in asking for information. This meaning looms large in survey designers' talk of *question wording*: Given the sort of information that we want, what's the best way of structuring the question as a linguistic object to get at those facts or opinions?

But once we begin to consider the possibility of alternative wordings, we seem to presume that there is *something* of which they are alternative versions: an abstract question that we can ask in different ways. If we're interested in finding out when someone begins his or her commute, we might ask (6), adapted from the Long Form used in the decennial census:

(6) What time did Calvin usually leave home to go to work last week?

But we could also use *Could you please tell me when Calvin usually left home to go to work last week?* or *On those days when this person worked last week, when did he or she usually leave home?* These versions have clearly distinct linguistic forms, but, at least in some situations, they get at the same information and should receive the same answer. It is not easy to say exactly what the something is that each of these items expresses in common, but according to recent theories of the semantics of questions (e.g., Groenendijk & Stokhof, 1997; Higginbotham, 1996), this shared aspect of meaning is what we will call a *space of uncertainty*. This space consists of a set of possibilities, each of which constitutes a potential answer to the question. One of these possibilities is the correct answer. For Question (6) and its variants, this uncertainty space might be the set of all propositions of the form *Last week, Calvin usually left home to go to work at time t* for all clock times *t*. A respondent could give any of these propositions as an answer to the question, although only one of them (perhaps *Calvin usually left home to go to work at 9:15 a.m.*) would be the correct answer. We indicate this space in a schematic way in Figure 2.1a, where the different points in the space correspond to different answer possibilities and the starred point indicates the correct possibility. If respondents also have response options for Question (6) (e.g., 7:00–7:59 a.m., 8:00–8:59 a.m., etc.), each option will correspond to a single possibility, collapsing the earlier set of points (see Figure 2.1b). If the response options are vague (e.g., *morning*, *afternoon*, or *evening*), then the possibilities may share some of their propositions, as in Figure 2.1c. We discuss this last case in Section 2.4.2.

An interrogative sentence is a common way to express an uncertainty space, but it's not the only way. We can ask the same (abstract) question using an imperative sentence (*Please tell me when Calvin usually left home for work last week*) or a declarative sentence (*I'd like to know*

What time did Calvin usually leave home to go to work last week?

Uncertainty Space a:

- Calvin leaves home at 12:00 a.m.
- Calvin leaves home at 12:01 a.m.
- ⋮
- * Calvin leaves home at 9:15 a.m.
- ⋮
- Calvin leaves home at 11:59 p.m.

Uncertainty Space b:

- { Calvin leaves home at 7:00 a.m.
- Calvin leaves home at 7:59 a.m.
- { Calvin leaves home at 8:00 a.m.
- Calvin leaves home at 8:59 a.m.
- * { Calvin leaves home at 9:00 a.m.
- Calvin leaves home at 9:59 a.m.

Uncertainty Space c:

- * morning { Calvin leaves home at 5:00 a.m.
- Calvin leaves home at 11:30 a.m.
- afternoon { Calvin leaves home at 11:00 a.m.
- Calvin leaves home at 5:30 p.m.
- evening { Calvin leaves home at 5:00 p.m.
- Calvin leaves home at 9:30 p.m.

Figure 2.1. A schematic view of the uncertainty space for the question *When did Calvin usually leave home to go to work last week?* Panel a shows the uncertainty space without response options; panel b, for precise response options (6–7 a.m., 7–8 a.m., etc); panel c, for imprecise options (*morning*, *afternoon*, and *evening*).

when Calvin usually left home for work last week). And interrogatives do not necessarily express an uncertainty space; they can express a statement (*Did you know that Calvin usually left home for work at 9:15 last week? How could you possibly think that I wiped my feet on your mouse pad?*) or request an action (*Could you please stop wiping your feet on my mouse pad?*), as well as laying out a space of possible answers.

Finally, we can define questions not as a certain type of form or meaning, but as the activity that people perform when they ask for information. Seen this way, questions are one sort of speech act (Searle, 1969). When someone asks when Calvin usually leaves home for work, he or she is usually making a request – that the listener provide information about the time Calvin leaves home for work. If the meaning of a question is an uncertainty space, then the request to the listener is to provide information about which possibility in the space happens to be true. The standard way to make this request is to use an interrogative sentence, and when that way is used we have the typical alignment of sentence form, meaning, and use that appears in Table 2.1 (adapting the view of Higginbotham, 1996). In using the interrogative sentence (6), the questioner is expressing a space of uncertainty (that Calvin usually left home for work last week at time t for all relevant values of t) and requesting that the listener provide information as to which of these possibilities is correct. As Table 2.1 also shows, this correlation between grammatical form, meaning, and use for questions parallels a similar correlation for statements (including answers to questions).

The correlation between form, meaning, and use in Table 2.1 is easy to break because interrogative sentences do not necessarily lay out an array of possibilities, as we've already noted. Likewise, one can use interrogative sentences without requesting information. Graesser, Huber, and Person (1992) distinguish four classes of grammatical questions, only one of which corresponds to a request for information of the sort found in surveys. The others monitor common ground in conversation

TABLE 2.1 Components of Questions and Statements

	Grammatical Structure	Meaning	Use
Questions	Interrogative sentence (e.g., "When did Calvin usually leave home to go to work last week?")	Space of uncertainty (e.g., the set of propositions <i>that Calvin usually left home to go to work last week at time t</i>)	Requesting information (e.g., requesting someone to inform you when Calvin usually left home to go to work)
Answers	Declarative sentence (e.g., "Calvin usually leaves home for work at 9:15")	Proposition (e.g., the proposition <i>that Calvin usually leaves home for work at 9:15</i>)	Informing (in response to a request) (e.g., asserting that Calvin usually leaves for work at 9:15)

(*Do you follow me?*), coordinate social action (by issuing instructions or seeking permission), or control conversation and attention (as with rhetorical questions). As Bolinger (1957, pp. 2–3) put it, “No one element suffices to define a [question]. . . . For persons who demand rigorous definitions, the term *question* cannot be defined satisfactorily. . . .”

But despite this play in the connection between them, when interrogatives, uncertainty spaces, and requests for information line up, we have something like a prototypical question. Although questions can deviate from the prototype in many ways, it provides a starting point for our discussion of survey questions in the sections that follow.

2.2 Two Views of Comprehension: Immediate Understanding versus Interpretation

Before tackling other aspects of question comprehension, we need to address one other preliminary issue: What *is* comprehension? What is the product of the question-understanding process? Unfortunately, the term *comprehension* is itself ambiguous. On the one hand, the meaning that we get from a word or a sentence must be relatively stable across people; how else could we understand each other? But this stability implies that the interpretation of sentences has to be at least somewhat immune to differences in the amount of knowledge about the concepts. When a survey includes a question about commuting times, both the transportation planner who formulated the question and the commuter who answers it must share some essential set of meanings, even though the two may differ in both the depth and kind of information they bring to bear on the concept *commuting*. But, on the other hand, it seems quite reasonable to think that a transportation planner attaches a much richer and more abstract meaning to *commuting* than the typical commuter does. So how can their interpretations of a question about commuting really be the same?

This discrepancy in our intuition about the stability of meaning across listeners mirrors a similar discrepancy in our intuitions about when we have successfully understood a sentence. In the normal course of a conversation, we process sentences in a seemingly effortless way, and we feel we have interpreted each sentence adequately as soon as (or perhaps a bit before) we come to the end. Unless we are brought up short by a difficult grammatical construction (as in garden path sentences such as *The horse raced past the barn fell*) or an unfamiliar word or phrase (say, *computational lexicography*), we understand the sentence immediately

and are ready to move on to the next one. But it is also clear that comprehending a sentence doesn't always end at the period, with the reader or hearer secure in the right interpretation. If someone says that *Bill gave another great sermon*, and the hearer realizes that Bill is neither a minister nor a priest, then he or she may interpret the sentence nonliterally – inferring that the speaker intends it ironically. But another hearer who doesn't know Bill might interpret the same statement literally. There is obviously plenty of room for misunderstanding and contradictory interpretations in ordinary talk.

2.2.1 Representation-of and Representation-about the Question

How can we accommodate these intuitions – that understanding is generally shared and immediate but that it can reflect idiosyncratic knowledge and change or deepen over time? We assume that the product of comprehension consists of two parts, one obligatory and the other optional. Both parts are mental representations centered on the sentence that a person has just read or heard, but they differ in their content. One representation consists of a specification of the underlying grammatical and logical structure of the sentence, together with the lexical representation of the individual words it contains. The other representation consists largely of inferences that the interpreter draws from the sentence in conjunction with other knowledge that he or she has available on that occasion. We call the first a *representation of the sentence* and the second a *representation about the sentence* (Rips, 1995). The *representation of* the sentence is more or less constant across individuals competent in the language. The *representation about* the sentence varies, however, depending on the interpreter's standpoint, knowledge of the subject matter, knowledge of the speaker or writer, knowledge of the context in which the sentence was uttered or written, and probably many other factors. The *representation-about* will also vary with the amount of time and effort that the individual devotes to interpreting it: The greater the amount of interpreting that goes on, the richer this representation will be.

Consider, once again, Question (6), *What time did Calvin usually leave home to go to work last week?* According to some current theories in syntax and semantics (e.g., Higginbotham, 1996; Larson & Segal, 1995), the underlying structure of this sentence (its logical form) is similar to that in (6'):

(6') [[Which(*t*)] [?[Last week, Calvin usually left home to go to work at *t*]]],

where *Which*(*t*) specifies the questioned element of the sentence, ? marks the construction as a question, and *t* is a variable ranging over clock times. We discuss this type of formulation in more detail in the next section, but for now we can assume that (6') gives the skeleton of the representation-of the question, the framework that people compile as the result of hearing it. In addition, the representation-of the Calvin question must also contain some information about the meanings of the words and other lexical items in (6'). For example, the representation has to specify that *Calvin* is an expression that refers to an individual, that *work* refers to an event, *t* to a time, and so on.

2.2.2 Constructing the Representation-of a Question

If we step back from the formatting details, it is apparent that deriving the representation-of a question involves several cognitive operations:

- Representing the question in some format (like 6') that makes its logical structure clear;
- picking out the question's focus ([*Which* (*t*)]);
- linking the nouns and pronouns to the relevant concepts in memory (e.g., associating the terms *Calvin* and *last week* with their cognitive representations);
- assigning meanings to the predicates in the underlying representation (*usually*, *leave home*, *go to work*).

Graesser and his colleagues include essentially these same operations in their model of the question interpretation process (e.g., Graesser, Bommarreddy, Swamer, & Golding, 1996; Graesser et al., 1994).

What's controversial about the representation-of is its lexical content, the concepts that represent the meanings of the noun phrases, pronouns, and predicates. This component must suffice to determine the range of potential answers – the uncertainty space of the question – but beyond that point there is disagreement. According to some theories (e.g., Fodor, 1981, 1994; see also Anderson, 1983), the mental representations of the lexical items are fairly similar to words in natural language. According to others (e.g., Jackendoff, 1991; see also Schank, 1975), the mental representations are deeper and more fine-grained, specifying both the

primitive conceptual elements that underlie words and the larger conceptual structures that these elements are embedded in.

2.2.3 The Representation-about the Sentence

Respondents do not stop interpreting a question when they have finished determining its representation-of. The question about Calvin's commuting time, for example, seems to imply that Calvin has some set pattern, a regular time when he leaves for work. For this reason, a respondent might infer that if Calvin works irregular hours, then the question doesn't apply to him. If the question is accompanied by a set of response options, the respondents may use the options to refine their interpretation of the question. If the response options are 6:00–7:00 a.m., 7:00–8:00 a.m., 8:00–9:00 a.m., 9:00–10:00 a.m., and "other," then they know that the question doesn't require an answer that's precise to the minute. Likewise, they may take the response options as tacitly specifying the usual range of answers that people give to such questions – the typical times people begin work. Perhaps they even assume that the actual frequency of starting times in the population is about equal for each of the response options.

There is an endless set of possible inferences that respondents can make about the question that could be included in their representation-about it. Graesser and his colleagues distinguish 13 types of inferences that readers can make as they read a story (Graesser, Singer, & Trabasso, 1994). Only two of them (inferences that identify the referents of pronouns and noun phrases and those that assign case roles to the noun phrases) are needed for the representation-of; the remainder all help elaborate the representation-about. Which inferences respondents actually make will depend on factors like the amount of time they have to think about the question, their understanding of the purpose of the survey, the amount of information they have about the topic, and so on. Although some of these inferences might be more common than others, it's unlikely that every respondent will draw exactly the same ones. Thus, the representation-about the question is likely to vary across respondents and may even vary for a single respondent across occasions.

2.2.4 Relation between the Two Representations

Although we are treating the two representations as distinct entities, we do not mean to imply that there is no interplay between them.

Certainly, people may use the representation-of the sentence as the basis of inferences that become part of the representation-about it. And it is possible that the representation-about the sentence is involved in constructing its representation-of. As a person listens to a question, he or she may form hypotheses about how it will continue, hypotheses that may guide the construction of a representation-of the question. However, these hypotheses are not themselves part of the representation-of the sentence, and the listener may need to revise or discard them later when more of the sentence comes in. Thus, we needn't assume that people first construct the representation-of and then the representation-about in strict sequence.

Both the representation-of and the representation-about the question have an impact on respondents' answers, but these effects come from different directions. Complex wording or complex logical requirements can prevent respondents from being able to compute the representation-of, and in such a case, respondents are in much the same situation they would be in if they had heard only a fragment of the item. They are missing basic information they need to determine the question's space of uncertainty and, as a result, they cannot be expected to come up with a relevant answer. Difficulties surrounding the representation-about the question, however, usually stem from too much information rather than from too little. Respondents may make unwarranted inferences about the question and use those inferences in constructing an inaccurate answer. Suppose, for example, that respondents infer that the response options provide the typical answers to the question and base their own answers on whether they believe they are above or below average. Then their answers will vary with the particular set of categories the survey designer has chosen, no matter what the correct answer happens to be (Schwarz, 1996). In general, then, respondents' problems with representations-of a question may require clarifying and supplementing the question itself. But problems with representations-about the question may require explicitly canceling inferences that the item seems to invite. In the remaining sections of this chapter, we make use of this representation-about/representation-of distinction in examining effects due to the interrogative form, meaning, and use of survey questions.

2.3 Syntactic Difficulties in Question Wording

Let's return to the interrogative form, the first component of typical questions in Table 2.1, to see what difficulties it can pose for respon-

dents. In processing this component, the respondent's job is to get the question into its underlying propositional format (as in (6')) and to identify the question's focus. Difficulties in accomplishing these tasks partly reflect surface features of the interrogative form. In addition, they may reflect syntactic ambiguity or excessive complexity.

2.3.1 Interrogative Form

As we noted earlier, interrogatives usually involve displacement of words from the positions they occupy in the corresponding declarative sentences. In *yes/no questions* (i.e., questions calling for a yes or no answer), these changes are small, involving a switch in the position of the sentence's subject and an auxiliary verb. For example, the interrogative *Have you had a mortgage on this property since the first of June?* [CE] corresponds to the declarative *You have had a mortgage on this property . . .*, where the auxiliary *have* has changed places with the subject *you*. When the declarative has no auxiliary, a form of the word *do* appears instead at the beginning of the question. *Do you have a home equity loan?* is the interrogative form of *You have a home equity loan*.

Matters are more complicated, however, for questions that begin with *wh-words*, such as *who*, *where*, *when*, *why*, *what*, *which*, and *how*. Most of these *wh-questions* shift the position of the subject and the auxiliary, just as *yes/no questions* do. For example, *What would you have to spend each month in order to provide the basic necessities for your family?* flips the order of *you* and *would*. But the more dramatic difference is the position of the *wh-word what*. The corresponding declarative seems to be of the form *You would have to spend X each month . . .*; so the *wh-word* has switched to the front of the sentence from the position *X* occupies in the declarative version. In fact, there can be many embedded clauses separating the *wh-word* in complex questions from its corresponding position in a declarative. According to current generative theories of grammar (see, e.g., Radford, 1997), *wh-questions* take shape through a process that moves the *wh-word* to the front of the sentence, leaving behind a silent (i.e., unpronounced) grammatical marker or *trace* in the original position. Thus, the representation of (7a) will contain a trace *t** in the position shown in (7b):

- (7) a. What would you have to spend each month in order to provide the basic necessities for your family?

- b. What would you have to spend t^* each month in order to provide the basic necessities for your family?
- c. I would have to spend \$1,000 each month in order to provide the basic necessities for my family.

The trick in understanding wh-questions is to determine the trace position, since it is the trace position that determines the question's focus – the information that needs to be filled in to answer the question. For example, (7c) can serve as an answer to (7a), where \$1,000 occupies the position of t^* in (7b).

In (7), the trace can occupy only one position, but questions can be ambiguous in this respect. The question in (8a), for instance, has two readings, depending on whether the question is asking about the time of the telling (trace at the position shown in (8b)) or the time the telephone will be repaired (trace at the position shown in (8c)):

- (8) a. When did Lydia tell Emily the telephone would be fixed?
- b. When did Lydia tell Emily t^* the telephone would be fixed?
- c. When did Lydia tell Emily the telephone would be fixed t^* ?

How do people determine the position of the trace in understanding wh-questions? Research on sentence parsing in artificial intelligence (AI) suggests some approaches to this problem (see Allen, 1995, Chapter 5, for a review). The basic idea is that when people encounter the wh-word at the beginning of a question, they store in memory information predicting that they will encounter a missing part in the remainder of the sentence. Which component will be missing is partly determined by the nature of the wh-component itself. Because the trace shares the grammatical properties of the wh-phrase, the trace and the phrase will belong to related grammatical categories. In (7), for example, the missing piece will have to be a noun phrase because *what* fills the role of a noun phrase. The missing part of (8) will be a prepositional phrase, since *when* fills the role of a prepositional phrase (i.e., *at what time*)? As they process the sentence, people look for the missing component. If they run into a stretch of the sentence that is ungrammatical because it lacks the predicted part, they can plug the trace in to fill the gap. If the sentence is grammatical without the predicted part, they must nevertheless find a spot where the trace can go.

Because monitoring for the trace position requires the listener to use working memory, this process is likely to make comprehension more difficult until he or she finds the correct location. (See Just & Carpenter,

1992, for an account of the role of working memory in sentence comprehension.)³ Psycholinguistic research suggests that people make guesses about the position of the trace in wh-argument questions as soon as they process the sentence's verb (e.g., Crain & Fodor, 1985). They may even make preliminary decisions about trace positions before they reach any potential trace sites (see Tannenhaus, Boland, Mauner, & Carlson, 1993).

Consider how this trace-location process works in comprehending (7) and (8). In the case of (7a), the wh-word *what* alerts the listener that a noun phrase will be missing later in the question. As he or she processes the rest of the sentence, the listener encounters the word *spend*, a transitive verb that is missing its object (*You spend* is not grammatical). Since the expected noun phrase can fill the role of this missing component, the trace must occur after *spend*, as in (7b). Parsing (8a), however, is a bit more difficult. One would expect a missing prepositional phrase because of *when*, but there are no clues about the location of the phrase in the rest of the question. *Lydia told Emily the telephone would be fixed* is perfectly grammatical by itself, so there is no need for the prepositional phrase to patch up the syntax. Since the phrase could attach to either verb in the sentence (*tell* or *would be fixed*), the trace could appear in either the position it occupies in (8b) or the one in (8c). To decide between these two readings, the listener must rely on plausibility, intonation, or other external factors (e.g., stressing *tell* – *When did Lydia TELL Emily the telephone would be fixed?* – favors (8b) over (8c)).

As these examples illustrate, wh-words differ in whether or not the component they represent is an obligatory or an optional part of the remaining sentence. The words *who(m)*, *which*, and *what* often begin components called *arguments* that are required by verbs or prepositions in the rest of the sentence. The words *where* (in what place?), *when* (at what time?), *why* (for what reason?), and *how* (in what manner?) often begin components called *adjuncts* that are optional parts of the question. Argument questions tend to provide more guidance than adjunct questions about where the trace should go. There is also evidence that ambiguous adjunct questions like (8a) are somewhat easier to understand when

³ According to some parsing theories, grammatical rules determine the position of the trace automatically without needing a separate monitoring process. We can assume, however, that the rules that carry out this process draw on extra memory resources in handling these questions, so the effect on working memory may be the same.

people interpret the trace as occurring in the main clause (i.e., (8b)) than when they interpret it as occurring in the subordinate clause (i.e., (8c)).

In summary, questions with different surface forms impose different demands on the listener. Relative to yes-no questions, *wh*-questions increase the load on working memory, since they require the listener to reconstruct the position of the queried component. Among *wh*-questions, those that concern arguments (questions about *who*, *which*, or *what*) may be somewhat easier to process than those that concern adjuncts (questions about *where*, *when*, *why*, or *how*). Finally, among adjunct questions, those that focus on an uncertain aspect of the main clause are easier to understand than those that focus on an uncertain aspect of a subordinate clause.

2.3.2 Ambiguity and Complexity

The examples we have touched on so far have already indicated two further sources of difficulty that can result from the syntax of a question – ambiguity and complexity. In the framework presented here, grammatical ambiguity arises because the missing trace (the focus of the question) can be linked to more than one component of the underlying representation of the sentence. For example, in (8), the queried time may involve when Lydia told Emily about the telephone or when the telephone will be fixed. In general, ambiguities arise in complex questions with multiple embedded clauses, so rewriting the question to eliminate the embedding seems a natural strategy for clarifying the intended meaning. Closed questions can also clarify things, since the answer options will point to the focal component of the question.

Even when the complex form does not introduce syntactic ambiguities, it may overload the processing resources of the respondents. Consider this example from Fowler (1992), modeled on a question in the Health Interview Survey:

- (9) During the past 12 months, since January 1, 1987, how many times have you seen or talked to a doctor or assistant about your health? Do not count any time you might have seen a doctor while you were a patient in a hospital, but count all other times you actually saw or talked to a medical doctor of any kind.

Both the syntax and semantics of the question are complicated. The question covers face-to-face and telephone consultations, with doctors

“of any kind” as well as with “assistants.” In addition, respondents are to exclude such consultations if they took place while the respondents were hospitalized (or if they didn’t concern the respondents’ health) and to restrict their responses to visits that took place during the time frame specified. Processing this question into its underlying logical form is likely to impose quite a burden on working memory, one that may exceed the capacity of the respondent.

Questions like (9) are the product of the practical constraints that questionnaire designers face. On the one hand, the questions need to specify the exact concepts the questions are trying to tap. In the case of (9), most of the question and the accompanying instructions aim to define a particular set of medical visits – outpatient medical visits that occurred during a one-year period, where *visit* is construed broadly to include telephone consultations. On the other hand, there is the need to save time. The cost of a survey is, in part, a function of the length of the questionnaire. So, rather than ask a series of simpler questions to get at the same information as (9), the survey designers compress all four of the main possibilities (face-to-face visits with a doctor, other face-to-face visits with medical personnel, telephone consultations with a doctor, telephone consultations with other medical personnel) into a single question.

In attitude surveys, there is another pressure that makes for complicated questions. Many survey researchers believe that balanced items like (10a) are better than items that state only one side of an issue (10b):

- (10) a. Some people feel the federal government should take action to reduce the inflation rate even if it means that unemployment would go up a lot. Others feel the government should take action to reduce the rate of unemployment even if it means the inflation rate would go up a lot. Where would you place yourself on this [seven-point] scale?
- b. Some people feel the federal government should take action to reduce the inflation rate even if it means that unemployment would go up a lot. What do you think? Do you agree strongly, agree, . . .

Question (10a) is taken from Converse and Presser (1986, p. 38), a widely used text on questionnaire design. The respondents are to indicate their answers on a seven-point scale, whose endpoints are labeled *Reduce Inflation* and *Reduce Unemployment*. A simpler alternative would be to present items like (10b) and the parallel question on reducing unemploy-

ment; however, Converse and Presser observe that a substantial number of respondents would agree with both of these items, while (10a) encourages such respondents to take a more definite position.

Aside from the conceptual complexity of the underlying representation of the question, several other variables affect the load a question imposes on working memory (Just & Carpenter, 1992). One is the degree of embeddedness; questions along the lines of *Is that the dog that chased the cat that ate the rat that Karen saw?* impose an especially heavy burden on processing capacity. Another is syntactic ambiguity, which increases the burden on working memory by forcing listeners to entertain two interpretations; garden path sentences (which require re-interpretation at the end) are similarly burdensome. A final variable is the individual respondent's working memory capacity. According to Just and Carpenter (1992), individuals differ sharply in how much they can hold in working memory; questions that overburden one respondent may pose no particular problem for another.

There are two main consequences to overloading working memory: Items may drop out of working memory (i.e., their level of activation may get so low that the item can no longer be used in ongoing processing) or cognitive processing may slow down (Just & Carpenter, 1992). Respondents may take a long time to deal with Fowler's item on doctor visits in the past year, their representation of the question may omit some part of the question's intended meaning, or both things may happen, with respondents taking a lot of time to come up with an incomplete interpretation of the question.

2.4 Semantic Effects: Presupposition, Unfamiliarity, and Vagueness

In the framework we have adopted, a question specifies an uncertainty space – a set of possibilities that correspond to the range of legitimate answers (see Figure 2.1). The survey designer's job is to ask a question in such a way as to convey the intended space, and the respondent's job is to reconstruct the space and say where the correct answer lies within it. This perspective is a handy one because it allows us to discuss some of the common semantic problems that can derail a survey question. The question can express a space whose possibilities are not exhaustive, providing no location that the respondent can identify as a correct answer. In the extreme, when the question uses terms that are unfamiliar to the respondent, the question may not express a space of possibilities

at all. The question and its response alternatives can also produce a space that is poorly specified – for example, one in which the regions in the space aren't mutually exclusive. Lack of exhaustiveness results from (faulty) presupposition; lack of exclusiveness results from vagueness.

2.4.1 Presupposition

We found that statements like (5) – *Family life often suffers because men concentrate too much on their work* – can carry not-so-innocent assumptions. The same is true of questions. Question (6) – *What time did Calvin usually leave home to go to work last week?* – presupposes that Calvin had a job and asks what time he usually left to go there. If Calvin has no job, the question is ill-posed, and no direct answer is possible. Instead, the respondent would be forced to object that the question is simply not applicable to Calvin. Presuppositions like these arise because questioners must somehow describe the event or state about which they seek information. To answer the question, the addressee must identify the relevant events (e.g., Calvin's departures for work during the last week), identify the queried property of those events (their usual time of occurrence), and search memory for information relevant to the answer. (See Graesser, Roberts, & Hackett-Renner, 1990, Graesser et al., 1994, and Singer, 1985, for theories of question-answering that run along these lines.) Descriptive information in the question allows the addressee to perform these tasks by specifying what the question is about. The presupposed information consists of things that the questioner and addressee normally assume to hold, since they are among the conditions that make the question meaningful. If the addressee does not agree with the information, however, then he or she must make some adjustment, either to accommodate or to reject it.

We can think of a question's presuppositions as limiting the uncertainty space that the question expresses. The uncertainty space that Question (6) expresses is roughly the set of propositions of the form *Calvin left home to go to work at t* for all clock times *t* (e.g., 8:15 a.m.), as we saw in Figure 2.1. For the question to be correctly posed, some proposition in this space must be true. If none is true (because Calvin has no job) or if more than one is true (because he has several jobs or works irregular hours), then the question has no good answer. Presuppositions restrict the range of possible propositions in the space: The more stringent the presuppositions, the narrower the range of allowable answers. *At what time of the morning does Calvin usually leave home*

to go to work? adds to the presuppositions of (6) that Calvin doesn't work a late shift. The uncertainty space of the question in Figure 2.1 is correspondingly narrowed to just those times that occur during the morning.

Research by Loftus and her colleagues has documented the effects of presupposition on people's memory for events. In these experiments, *leading questions* (i.e., questions containing a false presupposition about an event) can cause addressees to misremember the event as if the presupposition were true. A question like *How fast was the car going when it went through the yield sign?* can cause subjects to report the presence of a yield sign on a follow-up memory test, even if no such sign was part of the original traffic event that the subjects witnessed (Loftus, 1979). The cognitive basis of these false-memory effects remains controversial (for a taste of the controversy, see Belli, 1989, Tversky & Tuchin, 1989, and Zaragoza & McCloskey, 1989), but it is clear that under certain circumstances presuppositions can lead respondents to make incorrect inferences about what happened.

The effect of leading questions may be due in part to a normal reaction that occurs in conversation when a question contains a presupposition about which the addressee has no prior knowledge. Presupposed information is something that the questioner assumes (and believes that the addressee also assumes). When the addressee does indeed know it, then all is well, and he or she can proceed to answer without further ado. In some cases, however, the addressee may not know the presupposed information (e.g., may not know that Calvin has a job in the case of Question (6)). The presupposition is not necessarily rejected unless the addressee believes it to be false. In fact, if the addressee thinks the questioner is in a position to know the truth of the presupposition, the addressee may find it informative and remember it as factual. Although this information is not something the addressee knew before, nevertheless the addressee can *accommodate* to the presupposition, treating it as true (see Lewis, 1979, and Stalnaker, 1974).

In the context of a survey, of course, respondents are unlikely to suppose that the interviewer knows more about their personal circumstances than they themselves do, so they would be unlikely to interpret the interviewer's presuppositions in a question like (6) as news. Some survey questions, however, depend on more specialized information, such as facts about medical procedures and conditions (*Have you or any family members ever suffered a myocardial infarction or heart attack? Do you belong to a health maintenance organization?*) or knowledge of

public issues (*Do you favor or oppose the Agricultural Trade Act of 1978?*). If the interviewer is asking their opinion about the Agricultural Trade Act of 1978, then respondents may infer that this is something they could (or should) have an opinion about, rather than an issue that is deliberately obscure or even nonexistent.

How do respondents cope with questions that presuppose information respondents do not have? With opinion questions, some respondents (usually a majority) simply state that they don't know (Schuman & Presser, 1981, Chapter 5). Those respondents who do answer may look to the prior questions to support a guess about the meaning of the obscure issue (see chapter 7 for several examples of such context-based inferences with unfamiliar issues). When the item concerns a factual matter (*Do you or any members of your family have dental sealants?*), respondents seem to employ a variety of strategies (Lessler et al., 1989). They may ask for a definition of the unfamiliar term or state that they do not know. The question's apparent presupposition that the respondent ought to know the term may, however, encourage other strategies. (And, as we shall see, standard survey practice leaves interviewers little room to define unfamiliar terms.) Some respondents seem to assimilate the problematic term to a similar-sounding, more familiar one (*denture cream*). Others conclude that the answer must be no, reasoning that they would probably be more familiar with the term if it applied to them (Gentner and Collins, 1981, describe similar inferences based on lack of knowledge). Surely, we would know if we'd had a *myocardial infarction*. But each of these strategies for generating a substantive answer can lead to problems.

Presuppositions are inescapable in natural language questions because they are a necessary part of expressing the range of uncertainty that questions address. It is usually possible, however, to avoid most troublesome presuppositions through standard survey tactics. Questionnaires often include filter questions that route respondents around items that don't apply to them and issues they never heard of. Similarly, it's common to add "don't know" or "no opinion" options in attitude questions to reduce the pressure on respondents to fabricate opinions about issues they are not familiar with (Converse & Presser, 1986; Sudman & Bradburn, 1982). However, such tactics cannot eliminate all presuppositions. Even if we add a filter question that asks whether Calvin has a job, the main question about his commuting habits still presupposes that Calvin has only one job and that he leaves home for that job at a regular time. No question is presupposition-proof. The best any questionnaire can do

is to avoid presuppositions likely to be false in a significant number of cases within the intended population.

2.4.2 Vagueness in Questions and in Response Alternatives

Like presupposition, vagueness is impossible to avoid in natural language. Vagueness occurs when it is unclear whether or not some descriptor applies to an object or event. In the case of Question (3), repeated as (11a) later, it is unclear whether the descriptor *children* applies to teens (or to older offspring); there is no fact of the matter that could decide this issue. We might take a step toward making (11a) precise by stipulating the age boundary, as in (11b). Although this cutoff seems somewhat arbitrary, it at least manages to eliminate borderline cases that could make (11a) problematic for some respondents. However, what about *ill effects*, *programmes with violence*, and *ordinary Westerns*? To deal with *ill effects*, for example, we could try listing ill effects, as in (11c), but in addition to the fact that some ill effects may themselves be vague, it would be very difficult to spell out all the ill effects that are relevant to the question. Ill effects are inherently open-ended, so making the phrase precise by fiat means missing some clearly relevant symptoms. Perhaps the best we can do is to acknowledge this with an *and-so-on* at the end of the list, as in (11c). Much the same is true of *violence* in this context. We can go some way toward clarifying the concept by giving examples, as in (11d), but this is hardly precise enough to settle all questions about whether specific incidents constitute violence.

- (11) a. Do you think that children suffer any ill effects from watching programmes with violence in them, other than ordinary Westerns?
 b. Do you think that people under the age of 14 suffer any ill effects from watching programmes with violence in them, other than ordinary Westerns?
 c. Do you think that people under the age of 14 suffer any ill effects from watching programmes with violence in them, other than ordinary Westerns? By *ill effects* I mean increased aggression in school or at home, increased nightmares, inability to concentrate on routine chores, and so on.
 d. Do you think that people under the age of 14 suffer any ill effects from watching programmes with violence in them, other than ordinary Westerns? By *ill effects* I mean increased aggression in school or at home, increased nightmares, inability to concentrate on routine

chores, and so on. By *violence*, I mean graphic depictions of individuals inflicting physical injuries on others or on themselves, depictions of individuals wantonly damaging property or possessions, abusive behavior and language to others, and so on.

Vague Concepts

Like the earlier example about doctor visits (9), (11d) would clearly overwhelm the working memory capacity of most respondents, raising the issue of whether the effort to achieve precision is worthwhile. The danger in vagueness is supposed to be that some respondents will choose one way to make a vague question precise, whereas others will choose a different way, leading to uninterpretable variability in the responses. In commenting on his respondents' understanding of (11a), Belson (1981, p. 182) remarks that "It is . . . well worth noting that there was a high degree of variability in the interpretation of terms like 'children', 'ill effects', 'violence' – such that respondents who offer identical choice of answer may well have been considering rather different aspects of the matter concerned." He reports, in fact, that only 8% of respondents understood the question as intended. But the danger in making efforts to clarify vague concepts is that it produces lengthy, complicated questions that are also hard to interpret.

It is certainly possible that respondents to questions like (11a) sometimes adopt interpretations that differ radically from those of others. For example, according to Belson, some respondents understood *children* as kids eight years old or younger, whereas others understood *children* as those 19–20 years old or younger. Later, we will also see evidence that people interpret vague frequency expressions (e.g., *pretty often*) in different ways. Nevertheless, some degree of vagueness seems built into the meaning of important concepts such as *children* and *violence*. If these are indeed the concepts we're interested in, then we cannot avoid imprecision entirely. In fact, as we argue in Chapter 6, part of what it means to have an attitude is to construe an attitude object in a characteristic way; part of the reason why some people favor *welfare spending* and some oppose it is that they see the issue in different terms. Such differences in how attitudinal concepts are interpreted are partly what we seek to measure by asking attitude questions.

Moreover, some of the evidence of variability in understanding may be due in part to differences in the way respondents specify meanings when they are asked to do so after the fact (e.g., in cognitive interviews or follow-up questionnaires), not to differences in their immediate com-

prehension. It is quite possible that in computing the representation-of the question, respondents do not make a vague expression precise but deal with the expression in its own terms. It is only when asked to explain what they understood by *violence* or *children* that respondents reach for more specific explications – part of their representation-about the question.⁴ At that point, variability is unavoidable because there are many ways to draw arbitrary boundaries.

Since the early 1970s, research on categorization has stressed the gradedness of everyday categories, such as *furniture* or *flower* (e.g., Rips, Shoben, & Smith, 1973; Rosch, 1973). Subjects rate some members of these categories as being more typical than others (e.g., chairs are judged to be more typical as furniture than TV sets are), and these typicality ratings predict many other aspects of their responses to these categories. For example, people take longer to decide that atypical members belong to the category than typical members (e.g., it takes longer to judge that a TV set is furniture than that a chair is furniture), and they are more willing to generalize from typical than from atypical members.

Later research has made it clear that these typicality effects are not always due to vagueness about what counts as members of a category. A category member can be atypical without necessarily being a borderline case (Armstrong, Gleitman, & Gleitman, 1983). A dandelion may be an odd flower, for example, but it is clearly a flower nonetheless. In such cases, gradedness of typicality or exemplariness does not necessarily entail vagueness about category membership. For many categories, however, gradedness in typicality and gradedness in membership go hand in hand. A patch of color intermediate between orange and red is not only an atypical red but also a borderline red. TV sets are not just atypical pieces of furniture but also borderline furniture, since there are no technical facts about either furniture or TVs that would reveal their true category status (Malt & Johnson, 1992). For categories like these, attempts to draw the boundaries sharply may be useful for certain purposes, but they also falsify to some extent the nature of the categories

⁴ The situation in interpreting vague predicates would then parallel other instances in which people must introspect about their own mental processes (Nisbett & Wilson, 1977). In asking respondents what they understood by *violence* when they first heard Question (11a), we may be asking them to try to be linguists and to analyze the meaning of the term from a theoretical perspective. Since respondents typically have no training in performing such an analysis, the results may be highly variable. This is likely to be especially true when the follow-up question occurs on the day following the interview, as it did in Belson's (1981) study (see Ericsson & Simon, 1984).

themselves. In crafting questions, we may want to distinguish categories that are inherently vague (e.g., *children*, *violence*, *ill effects*) from categories whose fuzziness is due to other factors (e.g., respondents' lack of knowledge). Explanatory comments, such as those in (11c) and (11d), may be more valuable with unfamiliar concepts than with inherently vague ones.

Vague Quantifiers

Vagueness affects nearly all facets of language in surveys – not only the content of the questions, but also the wording of the response alternatives. Bradburn and Sudman (1979, Chapter 10) called attention to the fact that surveys often give their respondents a choice among ordered natural-language categories (e.g., *never*, *not too often*, *pretty often*, *very often*) that may not have exact or constant numerical equivalents. Scales of this sort include adverbial quantifiers for frequency (e.g., *never*, *not too often*), probability expressions (e.g., *very unlikely*, *unlikely*, *likely*), and quantifiers for amounts (e.g., *none*, *few*, *some*, *many*). Most of the expressions on these scales correspond, at best, to a range of numerical values. But both the range and the central tendency sometimes depend on the typical frequency of the quantified event (Pepper, 1981), the other alternatives on the scale (e.g., Newstead, 1988), and group and individual differences among respondents (e.g., Budescu & Wallsten, 1985; Schaeffer, 1991). Moxey and Sanford (1993) provide a comprehensive review of such factors. To take an extreme example, the statement that earthquakes occur *very often* in California implies a very different objective frequency (perhaps once a year) from the statement that someone sneezes *very often* (perhaps once a day). Bradburn and Miles (1979) show that *very often* seems to refer to a somewhat higher frequency as applied to incidents of excitement than of boredom. Thus, it may be difficult to compare numerically a *pretty often* response to one question with the same response to a different question (or from a different respondent).

The data in Figure 2.2 (from Schaeffer, 1991a, Table 2) illustrate the difficulty with vague response categories. The results come from a survey of 1,172 respondents who were asked, "How often do you feel . . . particularly excited or interested in something? Very often, pretty often, not too often, or never?" and "How often do you feel bored?" If the respondents gave an answer other than "never," they were then asked, "About how many times a week or a month did you mean?" (Bradburn & Sudman, 1979). The figure plots the mean numerical frequency that

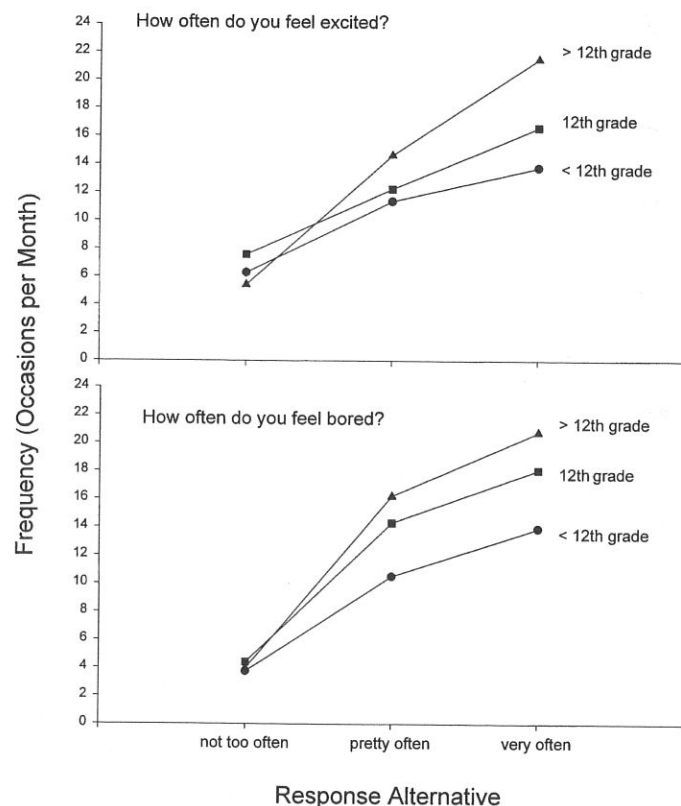


Figure 2.2. Absolute frequency of being excited (top panel) and being bored (bottom panel) for those who said they were excited (bored) “very often,” “pretty often,” and “not too often.” The X-axis shows the highest grade in school completed by the respondents. Data from Schaeffer (1991a, Table 2). Copyright © 1991. Reprinted with permission of the University of Chicago Press.

respondents gave for each of the categorical responses *not too often*, *pretty often*, and *very often*. For example, the top panel shows that respondents who said that they felt excited “very often” explained that they meant about 17 times per month, those who said “pretty often” meant 13 times, and those who said “not too often” meant 6 times. What’s of interest in these data, however, is that the numerical equivalent for “very often” and “pretty often” varied with the respondents’ level of education: More educated respondents apparently had in mind larger values for each of these two response categories. Much the same was true for the respondents’ age, with younger respondents giving

larger numerical values for both “very often” and “pretty often” than did older respondents. Schaeffer (1991a) interprets these results to mean that phrases like “very often” reflect an implicit comparison to a prototypical frequency that varies by group. “Very often” means roughly “more often than typically happens to people like me.”

One way to model vagueness in response alternatives is to allow the set of possibilities to overlap in the uncertainty space for the question, as we did in Figure 2.1c. For the response alternatives “never,” “not too often,” “pretty often,” and “very often,” the space consists of just four sets, each containing a range of propositions that specify the exact frequency. The set for “not too often” might contain the propositions *I feel bored 1 time a month, . . . , I feel bored 20 times a month*; the set for “pretty often” might contain *I feel bored 5 times a month, . . . , I feel bored 50 times a month*; and so on. Overlap among the sets indicates the fuzziness of the response categories. In line with this model are empirical attempts to map quantifiers like these to regions of a numerical scale by asking subjects whether (or to what degree) *not too often* applies to something that occurs n times a week. Results of these studies display extensive overlap between adjacent categories (Moxey & Sanford, 1993). Individual differences or differences due to the content of the question could be handled by reassigning the propositions to the available categories.

Although this uncertainty-space model seems natural, we need to be wary of the idea that the meaning of a quantifier like *pretty often* is equivalent to a distribution of numerical values. In the first place, as Moxey and Sanford (1993) note, the overlap in scale values for even six or seven of these quantifiers is usually so large that they could hardly serve any useful communicative purpose if this were their sole meaning. In the second place, as we discuss in Chapter 5, phrases like *pretty often* are often used to convey the fact that the respondents have only a vague sense of the relevant quantity; it is not as if Bradburn and Miles’s respondents knew exactly how often they were bored each month and decided that that exact frequency fell into the range covered by *pretty often*.

It might be more reasonable to think of these quantifiers in their own terms as specifying a relative position within an ordinal series given by the complete set of response options. In part, the meaning of each of these terms reflects the contrast set defined by the alternatives (Fillmore, 1999). Thus, what *pretty often* means in the context of the data of Figure 2.2 is just: more often than *not too often* and less often than *very*

often. A respondent's representation of the alternatives may amount to no more than this, as we suggested in the case of other vague terms. Still, we need a way to capture the plight of the respondent who does recall a specific number of incidents and is then faced with the task of deciding whether this number is *pretty often* or *very often* for the purpose of answering the question. In these terms, overlap among the possibilities in the uncertainty space provides a clear picture of the respondent's dilemma. Perhaps the most important issue, however, from the vantage point of survey designers is that the content of these possibilities changes with question content and with respondent characteristics, as we've noted. This means that translation of the response options back to numerical values is a complex undertaking. If designers want numerical information about frequencies or amounts, then it may be best for them to ask for it directly.

2.5 Survey Pragmatics and Its Effects on Comprehension

The purpose of a question in a survey is to make a request: to get the respondent to provide information as specified by the question's meaning. What the question conveys, however, often goes beyond what it literally says. Respondents can take the question as a starting point for a complicated set of inferences that then influence the answers they produce. Two kinds of factors trigger these inferences, both based on failures of respondents' expectations. The first concerns a question's apparent failure to meet standards of conversational informativeness; the second concerns failure of the interaction between the interviewer and respondent to conform to normal social patterns. We consider these two sources of inferences in turn.

2.5.1 Pragmatic Theory

In a series of lectures entitled "Logic and Conversation," the philosopher Paul Grice (1989) outlined a way to reconcile the strict meaning of sentences (what we have called the representation-of the sentence) with the somewhat different meaning that they sometimes convey (the representation-about the sentence). Grice's main purpose was to square the interpretation that logical words like *or*, *if*, and *some* have in formal logic and the rather different meanings they assume in everyday speech,

but his analysis carries over to other differences between what sentences mean and what they convey in conversational settings.

Grice's Conversational Maxims

Grice believed that conversations are governed by what he called a *cooperative principle* – a kind of implicit agreement between participants to make their contributions support the conversation's purpose. Since we often have conversations to accomplish useful things, it's in everyone's interest to make them work smoothly. To flesh out this cooperative principle, Grice divided it into several subprinciples or *maxims* that control the course of conversation. These are supposed to be general rules of reasonable behavior in conversations – a conversational etiquette. For example, Grice discussed a maxim of *Quantity* that asserts that you should (a) make your contribution as informative as is required (for the current purposes of the exchange) and (b) do not make your contribution more informative than is required. This maxim guards against someone's misleading his or her conversational partner by saying too little or derailing the conversation by saying too much. Grice also mentioned a maxim of *Quality* that enjoins: (a) Do not say what you believe to be false and (b) Do not say that for which you lack adequate evidence. There is also a maxim of *Relation* that says to be relevant and a maxim of *Manner* that says to be clear.

Grice's reason for spelling out these maxims was to show that people can convey information indirectly by saying things and expecting the listener to interpret them as cooperative. Suppose you're having a conversation with someone, and he or she utters sentence *S*. Let's suppose that you have no reason to think that your conversational partner is deliberately being uncooperative. So you should interpret *S* in a way that makes it seem cooperative and nonmisleading. This sometimes entails adding to what your partner said some additional information *I* that the partner must believe in order for the sentence to be truly cooperative. Grice calls this extra information a *conversational implicature* of the original utterance. It is an *inference* drawn from the original sentence that preserves its cooperativeness.

Grice gives a number of examples of how this process works. In one of them, the speaker tells Fred that he is out of gas, and Fred says, "There's a gas station around the corner." For Fred to be cooperative in this situation, he must believe that the gas station is not closed, and the inference that Fred believes this is a conversational implicature. More

interesting examples are cases in which a speaker deliberately violates (or flouts) one of the maxims. One example is that of a teacher writing a letter of recommendation for a job candidate. If the letter says merely, "Dear Sir, Mr. X's command of English is excellent, and his attendance in class has been regular. Sincerely yours, etc." then the teacher has deliberately violated the first maxim of quantity – make a contribution as informative as is required. Assuming that the teacher is being cooperative (otherwise, why would he write at all?), he probably believes that Mr. X is not a good job prospect because that belief would explain why he is flouting the maxim.

Grice's examples follow the pattern shown in Figure 2.3. If there is

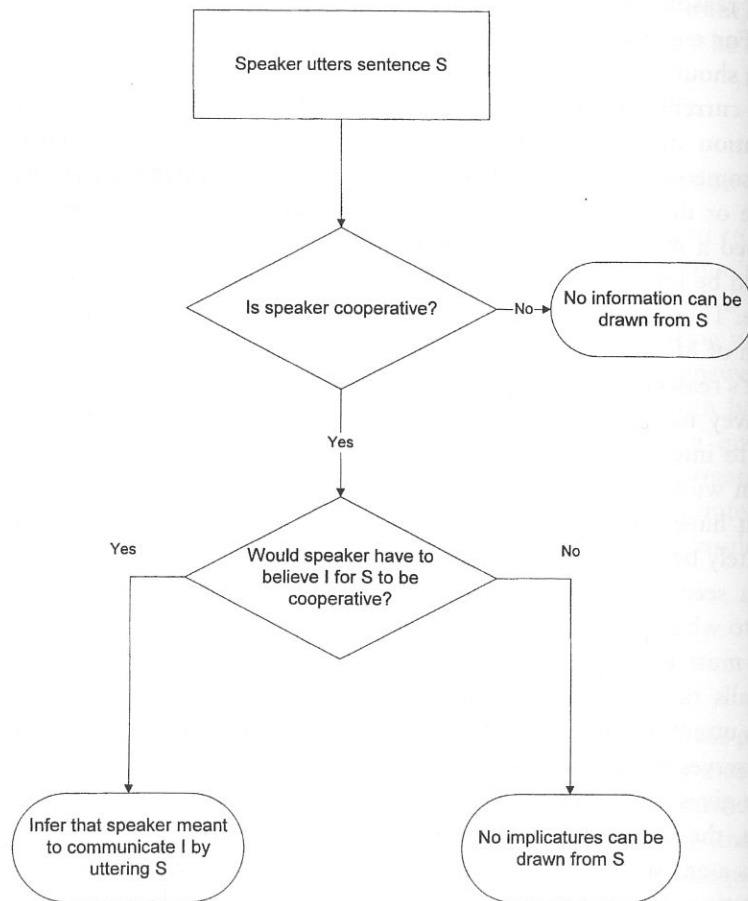


Figure 2.3. Steps in extracting implicatures (I) from a statement (S), according to Grice's cooperative principle.

evidence that a speaker is being uncooperative, then there may be no reason to take what the speaker said seriously. If the speaker is being cooperative, however, we can sometimes guess some belief *I* that the speaker must have (or some belief that the speaker must not have) in order for the assumption of cooperativeness to hold. That information is the conversational implicature. If the speaker violates a maxim for some reason (while still remaining cooperative), that often gives us a clue to what the belief might be. The belief that we attribute to the speaker is one that would best explain why he or she violated the maxim. It's not necessary that the speaker violate a maxim, however, to get an implicature across. In the gas station example, no maxim was violated. Still, flouting a maxim can provide clues to what the speaker intended to convey.

Implicatures in Surveys

In the context of surveys, implicatures often arise because the respondents assume (correctly or incorrectly) that the survey designer intends a certain feature of the question to be relevant to their response. That is, respondents assume that the question obeys the maxim of relation (Be relevant!) and thus conveys information *I* that they should take into account in determining their answer. Respondents may even lean on this assumption more heavily in survey interviews than in other forms of discourse; "after all, [the survey designer] prepared and edited the question carefully, and since he knows [respondents] have no way of getting clarification, he must think a question won't need clarification" (Clark & Schober, 1992, p. 27).

For example, respondents may assume that two very similar items convey different questions when they appear in the same part of the questionnaire. The similar items would otherwise be redundant, violating the maxim of relation and the cooperative principle as a whole. To demonstrate this point, Strack, Schwarz, and Wänke (1991) asked respondents to rate both their happiness and their satisfaction with their lives. In one condition, the questions – *How happy are you . . . ?* and *How satisfied are you . . . ?* – appeared next to each other in the same questionnaire. In a second condition, one of these questions appeared as the last item on an initial questionnaire, the second as the first item on a second questionnaire. In line with the prediction, ratings for the questions tended to be more similar when they appeared in *different* questionnaires than in the same one. According to Strack and his colleagues, respondents exaggerate small differences in meaning between happiness

and satisfaction when these questions appear in the same questionnaire so that the questions don't fail the cooperative principle. Chapter 7 on context effects in attitude surveys reviews other examples in which the order or grouping of survey items produces unintended Gricean implicatures.

Clark and Schober (1992) and Schwarz (1996) review findings that suggest that many incidental features of survey questions can produce similar implicatures and so affect responses. The overlap in meaning between survey items, the items' sequencing, the numerical range of the response alternatives, and the range and labeling of ratings scales can all produce implicatures about the survey designer's intent in asking the question, and these in turn shape the way respondents think about the questions. There are, however, many properties associated with each survey item; it is unlikely that respondents attend to all of them. Although some incidental aspects of a question (e.g., the range of numerical values on a rating scale) produce implicatures, respondents must ignore many properties, given normal limits to their ability to process information. They can't overinterpret *everything*. It seems reasonable to suppose that respondents have expectations about the amount of information they can extract from an item in a survey context. They may therefore continue processing the question until this expectation has been achieved or until further effort appears futile (see Sperber & Wilson, 1986). But apart from such general notions based on the respondents' limited capacity and interest, we currently have no method for predicting which elements of a question will produce a pragmatic effect. And without predictability, it is a little too easy to assign survey errors to pragmatic factors. Nearly any effect of wording can appear to be an implicature after the fact. As we'll see, a similar difficulty arises in extending pragmatics beyond implicatures to other aspects of the survey interview.

2.5.2 Pragmatics of Interviews

Grice's theory necessarily views comprehension as a process that's more complicated than passively decoding language: It includes a reasoning component through which listeners and readers infer the ideas behind the sentences. It is possible to go even further, however, and to describe comprehension as an active social endeavor.

Ongoing conversation is a dynamic process in which speakers and listeners collaborate to make sure they understand what's been said, at

least well enough for the purposes at hand. Speakers pause to ensure that their listener is still tracking, and listeners provide feedback to the speaker using backchannel phrases like *uh-huh*, *yeah*, or *okay* to signal that they're following. As Graesser and his colleagues (1992) note, a major use of questions in conversations is to elicit just such feedback. If the listeners have difficulty understanding what's been said, they can ask the speaker to clear up the problem before going on. Clark and Schaefer (1989) call this mutual checking process *grounding* and have shown that it is a feature of casual, spontaneous conversation. The opportunity to do this sort of checking is a benefit that conversation has over the more formalized comprehension that occurs in listening to a lecture or to TV. It also provides an advantage to conversational participants relative to passive overhearers of the same dialogs (Schober & Clark, 1989).

Compared to normal conversations, the survey interview is likely to seem stilted, resembling ordinary conversation in its one-on-one format but, at the same time, limiting drastically the conversational moves participants can make. Suchman and Jordan (1992) argue that these deviations have detrimental effects on survey quality. They argue that the regulations surrounding surveys straitjacket interviewers, keeping them from interpreting answers flexibly and from adapting questions to the respondents' needs.

The standard texts on survey interviewing caution against giving interviewers too much discretion (see, e.g., Fowler & Mangione, 1990). In many surveys, interviewers are not supposed to offer clarifications on their own initiative or to give much help to respondents who ask for aid in interpreting a question. In the interest of standardization, interviewers are often trained to repeat the survey item verbatim when the respondent asks for clarification and to deflect any further queries that respondents make.⁵ According to Suchman and Jordan (1992), this lack of responsiveness threatens the validity of survey questions by masking differences

⁵ At least officially. Some interviewers may simply ignore these restrictions. Others may find their way around them through paralinguistic cues, such as intonation and gestures, or by repeating a question if the respondent gets it wrong. Schober and Conrad (1997) offer the following example of a trained interviewer clarifying the respondent's uncertainty about the meaning of *bedroom* (words in all caps are emphasized):

Interviewer: How many bedrooms are there in THIS house?

Respondent: Uh, there are two bedrooms. And one den is being used as a bedroom.

Interviewer: How many BEDROOMS are there in this house?

Respondent: Two.

Interviewer: (continues)

between the meaning that the survey designer intended and the meaning the respondent extracts. Clark and Schober (1992) similarly contend that when survey interviews suspend normal processes of grounding, they force respondents to make arbitrary, possibly idiosyncratic, assumptions about the questions' meaning.

The Impact of Standardization on Comprehension

Some evidence for this point of view comes from a recent experiment by Schober and Conrad (1997) that deliberately varied interviewers' freedom to interact with the respondents. Experienced interviewers asked respondents 12 questions drawn from national surveys, such as *Has [name] purchased or had expenses for meats and poultry?* Each question included key concepts (e.g., *meat* and *poultry*) with official definitions that the interviewers knew but that were not included in the question itself. (In case you're wondering, *meats* and *poultry* include "beef, lamb, pork, game; organ meats, such as kidney, sweetbreads, chitterlings, heart, tongue; sausages and luncheon meats; poultry, such as chicken, turkey, pheasant, goose, duck. Include canned ham. Do not include other canned meats and canned poultry, or any prepared meats and poultry.")

Half of the interviewers in the experiment used the techniques prescribed by the Department of Commerce's *CPS Interviewing Manual*, asking questions verbatim, rereading questions if necessary, but not providing the definitions. The remaining interviewers also had to read the questions verbatim, but they were free to say anything that could help the respondent understand the question according to the survey designer's intent. The "respondents" in the experiment had been instructed to base their answers on fictitious information that the experimenters had given to them prior to the interview. Half of the time, this information allowed the respondents to answer the question on the basis of typical instances of the key concepts. For example, they answered the question about meat and poultry on the basis of their purchase of beef. The rest of the time, the information was atypical of the key concepts but clearly fell under the official definition or outside the official definition. On such a trial, for example, the respondents might have had to answer the meat-and-poultry question on the basis of their purchase of canned chicken. The experimenters measured the amount of time it took subjects to answer the question and measured their accuracy, as judged by the official definitions. When the stimulus information was typical (e.g., hamburger), both flexible and standardized interviewing yielded

highly accurate responses (98% and 97% correct, respectively). When the information was atypical, however, flexible interviewing led to many more accurate responses (87%) than did standard interviewing (28%). There was a drawback to flexibility, though; flexible interviews were more than three times longer than standardized ones.

Limitations on Flexible Interviewing

These results demonstrate that accuracy improves when interviewers know the official definitions of the key terms and can inform the respondents about them. They leave open the question of whether flexible interviewing is also helpful when interviewers are themselves uncertain about the questions' thrust. This latter situation must arise fairly often in interviews, since many words in natural language are inherently vague, as we noted earlier. Not all the key terms in a question will have definitions to make them precise. (We now know the official meanings of *meat* and *poultry*, but what about *had expenses* in the sample question, *Has [name] purchased or had expenses for meat and poultry?*) Even when definitions are available, they probably do not cover all situations that arise, since the categories mentioned in the definition are themselves open-ended. What, for example, are the boundary conditions for *prepared meats or poultry*? It is unclear whether flexible interviewing will produce greater stability or greater variability when both respondent and interviewer are uncertain about the fit between the question and the respondent's situation. Variations across interviewers can be a very large source of error in survey estimates; that's why survey researchers attempted to standardize interviews in the first place (see Tourangeau, 1990).

Moreover, some of the arguments for flexible interviewing depend on some controversial premises. Critics of the standard survey interview (Clark & Schober, 1992; Suchman & Jordan, 1992) assume that words and phrases in questions have no inherent meaning (or that if such inherent meaning exists, it is irrelevant in understanding survey outcomes). They argue that the meaning respondents extract is always a product of the survey situation in which they find themselves. According to Clark and Schober (1992), effects of question wording seem mysterious because of "the common misconception that language use has primarily to do with words and what they mean. It doesn't. It has primarily to do with people and what *they* mean. It is essentially about speakers' intentions. . . ." (p. 15). Similarly, Suchman and Jordan (1992) assert that "the meaning of an utterance does not inhere in the language, but

is a product of interaction between speakers and hearers" (p. 256). If the meaning of a survey question must be constructed on the fly through negotiation between the interviewer and the respondent, then standardized interviewing, which deliberately hamstring this negotiation process, leaves the meaning of the question indeterminate.

No doubt, respondents draw inferences about the intended meaning of the question, but does this imply that meanings are underdetermined in settings where grounding is restricted or impossible? If meanings have to be continually renegotiated by participants, then it would be impossible for us to understand lectures, speeches, TV shows, movies, plays, and other forms of noninteractive speech. We couldn't read a book or understand a newspaper or magazine, since we cannot negotiate meanings with the writers in any significant way. We seem forced to conclude that if comprehension of questions and statements includes what goes on when we listen to speeches and read books, then some large portion of comprehension must be independent of ongoing collaboration. It is possible to maintain that word and sentence meanings depend on cultural processes and facts about language use, but these processes and facts must nevertheless be stable enough to support interpretation in noncollaborative settings.

A related issue for these interaction views is that there are clear limits on the negotiability of the meaning of questions. The collaboration between the interviewer and the respondent might converge on some novel interpretation of a word or phrase during the question-answering process, but not everything goes. Sometimes these ad hoc meanings will be incorrect. Consider a hypothetical survey interview that proceeds along these lines:

Interviewer: Has Martha purchased or had expenses for meats and poultry?

Respondent: What exactly do you mean by *meat*?

Interviewer: Okay, meat and poultry include beef, lamb, pork, game; organ meats, such as kidney, sweetbreads, chitterlings, heart, tongue; sausages and luncheon meats . . . (reads rest of definition)

Respondent: How interesting. So chitterlings are organ meats. I never would have guessed. But what about peanut butter?

Interviewer: Peanut butter? You think peanut butter is an organ meat?

Respondent: No, not an organ meat. But it is the protein equivalent of beef and pork, so it seems fairly similar to the items on your list.

Interviewer: I suppose that's true. Let's see, it doesn't mention peanut butter here. I'm not sure, um, . . .

Respondent: I know that Martha bought a jar of peanut butter a couple of days ago, so I guess she did have expenses for meat and poultry.

Interviewer: Well, allrighty then. (continues with the next question)

Surely this type of agreement is possible, and we could say that for this pair of people meat and poultry includes peanut butter (in the context of this exchange). But most of us have the strong intuition that it's simply mistaken to think that the meaning of meat and poultry extends to peanut butter, no matter what these participants decide (and no matter what's in the official definition or what the question designer intended). This intuition, along with the possibility of noninteractive uses of language, should make us cautious about the idea that what really matters in the survey context is what the participants intend or mutually decide to believe. Pragmatic inferences about questions are still inferences (part of the respondents' representation about the question), and they can be incorrect in the same way as other nondeductive conclusions. It might be best to view these conclusions not as constituting the question's meaning, but as supplementing a more stable framework of meaning that is nonnegotiable.

These considerations do not mean that flexible interviewing is not an improvement over current survey practice, and we certainly do not mean to discourage researchers from studying the pragmatics of survey interviews. We do think, however, that the costs and benefits of flexible interviewing will have to be determined by empirical research; they are not an inevitable outgrowth of the nature of language use. And, as with most attempts to improve survey practice, flexible interviewing is likely to involve trade-offs. In this case, improved comprehension is likely to come at the price of longer interviews with better-trained interviewers who will now need to learn the official definitions for all the key terms.

2.6 Summary

Questions are complex linguistic constructions, and understanding them requires nearly all of our language skills. At the level of English grammar, questions pose special difficulties. Depending on their form, they may require burdensome processing to determine the focus of the question. Attempts to clarify key concepts or to achieve balance in attitude questions can lead to complicated questions that overload the working memory capacity of some of the respondents. At the level of meaning,

questions can be too restrictive, carrying presuppositions that may be inappropriate for some listeners. They can also be too unconstrained, including imprecise words and phrases that make a relevant answer difficult. At the level of language use, questions can present incidental features that listeners overinterpret, especially in the context of survey interviews, where respondents receive little guidance from the interviewer.

These difficulties fall into two functional groups that require different remedies. Some difficulties of grammar and meaning make it hard for respondents to build a basic representation of the question. If respondents can't figure out which part of the sentence is being queried (as in ambiguous examples such as (8a)), then they will be unable even to begin answering the question. This sort of problem requires careful pretesting, perhaps using online techniques to discover exactly where in the sentence respondents begin to go astray.

The other sort of difficulty arises from respondents' tendency to use inference to supplement a question. These inferences can occur at all levels of question interpretation, from hypotheses about grammar to hypotheses about the questioner's intentions. In some cases, these inferences may do no harm, even when they are incorrect. For example, it's at least possible that the different interpretations respondents give for complex words like *violence* or *ill effects* in Question (3) may be no worse than the usual problems we all have in defining terms that have no exact synonyms. Such problems do not necessarily mean that we have not grasped the concepts correctly. Nevertheless, there is plenty of evidence to suggest that inferences do sometimes lead respondents to misinterpret questions and to produce discrepant answers. Although we can sometimes use general pragmatic principles to describe after the fact how these inferences arose, we don't currently have an adequate way to predict them. Detecting the inferences may require educating interviewers to detect and cancel the inferences when they arise, but we need to beware of the possibilities of *foibles à deux*. Perhaps the right approach is pretesting with a sample of respondents about whom the relevant background information is independently known, since we then have a direct comparison between the answer to the question we intended to ask and the answer to the question the respondent understood.

Survey texts sometimes offer specific pointers for writing survey questions that avoid the problems discussed here. The guidelines they offer are generally consistent with the evidence and theoretical analyses presented in this chapter (see Bradburn & Sudman, 1979, and Converse &

Presser, 1986, for two particularly good examples). Here's our own attempt to distill the implications of our discussion of the comprehension of survey questions into practical advice for questionnaire designers:

- Keep questions simple;
- Avoid complicated syntax, including adjunct wh-questions and constructions with embedded clauses;
- Decompose questions that cover multiple possibilities into simpler questions that cover a single possibility apiece;
- Define ambiguous or unfamiliar terms;
- Avoid vague concepts, and provide examples when such concepts must be mentioned;
- Replace vague quantifiers with ranges that specify exact probabilities, frequencies, and so on;
- Train interviewers to recognize and repair misunderstandings.

As this chapter has already made clear, there will be exceptions to each of these rules, potential conflicts among them, and trade-offs involved in implementing them. As a result, crafting good survey questions is likely to remain an empirical enterprise, one that requires the testing of draft questions against the data from pretests.