

CHAPTER 3

A Collaborative View of Standardized Survey Interviews

Michael F. Schober and Frederick G. Conrad

What role should interviewers play in standardized surveys? Under the prevailing view, interviewers should be little more than purveyors of scripted questions; survey data are most valid when interviewers are least likely to have biased the answers, or at least to have biased them in different ways. This view, we propose, rests on problematic underlying assumptions about communication: that words contain meaning and that sameness of words entails sameness of meaning. Here we examine these assumptions more closely and demonstrate how interviewers can influence responses even when they follow the most strictly standardized prescriptions to the letter. We present a more collaborative view of interviewer–respondent interaction and explore how and when it might be beneficial to adopt such a view.

STANDARDIZING WHAT INTERVIEWERS SAY

Most large-scale surveys—including the one on which this volume focuses—try to standardize what interviewers say. As described by Fowler and Mangione (1990) in an influential work on the theory and practice of standardized survey interviewing, interviewers must read questions exactly as worded, they must probe neutrally, and they must never allow their own ideas to influence the respondents' answers. According to Fowler and Mangione (1990), this practice reduces or eliminates what they call *interviewer-related error*—any systematic effect of particular interviewers on survey responses.

Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview
Edited by Douglas W. Maynard, Hanneke Houtkoop-Steenstra, Nora Cate Schaeffer,
and Johannes van der Zouwen
ISBN 0-471-35829-0 Copyright © 2002 by John Wiley & Sons, Inc.

These prescriptions reflect how most researchers have conceived of the sources of measurement error in survey interviews. Under the usual view, various kinds of error contribute to overall measurement error: interviewer error, respondent error, sampling error, problems of question design (including wording problems), and errors in coding or classifying responses (see Groves, 1989, 1991). Some of these kinds of errors are more under researchers' control than others. Sampling, question design, and interviewer training are up to the researchers, but other kinds of error are less clearly controllable, for example, those due to the accuracy of respondents' memories or their willingness to respond honestly.

Under this view, the different kinds of error are seen as independent (although they can affect one another), and they are reduced by different methods. Sampling error can be reduced by improved statistical methods. Question design problems can be reduced through laboratory pretesting. Interviewer error can be reduced through standardized interviewing practices. Classification errors can be reduced by more strictly monitoring or automating the process of classifying the data.

Error due to the respondent, while not directly controllable, can be minimized by reducing the other sorts of error. So if questions have been well pretested, respondents should be less likely to misunderstand them. And if interviewers behave in a standardized way, that is, expose respondents to the same question stimulus, researchers can be confident that differences in the answers stem only from the respondent, and not from anything in the interviewer's behavior (Fowler and Mangione, 1990:14–15).

THE MESSAGE MODEL OF COMMUNICATION

By completely separating interviewer behavior, respondent behavior, and question wording, this prevailing approach relies on a view of communication that has been discredited, at least for ordinary spontaneous conversations. This view, which dates back to John Locke or even earlier, has been called, variously, the *message model* (Akmajian et al., 1990), the *conduit metaphor* (Reddy, 1979/1993), and the *meaning-in-words assumption* (Schober, 1998). On this view, speakers encode their thoughts into linguistic messages and send these messages to recipients, who decode them into their own thoughts. Thoughts and conceptual material are thus transferred from one head to the other via words.

Although the message model captures most people's intuitions about how communication works, it cannot account for all of what goes on in ordinary conversations (see also Clark, 1992, 1996; Gibbs, 1994; Maynard and Whalen, 1995). The problem is that the message model assumes that the meaning of speakers' words is in the words themselves. But it isn't. Rather, the same words can have vastly different meanings depending on what speakers intend them to mean on particular occasions (see Akmajian et al., 1990). In fact, meaning is even more complicated than this: Utterance meaning depends not only on what speakers mean on particular occasions but also on how addressees take those meanings — what speakers and addressees together determine the words mean. If this is true, the meanings that a speaker and addressee jointly determine may

differ from the meanings inferred by someone who isn't participating in their conversation.

An empirical demonstration of this is found in a laboratory experiment by Schober and Clark (1989). In the study, pairs of strangers performed a matching task. One person, the *director*, saw an array of unfamiliar abstract geometric figures. The other person, the *matcher*, sat behind a screen with a scrambled set of the same abstract figures. The task was for the director to get the matcher to arrange his figures in the same order as the director's figures, saying anything it took for the two to perform the task. Unbeknownst to the director and matcher, a third person, the *overhearer*, listened to every word they said to each other from the moment they met. As the director and matcher conversed to arrange the figures, the overhearer tried to arrange the same set of figures based on what she heard.

The logic of the study was this: If a speaker's meaning is contained in the words themselves, then overhearers should match the directors' (speakers') array of figures at least as well as matchers do, because the overhearers hear every word the matchers (the addressees) hear. In fact, overhearers may perform better than matchers, because the overhearers don't have to expend any attentional resources on letting directors know that they have understood. In contrast, if speakers and addressees jointly determine the meaning of what is said (i.e., if comprehension requires participation in the interactive process), then overhearers should not match the figures as well as the participating matchers, even though they have heard every word uttered from the moment the director and matcher have met.

The results were plain: Conversational participants (the matchers) matched figures with nearly perfect accuracy. Nonparticipants (overhearers) matched figures reliably less accurately. And this continued to be the case over time. After directors and matchers completed the task, the directors' figures were reshuffled and the director and matcher began again. Overhearers' performance was poorer than the matchers for even the sixth reshuffling, even though they heard everything the conversational participants said. And another set of overhearers listening to tape-recordings of the conversations, who could press the pause button and stop and think as long as they liked, performed no better.

A message model can't account for findings like these. If meaning is in the words themselves, overhearers should understand at least as well as addressees. Because on some occasions they don't, the conclusion must be that meaning is not only in the words; the words themselves do not always bring speakers' and listeners' conceptualizations into sufficient alignment [for further discussion see Schober (1998)]. We need an alternative model to explicitly include what both parties in a conversation do together.

One proposal is Clark and Wilkes-Gibbs' (1986) *collaborative* model (see also Clark, 1992, 1996; Clark and Brennan, 1991; Clark and Schaefer, 1987, 1989; Schober and Clark, 1989). This model generalizes observations by Grice, Schegloff, and others, bringing them into the psychological realm in ways that can be modeled and tested precisely (e.g., Cahn and Brennan, 1999; Traum, 1994). Under this approach, as people speak they carefully monitor their addressees for evidence of understanding or misunderstanding, and they adjust their utterances,

moment by moment, to ensure that their addressees are understanding them well enough for current purposes. Addressees, by providing such evidence, help mold the utterances speakers produce.

Speakers and addressees thus coordinate in speaking much as dance partners coordinate in dancing or acquaintances coordinate in shaking hands (see Clark, 1992, 1996; Clark and Brennan, 1991). Consider shaking hands. When you extend your hand to initiate a handshake, you can't begin the handshake until your partner has extended her hand in just the right way. You can only continue shaking hands as long as your partner is also engaged in the activity in the right manner, with appropriate hand pressure and movement. You adjust the force and character of your handshake to coordinate with what your partner is doing. And you can only stop shaking hands once your partner implicitly agrees to stop. The argument is that language use is much like a handshake or any other kind of joint action: Both parties together continually adjust their behavior on the basis of what their partners do.

On this view, no utterance is complete until it has been *grounded*—until both participants have accepted that it has been understood. Understanding a reference in any particular utterance requires active participation by both speaker and addressee, and this can take several turns. Note that the point isn't that words don't have conventional meanings; the conventional meanings of words provide important constraints on speakers' meaning. But speakers regularly use words in idiosyncratic ways that go far beyond dictionary definitions (see Clark, 1991; Clark and Gerrig, 1983). Speakers produce utterances based on their *common ground* with their conversational partners—that is, what they presume that they and their conversational partners mutually know, believe, and assume.

THE MESSAGE MODEL AND STANDARDIZED INTERVIEWS

So the message model doesn't sufficiently account for ordinary conversations. [See Akmajian et al. (1990) for other reasons that a message model is both theoretically and empirically untenable for describing language use. For further arguments about how meaning is socially or interactionally constructed see Button (1987), Cicourel (1973), Goffman (1981), Goodwin (1981), Gumperz (1982), Heritage (1984), Krauss and Fussell (1996), Maynard and Whalen (1995), Rogoff (1990), Sacks, Schegloff, and Jefferson (1974), Schiffrin (1994), among many others.] But could the message model hold for survey interviews?

On the face of it, it could: Standardized survey interviews differ from ordinary conversation in several ways that have been documented elsewhere (see, e.g., Clark and Schober, 1991; Schaeffer, 1991; Strack and Schwarz, 1992; Suchman and Jordan, 1990, 1991). The message model might hold in standardized surveys because survey interviewers (according to the theory of standardization), unlike ordinary speakers, use words carefully scripted to be interpretable by the general population. Enough respondents may understand questions as intended, given that the questions have been carefully pretested, that a message model really can describe communication in survey interactions. In a large sample survey,

misunderstandings may be sufficiently rare that noncollaborative, standardized exchanges lead to adequate responses overall.

Evaluating the message model's applicability to standardized surveys requires closer examination of standardized interviewing techniques. In strictly standardized interviews, interviewers are warned that interpreting questions for respondents often can change the meaning of the question, and so they are required to follow standardized *neutral probing* techniques. Consider these sample probes from the interviewer training materials from the University of Wisconsin Survey Center (October 1994), which are consistent with Fowler and Mangione's prescriptions. Interviewers are advised to say:

"Is there anything else?"

"Yes, I see" or "Uh-huh" in an expectant manner and followed by a pause.

"Could I read back what I have taken down to be sure I have exactly what you wanted to say?"

"What do you mean by that?"

"Could you be a little bit more specific?"

"Which comes closest to the way you feel?"

If a respondent asks what a question means, the interviewer can say:

"Whatever it means to you."

"I'm sorry, but I don't have that information."

"It's important that the question be answered as best you can, in terms of the way it's stated. Perhaps I could read it to you again."

These probing techniques are designed to prevent the interviewer's point of view from intruding on respondents' interpretations of questions (and thus their answers). But a little reflection shows that they nonetheless can. Each of these techniques can provide some information to the respondent about how the question should be interpreted—they aren't entirely neutral (see also Marlaire and Maynard, 1990; Maynard and Marlaire, 1992; Schaeffer and Maynard, 1996, this volume, Chapter 12; Suchman and Jordan, 1990, 1991). (Whether the information potentially provided is *harmful* or not we will take up later).

If an interviewer says "Is there anything else?" she provides the respondent with some important information, namely that it is entirely possible that the answer the respondent has thus far provided is insufficient, and that the respondent should work to elaborate the answer if possible. In contrast, if the interviewer doesn't use this neutral probe, she may have signaled that what the respondent provided thus far is sufficient. The same logic holds for saying "I see" or "uh-huh" in an expectant manner, which can signal that more is expected of the respondent. While this may seem to be an innocuous intervention, it can actually substantively mold the response; if the interviewer didn't say "uh-huh" expectantly, the respondent might infer that the initial response was sufficient (see Cannell, Miller, and Oksenberg, 1981, for a discussion of how other types of interviewer feedback may be able to improve response accuracy).

Other neutral probes don't simply request more information; they suggest that the respondent ought to provide a *different* answer than he has already given. If an interviewer, following the suggestion in the training materials, says "Could I read back what I have taken down to be sure I have exactly what you wanted to say?" this can suggest that there is something inadequate about the answer. Saying "What do you mean by that?" more directly implies that the answer is unclear or unacceptable and requires the respondent to present more.

Even those probes that seem as if they *must* be neutral—the "whatever it means to you" sort that inform the respondent that the respondent's own interpretation of the question is required—provide real information about question interpretation. By using such probes, interviewers indicate that the respondent's own idiosyncratic interpretation of the question is the desired one. This can lead to respondents interpreting questions quite differently than survey designers intend (see Belson, 1981, 1986; Schober and Conrad, 1997). Take this example:

(1) AW01:620

```

1  IV:  ... and >now we'd would like to ask about< y↑our
2        ↑employment ↑status .hhh did you do any work (.) for pay
3        (.) last week {q359}
4        (1.1)
5  FR:  °.hh° Well
6        (0.6)
7  FR:  I'm still getting paid but school's out (.) So
8        (0.7)
9  IV:  Okay s:○ (.) would you say (0.4) I mean (0.3) [it's=
10 FR:                                     [well
11 IV:  it's your c[all
12 FR:                                     [I got paid (.)
13 FR:  [for work (.) But I was]n't at work
14 IV:  [Ok(h)ay(h) huh huh huh huh]
15 IV:  .hhh o↑kay (.)

```

Here the respondent doesn't answer the question with a "yes" or a "no," but rather describes her circumstances ("reports," in Drew's [1984] terms; see also Schaeffer and Maynard [1996]), implicitly leaving the interpretation of "work for pay" up to the interviewer. The interviewer refuses to interpret "work for pay" for the respondent, telling her that "it's your call" (line 11). This indicates that the respondent's interpretation of "work for pay" is adequate. The respondent answers that she was "paid for work," which the interviewer takes as a "yes" answer (we can tell this from the subsequent questioning). The interviewer's neutral probe molds the respondent's answer such that she picks one interpretation. (Later in the interview, it turns out that the respondent's personal interpretation of "work for pay" here was not appropriate, and the response must be changed.)

So interviewers, following standard neutral probing techniques, provide evidence for respondents of whether they have interpreted questions appropriately, and thus whether they have provided sufficient responses.

The issue here is an even larger one. It isn't just that neutral probing techniques aren't really neutral: Interviewers can influence answers even when they do *not*

probe (see also Marlaire and Maynard, 1990; Maynard and Marlaire, 1992, for discussion of this phenomenon in the domain of standardized testing). Consider this example, where all the interviewer does is accept the respondent's answer by saying "okay":

(2) AW02:691

```

1  IV:  [°Thee(gh) U S .hhh ar-my° .hhh what wr:- what were]=
2      [ ((typing)) ]
3      =your [most important activities? or duties on this=
4      [#
5      =jO:b {q359f}
6      (1.8)
7  MR:  uh:: This past week I had to lead uh: my
8      [platoon out to uh (0.2) field exercise]
9      [ ((typing)) ]
10 IV:  [.hhh Oka(gh)::y]
11      [((typing)) ]
12      (3.7) ((typing))
13 IV:  Field exerci(gh)::se:
14      (5.1) ((typing))
15      [((typing))]
16 IV:  [.hhh ] -Oh kay:

```

By saying "okay" (line 16) the interviewer seems to indicate that the respondent's answer was sufficient. In conversational analysis terms, *response tokens* such as "okay" and "uh-huh" suggest that a speaker was understood and doesn't need to initiate a repair. In the second example, if the interviewer hadn't said "okay" and had asked for further or different information, the respondent would almost certainly have produced a substantively different answer.

Even when interviewers say absolutely *nothing*, they are informative about how questions should be interpreted. In the next example, after the respondent provides an answer in line 8 ("repair and maintaing payphones") the interviewer types the answer and goes on to the next question:

(3) AW21:643

```

1  IV:  ((typing)).hh And what were your most important
2      activities or duties on this job? {q359f}
3      (3.7)
4  MR:  Mos' importan'
5      (0.7)
6  MR:  .hhh duties er activities=
7  IV:  =mhmm
8  MR:  repair and maintaing (.) -payphones
9      (8.5)
10 IV:  ((typing)).hh Do you work for yourself in a family
11      business or for someone else. {q359h}

```

Merely by going on to the next question without comment the interviewer implicitly indicates that the answer was appropriate for current purposes (see also Button, 1987; Heritage and Sorjonen, 1994; Schaeffer, 1991, among others). So even when interviewers don't probe, they indicate whether respondents have

interpreted questions appropriately. The interviewer's behavior is informative in light of what *else* the interviewer might have done — the evidence the interviewer did *not* give that the answer was inappropriate.

All this shows that even the most seemingly neutral interviewer behaviors may not be neutral (see also Schaeffer and Maynard, 1996; Schober and Conrad, 1997:592–593). As conversational analysts have pointed out in nonsurvey contexts (e.g., Button, 1987; Marlaire and Maynard, 1990), in standardized interactions, everything a questioner says—and doesn't say—can affect the respondent's interpretations. This leads us to conclude that the message model of communication—a model that assumes that interviewer and respondent behaviors are separable—does not sufficiently characterize survey interactions to be a theoretically appropriate account.

CONSEQUENCES OF ADOPTING A COLLABORATIVE MODEL

What are the practical consequences of changing our underlying assumptions, of adopting a collaborative model and partitioning the measurement error differently? Does changing our assumptions about meaning require us to change what we train interviewers to do?

The answer to this question isn't straightforward; in fact, there is a range of possible answers. To take one extreme position, one could argue that absolutely strict standardization should be the norm. Even if the message model is inappropriate for survey interviews, the prescriptions of standardization are nonetheless practically useful, or even collaboratively appropriate. Perhaps the standardized prescriptions lead to desirable interactions and deviations from standardized practice lead to undesirable interactions. This, of course, requires elaboration of what counts as a desirable interaction.

To take an alternative extreme position, one could argue that the prescriptions of standardized interviewing should be entirely rethought. Since standardization's underlying model of communication is problematic, interviewers should be trained to use the full range of conversational resources that ordinary speakers use to make sure they are understood, even if this requires deviating from an official script. If there is no longer such a clear dividing line between those interviewer behaviors that affect a response (unlicensed probes) and those that supposedly don't (neutral probes)—because every interviewer behavior can influence responses—then perhaps no probes should be out of bounds.

To evaluate these two extreme positions, and the range of possibilities in between, we first examine standardized interviewer behaviors, sanctioned and unsanctioned, from a collaborative viewpoint. To do this, we outline further details of a collaborative view of spontaneous (nonsurvey) conversations. Then we examine interviewer behaviors in standardized surveys from this viewpoint. This analysis shows that the answer to our question—are the prescriptions for standardized interviewing appropriate?—depends on how informative interviewers' probes are about question meaning.

A Model of Grounding

Conversational partners continuously provide each other with evidence of whether the conversation is on track—whether they are understanding each other well enough for current purposes (see Brennan, 1990; Clark and Brennan, 1991). The conversation can go off track at several levels (Clark and Schaefer, 1987). A speaker can produce an utterance that the addressee doesn't even notice. An addressee can notice that there was an utterance but fail to have heard it. Or an addressee can hear an utterance (understand all the words) but fail to have understood what the speaker meant.

How do people provide each other with evidence of understanding or misunderstanding? They do it by accepting (or failing to accept) each presentation their partner makes. The idea is this: Every utterance a speaker produces constitutes the first part—the *presentation* phase—of what Clark and Schaefer call a *contribution*. For a presentation to become part of what both parties believe has been understood in the conversation, it must be *accepted* by the other party. And then *that* acceptance must be accepted as well.

The contribution structure of a conversation (the details of which are only evident in retrospect) is hierarchical. Consider the following episode, which in Schegloff's (1972) terminology would involve an *insertion sequence* within an adjacency pair. Speaker A asks partner B a question. Rather than answering the question, partner B asks A another question that requires an answer. Only once A has answered B's question does B proceed to answer A's original question. In Clark and Schaefer's terms, this final answer constitutes B's acceptance of A's original question; the question and answer that formed the insertion sequence form their own presentation and acceptance at a lower level of the hierarchy. So acceptances occur at the same level of the conversational hierarchy that their presentations occurred in (for details, see Clark and Schaefer, 1989).

It is only once a speaker and addressee have given each other evidence that a particular utterance has been understood that the content of that utterance officially becomes part of the conversational record, or part of the pair's common ground—the set of beliefs, assumptions and knowledge they can presume they mutually know. When presentations are attempted but aren't accepted, they can't be considered to be a part of common ground; if a speaker assumes that an unaccepted presentation has succeeded, this can lead to misunderstanding.

Speakers and addressees use several techniques to show each other whether they have understood each other well enough for their current purposes—that is, whether they have *grounded* their utterances. By using any of the following five techniques (from Clark and Schaefer, 1989:267) an addressee B shows he has understood what speaker A means; by not presenting any of these kinds of evidence, he shows that the conversation has gone off track.

1. *Continued Attention*. B shows he is continuing to attend and therefore remains satisfied with A's presentation.

2. *Initiation of the Relevant Next Contribution.* B starts in on the next contribution, which would be relevant at a level as high as the current one.
3. *Acknowledgment.* B nods or says "uh huh," "yeah," or the like.
4. *Demonstration.* B demonstrates all or part of what he has understood A to mean.
5. *Display.* B displays verbatim all or part of A's presentation.

As an illustration, consider Clark and Schaefer's example from the London-Lund corpus (p. 270):

- A. How far is it from Huddersfield to Coventry .
 B. um . About um a hundred miles -
 A. So, in fact, if you were . living in London during that period, . you would be closer - .

According to Clark and Schaefer, in saying "um . about um a hundred miles" B presents evidence that she has understood A's question. By initiating the next relevant contribution (by answering the question), B shows that she has understood that A presented a question (as opposed to some other sort of speech act), and the content of her answer shows how she interpreted the question. A must also accept B's acceptance of A's question; here she does this by going on with the next relevant contribution, rather than correcting A's interpretation of the question (e.g., "No, I mean how long does it take to get there?").

Note that with these ordinary grounding techniques *both* parties have input into what is understood. Each party's actions ultimately affect the other's, and we would be missing out on a crucial part of the process if we focused only on what A does or what B does. B's conversational move of responding to the question as she does affects A's actions; if B had displayed any evidence of a problem, A would have repeated the question (if B hadn't heard it) or rephrased it (if B had misunderstood it). A's subsequent acceptance of B's answer similarly affects B's actions; if A hadn't gone on to a relevant next utterance but had demonstrated that he hadn't understood B's utterance, B would have repeated or reframed her utterance.

A final point about the contribution model: Speakers and addressees ground utterances "to a criterion sufficient for current purposes" (Clark and Wilkes-Gibbs, 1986). When both parties desire extremely precise and accurate understanding—say, when a telephone caller wants a telephone number from an operator (Clark and Schaefer, 1987)—people present explicit displays of understanding, like overt "uh-huh"s. When people are engaged in small talk or banter, demonstrating precise understanding may be less important, and less explicit grounding techniques like continued attention may be more prevalent.

The most important point, for our purposes, is that the use of any grounding technique by one party *always* affects what the other party does. No grounding technique is neutral: A partner's response presents evidence about his or her states of understanding and thus shapes what the other party says.

Grounding in Standardized Surveys

Licensed Probes

If we examine Fowler and Mangione's (1990) prescriptions for interviewer behavior from a collaborative viewpoint, one feature is striking. All the neutral probing techniques — with one major exception — are part of speakers' ordinary repertoire of ways to provide evidence of how well they have understood each other. Ironically, standardized interviewers are licensed to provide (or at least not prevented from providing) exactly the kinds of evidence of understanding that Clark and Schaefer (1989) have outlined, and that have been shown to affect addressees' interpretations of speakers' utterances: continued attention, initiation of the relevant next contribution, acknowledgment, demonstration, and display. And, just like ordinary conversationalists, interviewers are licensed to probe when they suspect that respondents have not understood questions as intended or when respondents fail to provide appropriate answers. Consider again those neutral probes listed in the Wisconsin training materials. Each of them implements one or more of the kinds of grounding techniques delineated by Clark and Schaefer (1989). Some correspond to techniques for presenting positive evidence of understanding. For example, when interviewers say "Yes, I see" or "Uh-huh," in an expectant manner and followed by a pause, they are presenting an acknowledgment (technique 3) of what the respondent has said thus far, but by not going on to the next question they are failing to fully accept the respondent's answer and indicating that more is required. When interviewers say "Could I read back what I have taken down to be sure I have exactly what you wanted to say?" they are proposing to display (technique 5) the respondent's answer verbatim, allowing the respondent to correct the display if it doesn't accurately reflect what the respondent intended.

Other neutral probes facilitate grounding because they allow interviewers to give negative evidence, that is, evidence that they are not willing to accept their partner's presentation. When interviewers say "What do you mean by that?" they demonstrate that they are not yet willing to accept the respondent's answer. The same is true when interviewers say "Is there anything else?" or "Could you be a little bit more specific?" The probe "Which comes closest to the way you feel?" can be used when a respondent presents an answer that doesn't conform to a question's response categories, for example, an "I'm not sure" in response to a question about "Do you think the economy is improving, staying about the same, or getting worse?" This probe has the same character as the other negative probes: It shows that the interviewer is unwilling to accept the respondent's current formulation of the answer.

The exceptions to this — the neutral probing techniques that *don't* correspond to the kinds of grounding techniques used in spontaneous conversation — are the scripted responses to requests for clarification: "Whatever it means to you" or "I'm sorry, I don't have that information." In spontaneous conversations, speakers generally take some responsibility for what they mean (Clark and Schober's [1991] *principle of speaker's meaning*). If a speaker's question (presentation) isn't immediately understood (accepted) by the addressee, it is most often

up to the speaker to help clarify what she means. When a standardized interviewer apologizes that she doesn't have the information the respondent needs, she behaves in a way that ordinary conversationalists do not: She is refusing to help clarify the terms of her own question.

The "Whatever it means to you" probe is even stranger—it doesn't even attempt to excuse the interviewer's failure to help. Imagine how odd you might find this invented interchange with a friend:

Friend: How many hours per week do you usually work?

You: Well, that depends...What do you count as work?

Friend: Whatever it means to you.

Your friend's answer would be highly uncooperative; in fact, Grice (1975) would characterize your friend as opting out of the cooperative principle that underlies conversation altogether. You would be justified in retorting "What are you talking about? *You* asked the question!" Whether or not survey respondents perceive these "whatever it means to you" probes as odd, they certainly differ from the other types of neutral probes. They differ because they place the burden of interpreting the interviewer's meaning on the respondent (or rather, to be more accurate, they place the burden of interpreting the *survey designer's* meaning on the respondent—see Clark and Schober, 1991; Schober, 1999; Suchman and Jordan, 1990, 1991) and they prevent the grounding of meaning. The other probes are consistent with the ordinary conversational assumption that speakers are responsible for what they mean. Unlike the "whatever it means to you" probes, they help the respondent to infer whether they have interpreted the questions as the survey designers intended.

Deviations from Standardization

What do deviations from standardization look like from a collaborative viewpoint? Many deviations from standardization seem to serve collaborative functions (see also Houtkoop-Steenstra, 2000; Schaeffer and Maynard, this volume, Chapter 12; Schober, 1999; Schober and Conrad, 1998, Suchman and Jordan, 1990, 1991). That is, interviewers' deviations from the script can be seen as implementations of grounding techniques from spontaneous conversations, often in the form of demonstrations. Take this example:

(4) AW21:370

- 1 IV: mmmm # .hh And how about a year from now, do you expect
- 2 business conditions will be better or worse than they
- 3 are at present, or just about the same {q0080}
- 4 (0.6)
- 5 MR: tch I don't believe they'll improve a whole lot
- 6 IV: mmmm s- (.) So do you think
- 7 (1.0)
- 8 IV: is that worse or about the same
- 9 (1.8)
- 10 MR: I:t (0.4) probably (would) stay about the same [as] it
- 11 is now

The interviewer's probe "So do you think is that worse or about the same" (lines 6–8) is nonstandardized because the interviewer fails to re-present *all* the response alternatives (better, worse, or just about the same). Standardization theory considers probes like this potentially biasing because the meaning of the response alternatives depends on the entire set; by limiting the response alternatives the interviewer is providing an interpretation of what the respondent has said thus far (of "I don't think they'll improve a whole lot"). Such probes can lead respondents to pick the mentioned alternatives more often than they otherwise would (e.g., Schaeffer and Charng, 1991; Smit, Dijkstra, and van der Zouwen, 1997).

From a collaborative viewpoint, on the other hand, the interviewer's probe is simply an example of ordinary grounding behaviors. The interviewer's probe does indeed indicate an interpretation of what the respondent said: It is precisely through this interpretation that an interviewer demonstrates her current state of understanding (technique 4) and helps ground the response. By presenting this demonstration, the interviewer also provides the respondent with the opportunity to reject the interviewer's interpretation.

Here is another example of an unlicensed probe that can be seen as helping ground understanding:

(5) AW21:488

1 IV:h um On how many days during the past week did you
2 not feel like eating-your appetite was poor {q0155}
3 (0.7)
4 MR: I don't have a problem with that
5 IV: hhh .h So zero?
6 (1.0)
7 MR: Zero=
8 IV: =mhmm # .

According to norms of strict standardization, the interviewer has potentially biased the response by proposing the answer ("So zero?" line 5) for the respondent.¹ On a collaborative view, the interviewer has simply demonstrated understanding of the respondent's utterance "I don't have a problem with that" (line 4). The respondent, by repeating "zero" (line 7), then confirms (or at least does not challenge) that the interviewer's interpretation was appropriate; for Clark and Schaefer (1989), the respondent has accepted the interviewer's interpretation through a display (technique 5).

Interviewers' deviations from standardization don't take the form only of demonstrations. In this example, the interviewer uses acknowledgment (technique 3) and display (technique 5) to accept the respondent's interpretation of the question:

(6) AW02:133

1 IV: ... A:nd how much have you heard or read .hhh about the
2 issue: of federal agencies sharing information about
3 individuals .hhh a great deal?
4 (0.5)

- 5 IV: some (0.4) not very much or nothing at all? {q6}
 6 (0.9)
 7 MR: uh:: Lately I haven't heard anything at [all
 8 [# #
 9 [#
 10 IV: [.hhh A:nd how concerned would you say you are about
 11 this issue? (0.3) very concerned somewhat -concerned
 12 not very concerned or not at all concerned? {q7}
 13 (1.0)
 14 MR: Oh you're talkin' about individuals you're (talk-)
 15 like- you would u:se- like my n- my name in and to say
 16 .hhh this person said so and so (.) a:nd the:n (.)
 17 they'd give it to another agency to like follow up or
 18 (0.8)
 19 MR: use for their purposes?
 20 (0.6)
 21 IV: Y[eah
 22 MR: [Is that what you're- (0.3) (l[like)
 23 IV: [That-
 24 IV: [That's what's
 25 MR: [the actual individual
 26 (0.5)
 27 IV: Thee individual right
 28 (3.7)
 29 MR: I:'m not very concerned about it 'cause [I don't-
 30 IV: [huh
 31 MR: the census really can't
 32 (0.8)
 33 MR: .hhh
 34 (0.8)
 35 MR: (it really doesn't come up with)
 36 (2.5)
 37 MR: anything that would be too damaging I guess to a person
 38 IV: Okay? So (0.4) you said not very concerned about it?
 39 (0.5) ((#))
 40 MR: No
 41 IV: .hhh -Okay

When the respondent proposes an interpretation of the question (the utterance that ends "they'd give it to another agency to like follow up or use for their purposes," lines 16–19), the interviewer accepts this interpretation in line 21 with "yeah" (an acknowledgment) and then accepts a further refinement (the respondent's "the actual individual," line 25) with a display ("thee individual right," line 27).

Interviewers also use ordinary grounding techniques when they answer a respondent's request for clarification substantively rather than saying "whatever it means to you":

(7) AW02:104

- 1 IV: .hhh -A:nd is the census used to decide how many
 2 representatives each state will have in congress {q5d}
 3 (1.1)
 4 MR: (n- that-) You said this is a business census?
 5 IV: .hhh No [this is-

- 6 MR: [or just population census?=
 7 IV: =Th- this is a population census the U S
 8 (0.8)
 9 IV: population census done by: (.) the census bureau
 10 (0.6)
 11 MR: Yes I think it-
 12 (0.5)
 13 MR: it's used for uh:
 14 (0.7)
 15 MR: alocatin' seats in the house
 16 IV: Okay:?

By norms of standardization, this interviewer should have repeated the question or left the interpretation up to the respondent; certainly the interviewer should not have defined "the census" (lines 7–9). On a collaborative view, of course, the interviewer has simply taken some responsibility for explaining her own utterance (Clark and Schober, 1991).

So to sum up, a collaborative analysis of interviewer behaviors in standardized surveys shows that most licensed neutral probes, and many prohibited nonstandardized probes, fulfill ordinary grounding functions. The one kind of probing that does not is the "Whatever it means to you" kind of probe.

Interviewer Training from a Collaborative Perspective

What does this analysis suggest for training interviewers? As we noted earlier, accepting a collaborative viewpoint still allows for a range of possible approaches to training interviewers. On one extreme (let us call it approach 1), one might argue that standardized prescriptions lead to desirable interactions and deviations from standardized practice lead to undesirable interactions. On the other extreme (approach 2), one might argue that a collaborative viewpoint requires us to abandon the notion that standardized meaning can be achieved through standardized wording and neutral probes. Since all interviewers' actions influence respondents' interpretations of questions, and thus their responses, no particular ways of clarifying meaning should be forbidden.

The real question for approach 1 is: What constitutes a desirable interaction? The logic underlying standardization is that desirable interactions are those in which interviewers don't bias responses—or at least one interviewer doesn't bias responses differently than another. By this logic, the "whatever it means to you" probes lead to desirable interactions because interpretations are left up to respondents. Paradoxically, our analysis has shown that the other neutral probes can lead to interactions that are undesirable because they are not standardized; interviewers can affect responses idiosyncratically by probing or not probing, by choosing one probe instead of another, or because one probe affects one respondent differently than another. A collaborative viewpoint suggests a very different notion of what constitutes a desirable interaction: Desirable interactions are those in which respondents interpret questions accurately—that is, as intended by the researchers (see Suchman and Jordan, 1990, 1991). By this logic, any interviewer

probes that help respondents interpret questions accurately are desirable. Those probes that don't lead to accurate interpretations are undesirable. So "whatever it means to you" probes are highly *undesirable* because they leave interpretations up to the respondent; any probes that facilitate grounding are desirable.

As for approach 2, if interviewers are licensed to say anything at all to clarify question meanings, we have no idea what the ultimate effects might be. Standardization theorists like Fowler and Mangione claim that the effect would be a disastrous return to all the ills that led to standardized practice in the first place (see Beatty, 1995, for a review). Interviewers, for example, might be more likely to influence respondents' opinions, as Dijkstra's (1987) results (see also Smit, Dijkstra, and van der Zouwen, 1997) suggest can happen when interviewers allow their own opinions to become evident. Perhaps standardization's goal of "holding potentially important influences on answers constant across respondents" (Schaeffer, 1991:377, note) is worth upholding at all costs. But an alternative view (e.g., Mishler, 1986; Suchman and Jordan, 1990, 1991) is that empowering interviewers would have beneficial effects. Who is right?

Corpus-based evidence doesn't resolve the issue. From the Wisconsin transcripts, for example, we can find interactions that deviate from strictly standardized prescriptions and that seem to help ground understanding, presumably resulting in more accurate responses. Take this example:

(8) AW01:907

- 1 IV: .hh And what is the percent chance that your own total
 2 income before taxes will be under forty thousand
 3 dollars in the next twelve months {q461n}
 4 (2.6)
 5 FR: Say that one again
 6 IV: .hhh What [is the percent chance that your own total=
 7 FR: [huh
 8 IV: =income before taxes will be under forty thousand
 9 (0.7)
 10 FR: before taxes? {q461n}
 11 IV: mm hmm
 12 (2.0)
 13 FR: Uh::
 14 (1.3)
 15 FR: So what you're asking is forty- (0.4) A:m I gonna make
 16 forty thousand dollars
 17 IV: .hhh Well they're asking- (0.6) I mean just what the
 18 question says what is the percent chance that your
 19 income before taxes is gonna be under forty thousand
 20 (0.6)
 21 IV: in the upcoming twelve months {q461n}
 22 FR: Zero percent
 23 (0.6)
 24 IV: eh- {(0.4) Okay (.) S:o
 25 [#
 26 (0.6)
 27 IV: there's- (0.4) Do you think it will be under forty
 28 thousand

29 (0.6)
 30 FR: Yes
 31 IV: Then you would want a high percent [<cause there's=
 32 FR: (Oh
 33 IV: =a good chance [(it=ll be-)
 34 FR: [hundred percent
 35 IV: oh(gh) kay(gh) [.hhh it's a confusing section I know=
 36 [## #
 37 IV: =a lot of people get real messed up (h)on th(h)at .hhh
 38 okay what was your age at your last birth day? {q800}

In this example it seems that the interviewer's extremely unstandardized interventions (e.g., "do you think it *will* be under forty thousand?" lines 27–28, and "then you would want a *high* percent," line 31) helped the respondent to provide an accurate answer; if the respondent had not been guided by the interviewer, she would have given a response exactly wrong for what her circumstances warranted.

But for every positive example like this, one can find other examples where nonstandardized interventions may well be having ill effects, obscuring the intended meaning of questions and leading respondents to answer based on inappropriate interpretations. Consider this example:

(9) AW01:698

1 IV: .hh And what were your most important activities or
 2 duties on this jo(gh)::b hh {q359f}
 3 (2.8)
 4 FR: Well you mean of teaching?
 5 IV: .hhh ye- Well I mean I know I just need you to tell me
 6 (.) so I can put it in the computer
 7 (0.5)
 8 IV: What do you do
 9 (0.7)
 10 IV: in your jo(gh)::b what do you do teaching
 11 (1.1)
 12 FR: (h)Oh:: [well] Tha- that could take a long time
 13 IV: [hhh]
 14 IV: Okay [so-
 15 FR: [You mean you wanna know all the things that I
 16 teach?
 17 IV: No: let's say top three
 18 (1.0)
 19 FR: [uh::m:
 20 IV: [jst-
 21 (0.6)
 22 IV: What are your: activitie[s
 23 FR: [In health? I- eh:: my top
 24 [three things I probably teach are: .hhh drugs (h)]
 25 { ((typing)) }
 26 (1.2) ((typing))
 27 FR: [sex education
 28 [((typing))]
 29 (1.8) ((typing))

30 FR: [uh:::m]
 31 [((typing))]
 32 (0.5) ((typing))
 33 FR: C P R?
 34 (1.0) ((typing))
 35 IV: Okay?
 36 (1.4) ((typing))
 37 FR: [a::nd in physical education:]
 38 [((typing))]
 39 (0.6) ((typing))
 40 FR: [.hhh Oh gosh I teach every sport]
 41 [((typing))]
 42 (0.5) ((typing))
 43 IV: Okay? [So teaching sports that's: good] enough
 44 [((typing))]
 45 FR: [Okay
 46 [# #
 47 (4.0) ((typing))
 48 IV: .hhh A::nd do you work for yourself in a family business
 49 or for someone else:? {q359h}

This interviewer explicitly defines "most important activities or duties" as meaning the top three (line 17); she only goes on to the next question once all three activities are reported and recorded. But was this intervention desirable? Might the respondent have reported more or other activities if the interviewer had not intervened, or had intervened differently?

As we see it, response accuracy should be the ultimate arbiter of which approach we should take; the whole point of doing a survey is to find out about what is really going on in respondents' lives. By our view, accuracy involves the way respondents map the intended question meaning to their own circumstances. For example, if a respondent is asked whether he has been "out of work" in the last 12 months and answers affirmatively, the accuracy of his response depends on whether his interpretation of his circumstances actually corresponds to the researcher's definition of "out of work." If cooperative respondents understand the question as intended, there is a high probability their responses will be accurate. In this sense, accuracy reflects comprehension. (Of course, respondents can misunderstand a question and answer accurately just by chance or they can understand a question as intended and respond erroneously, including lying about their circumstances.)

Response accuracy can be quite difficult to measure (van der Zouwen, Dijkstra, and Smit, 1991; Wentland, 1993). It is especially difficult to measure by relying only on conversational examples. For our "out-of-work" respondent, unless he happens to describe the details of his employment situation in the course of answering the question, the transcript will not provide sufficient evidence for us to evaluate this correspondence and, therefore, the accuracy of the response. This is one reason we adopt the experimental approach described in the next section. From a practical standpoint, approach 1 may be easier to adopt than approach 2, since it may be closer to standard practice in many organizations (although practice can vary substantially — see Schober and Conrad, 1998). But under either

approach — or under some intermediate approach — taking a collaborative viewpoint may require redefining what we mean by the right and wrong kinds of interviewer influence.

EXPERIMENTAL EVIDENCE

In this section we describe a series of experiments in which we have evaluated response accuracy and interaction patterns for strictly standardized and more collaborative *conversational* interviews, and for various interviewing styles in between. In these studies, professional interviewers telephoned naïve respondents either in the laboratory (Schober and Conrad, 1997; Schober, Conrad, and Fricker, 1999) or at home (Conrad and Schober, 2000) and asked questions from large US government surveys. In the laboratory studies, the “respondents” answered the questions on the basis of scenarios that described the work, housing, and purchases of fictional people. Because we created the scenarios, we knew the correct answer, according to official definitions, for each question–scenario combination and so could determine response accuracy. In the household study, the respondents answered about their own lives; we evaluated the response accuracy by less direct methods.

In each study, respondents participated in either strictly standardized interviews, following Fowler and Mangione’s prescriptions (1990), or in various sorts of less standardized, more collaborative interviews. In all interviews, questions were first posed exactly as worded. In the more collaborative interviews, interviewers were then encouraged to ground understanding of question meaning, for example, by providing scripted definitions when respondents explicitly asked for them (Schober, Conrad, and Fricker, 1999) or by using whatever words interviewers liked to make sure respondents understood the questions as intended (Schober and Conrad, 1997). We believe such studies are necessary because of the variability in interviewer training across (and within) different organizations that consider themselves to practice standardization (Conrad and Schober, 2000) and because of the variability in interviewer behavior within organizations (Schober and Conrad, 1998). Therefore in these studies we have implemented “pure” versions of various techniques in order to see the effects of each technique clearly.

In all studies our basic question is which interviewing technique leads to more accurate responses. Of course, even in purely standardized surveys interviewers can affect responses. What we have tested here is how the kinds of influence that occur in strictly standardized interviews (approach 1) affect response accuracy, as compared to the kinds of influence that occur in more collaborative interviews (approach 2).

Mappings Between Scenarios and Questions

In all studies we used pretested questions from ongoing surveys whose words and grammar have been shown to be understandable, but whose interpretation for some respondents on some occasions might be unclear. Consider a question

like "During the past year, have you purchased or had expenses for household furniture?" A respondent who has bought an end table shouldn't have much trouble answering "yes," but a respondent who has bought a floor lamp may be less sure. Or consider a question like "Last week, how many hours did you work?" This should be clear for a respondent who has a 9-to-5 job that includes overtime, but less clear for a respondent who does business over lunch or solves work-related problems while jogging. We presume that the second respondent would be more likely than the first to request clarification such as "What do you mean by work?" or "Should I consider business lunches to be work?" Interviewers following strictly standardized procedures cannot answer these questions in a meaningful way; they would be obliged to use the "whatever it means to you" neutral probing techniques. More conversationally flexible interviewers could provide information to the respondent that would help clarify what the author of the question had in mind.

In the laboratory studies, we designed the fictional scenarios so that half corresponded to the questions in a straightforward way and, for the other half, the mapping was more complicated. In the household study, we had no control over the frequency of complicated mappings. Our prediction was that response accuracy would be high for both standardized and more conversational interviewing when the mappings were straightforward; when the mappings were complicated, accuracy should suffer for strictly standardized interviewing but not as much—maybe not at all—for conversational interviewing. This pattern of results in the household study, where mappings were not under our control, would indicate that complicated mappings are frequent enough in the real world to warrant further exploration of collaborative techniques.

Concept Training

To assure that conversational interviewers could answer respondents' substantive questions, we needed to teach them the official definitions of key concepts in the questions. Providing standardized interviewers with this knowledge might seem to violate the principles of standardization: The only role for definitions in the standardized interview is for them to be read in their entirety to all respondents on all occasions. But the logic of our experiment required us to train all interviewers together on the concepts so that any accuracy differences could not somehow be attributed to different levels of knowledge between standardized and conversational interviewers. (Standardized interviewers were told that the concepts training was necessary so that interviewers would be able to judge when respondents had answered a question completely; see Beatty, 1995.) In all studies the training lasted about 90 minutes; interviewers first studied the official definitions and then actively carried out exercises to ensure that they had grasped the concepts in detail.

Interviewing Techniques

After the concepts training, interviewers were then trained in their respective interviewing techniques. The standardized instructions were based on guidelines that appeared in an interviewing manual for a survey on which many of these

interviewers regularly worked, and were consistent with Fowler and Mangione's (1990) approach. Using this material, we reviewed standardized question-asking and neutral probing techniques and supplemented this with role-playing exercises.

The interviewers who were trained to use more conversational techniques were instructed to initially read the question as worded. Then (depending on the study) they could substantively answer respondents' requests for clarification, either following a script or in their own words; or they could also provide unsolicited clarification (scripted or in their own words) when respondents seemed to need it, even if respondents hadn't asked for help. In another study (Schober and Conrad, 1998), interviewers were not trained in a particular technique but were told to do whatever they ordinarily do.

Procedure

In the laboratory studies, 130 respondents were given a packet of scenarios to study, and then they were questioned over the telephone about the scenarios. The 87 respondents in various sorts of conversational interviews were instructed to work with the interviewers to make sure they had interpreted the questions as the survey designers intended; they were encouraged to ask for clarification if they needed it. Response accuracy was measured as the percentage of questions for which responses matched what the official definitions required.

Although the interviewers knew that respondents were answering on the basis of fictional situations, they were not familiar with the content of the individual scenarios, and so knowledge was allocated much as it is in real surveys: Interviewers knew the questions, and respondents knew about their own circumstances. We counterbalanced the items so that the respondents who were assigned to a particular interviewer always received different versions of the scenarios. This way the interviewers could not become familiar with the scenarios based on anything the previous respondents might have said.

In the household study 227 respondents were telephoned at home and asked about their own lives; no scenarios were involved. Because we could not directly determine the accuracy of the respondents' answers, we designed the experiment to provide two indirect measures. One measure was response change between interviews. All respondents participated in two interviews: The first was strictly standardized for all respondents; the second was strictly standardized for half of the respondents and conversational for the other half. If respondents' circumstances mapped in a complicated way to the question concepts, they should be more likely to change their answers between an initial standardized interview and a subsequent conversational interview than between two standardized interviews. This is because in the conversational interviews the interviewers should clarify question meaning, correcting respondent misconceptions, which in turn should lead to different answers than in the initial interview. In a second standardized interview, in contrast, the interviewers should not clarify meaning, and initial respondent misconceptions should remain uncorrected and responses unchanged.

The other measure in the household survey was the "legality" of respondents' explanations for their responses. If respondents answered "yes" when asked if

they had made certain types of purchases, they were asked to briefly describe the purchase(s). These were then coded for their consistency with official definitions—their legality. When interviewers could clarify question meaning (in a second conversational interview), respondents should be more likely to base their responses on legal purchases than when interviewers could not clarify question meaning (in any of the standardized interviews).

Results

In each study we first verified that interviewers had followed our instructions and implemented the appropriate technique, based on transcripts of the interviews. One way we showed this was to code the interviewers' various deviations from strict standardization: Rephrasing all or part of the question, providing all or part of a definition (either verbatim or paraphrased), classifying the respondents' descriptions of their circumstances (the fictional scenario in the laboratory studies), offering to provide clarification, confirming or disconfirming the respondent's interpretation of the question, and requesting particular information about the respondent's circumstances. For example, in the following exchange (from Schober and Conrad, 1997), the conversational interviewer paraphrased the long government definition of "household furniture" to answer the respondent's question:

- I: Has Kelly purchased or had expenses for household furniture.
R: Um...is a lamp furniture?
I: No sir, we do not include lamps and lighting fixtures.
R: Okay, no.
(I goes on to next question)

In strictly standardized interviewing, the interviewer should not have answered the respondent's request for clarification because by doing so she interpreted the survey question for the respondent. Across the various studies our coding leads us to be confident that our interviewer training led to fundamentally different types of interaction. In the Schober and Conrad (1997) study, for example, 85 percent of the question-answer sequences in conversational interviews contained deviations from standardization, compared with only 2 percent in strictly standardized interviews.

We can now turn to response accuracy. Again, an accurate response in our experiments is one that is consistent with the official definition of the relevant concept. In our earlier lamp example, the correct answer is "no" because a lamp purchase does not qualify as a furniture purchase, and so the respondent's answer was accurate.

Across all our lab studies, a general pattern emerges. When mappings between question concepts and people's circumstances (scenarios) are straightforward, all interviewing techniques lead to nearly perfect accuracy; virtually all respondents interpret question concepts in the ways that survey designers intended. But when mappings between question concepts and people's circumstances are complicated, strictly standardized interviewing leads to quite poor response accuracy (28 percent in Schober and Conrad's 1997 study). Response accuracy improves when interviewers provide clarification on request (54 and 59 percent

in Schober, Conrad, and Fricker, 2000), and it improves substantially more when interviewers are also empowered to use the full resources of ordinary conversational interaction (87 percent in Schober and Conrad, 1997). It does not matter much whether interviewers' clarification is scripted or unscripted (Schober, Conrad, and Fricker, 2000). Even when interviewers do what they ordinarily do (Schober and Conrad, 1998) response accuracy is much higher when they provide clarification (79 percent) than when they leave the interpretation of questions entirely up to respondents (23 percent).

These results are mirrored in the household study (Conrad and Schober, 2000) where we used indirect measures of response accuracy. Respondents changed their answers more when their second interview was conversational (22 percent) than when it was strictly standardized (11 percent). In addition, more responses were based on legal purchases when the second interview was conversational (95 percent) than when it was standardized (57 percent). The improved response accuracy in conversational interviews suggests that respondents' actual circumstances (as opposed to the fictional scenarios presented in the lab studies) are complicated often enough—at least for these questions—to justify exploring the technique further.

Fowler and Mangione (1990) have raised the concern that interviewers empowered to collaborate will mislead respondents by providing inaccurate information. In our experiments this hasn't seemed to be the case. For example, in the Schober and Conrad (1997) study, the information about the question that conversational interviewers provided was accurate (conformed to the official definitions) in 93 percent of the cases where it was given. For 87 percent of all cases respondents received accurate information from interviewers and provided accurate answers; for only 6 percent of all cases did respondents receive accurate information from interviewers but provide inaccurate answers. For the 7 percent of cases where interviewers provided inaccurate information, respondents were still accurate about half the time. In 4 percent of all cases, respondents received inaccurate information from interviewers yet provided accurate answers; in only 3 percent of all cases did respondents receive inaccurate information and answer the questions inaccurately. So overall, conversational interviewers provided highly accurate information. When they provided inaccurate information, this did not necessarily lead respondents to produce incorrect answers; in fact, respondents produced incorrect answers resulting from inaccurate information only 3 percent of the time.

Closer analysis of the interviewer–respondent interaction (see Conrad and Schober, 2000; Schober and Conrad, 1997) shows that it really was interviewers' deviations from standardization that led to the increases in response accuracy. And it seems that interviewers' interventions improved response accuracy whether respondents had requested clarification or not (i.e., even when interviewers provided the information without the respondents having asked for it). So, for example, in the Schober and Conrad (1997) study, for the 64 complicated-mapping cases where interviewers provided unsolicited help, respondents produced accurate answers for 55, a rate of 86 percent accuracy. In contrast, for the 11 complicated-mapping cases where interviewers did not provide any help, respondents only

produced four accurate answers, a rate of 34 percent. This figure is close to the 28 percent accuracy for complicated mappings in standardized interviews, suggesting that when conversational interviewers behave like standardized interviewers and don't provide clarification, response accuracy will suffer.

Across our studies, this improvement in response accuracy came at a significant cost. Conversational interviews took much longer than standardized interviews (from 80 to 300 percent longer, in the various studies), and this was true for both straightforward and complicated mappings. It remains to be seen whether this is a necessary by-product of conversational interviewing or whether conversationally flexible interviewers could be trained to achieve the same improvements in response accuracy in shorter interviews.

The results of our studies should not be taken as the final word on the issue, nor should they be taken as showing definitively that conversational interviewing is always a good idea. Our studies have examined nonsensitive fact-based questions, and the results may not generalize to questions about sensitive topics or opinions in a straightforward way. Just like fact-based questions, opinion questions contain phrases with alternative possible interpretations—consider “abortion” or “approve”—and thus opinion surveys might benefit from more collaborative approaches to interviewing. But whether this can be done without influencing the opinions is unclear, especially because response accuracy for opinions can't be validated as directly as it is for the fact-based questions in our studies.

Our results also don't take into account the potential real-world costs of implementing more collaborative interviewing techniques. Beyond the potential expenses associated with increased interview length, interviewer training might have to be more intensive than it often is now. Interviewer behavior would have to be monitored even more closely to ensure that question meanings were being clarified appropriately and uniformly, without increasing interviewer variance. Far more effort would have to go into developing clear definitions for question concepts. To the extent that respondents found increased collaboration a burden (which could depend on several factors—see Schober, Conrad, and Bloom, 2000), response rates could be affected.

Ultimately the generalizability of our experimental findings depends on the frequency of complicated mappings between questions and respondents' circumstances in real surveys, and this may vary from survey to survey. If complicated mappings are known to be rare, then strictly standardized techniques should be optimal, leading to accurate responses at reasonable costs. If the complicated mappings are known to be frequent, or if (more realistically) their frequency is unknown, more collaborative techniques might be worth the increased costs their use would, no doubt, entail.

CONCLUSIONS

We have seen that the theory of standardization assumes a view of communication, the message model, that does not hold in spontaneous conversations. We have argued on conceptual grounds that it doesn't hold in

standardized surveys either. Interviewers can't help but influence responses even if they use only the neutral probing techniques that are supposed to prevent such influence. Because different probes have different collaborative effects, even allowing interviewers to choose which probes to provide could lead one interviewer to bias responses differently than another.

Since interviewers always influence responses, this raises the question of which kinds of influence are benign and which are not. We argue that the criterion should be how interviewer behaviors affect response accuracy—that is, how well responses correspond with the definitions the survey authors had in mind. Using this criterion, we have demonstrated experimentally that standardized interviewing techniques lead to poor response accuracy when respondents are unclear about how questions map onto life circumstances. Interviewers empowered to use the full range of techniques for grounding understanding in natural conversation—providing substantive clarification, both solicited and unsolicited—help improve response accuracy.

We are aware that in some surveys such clarifications are considered part of standardized practice. But under the purest definition of standardization, interviewer clarifications are not standardized. Survey researchers who allow clarification should be aware that they are advocating a certain degree of nonstandardization. While we believe our experimental data and our conceptual arguments support some version of conversational interviewing, it is unclear whether the mix of standardized and conversational interviewing in current practice is optimal.

Adopting a collaborative model of interviewer–respondent interaction forces us to redraw various theoretical boundaries. We need to reconsider the assumption that respondent error, interviewer error, and error due to question wording are independent. We need to dispense with the notion that interviewers can behave in ways that don't influence responses. We need to examine further whether the deviations from strict standardization that occur in current interviewing practice are harmful or helpful. If we abandon the underlying message model assumption that meaning resides in words, we are forced to take seriously the proposal that interviews are only standardized when respondents interpret questions the same way (see Suchman and Jordan, 1990, 1991). Ultimately we may need to redefine what standardization ought to be.

NOTES

1. Some organizations that consider themselves to promote standardized interviewing would not consider this a deviation from standardization. But technically, because not every respondent is provided with this feedback, this intervention isn't standardized, in the strictest sense (Fowler and Mangione, 1990).

REFERENCES

- Akmajian, A., R. A. Demers, A. K. Farmer, and R. M. Harnish. 1990. *Linguistics: An Introduction to Language and Communication*, 3rd ed. Cambridge, MA: MIT Press.

- Beatty, P. 1995. "Understanding the Standardized/Non-Standardized Interviewing Controversy." *Journal of Official Statistics* 11: 147-160.
- Belson, W. A. 1981. *The Design and Understanding of Survey Questions*. Aldershot: Gower.
- Belson, W. A. 1986. *Validity in Survey Research*. Aldershot: Gower.
- Brennan, S. E. 1990. "Seeking and Providing Evidence for Mutual Understanding." Unpublished Ph.D. dissertation, Department of Psychology, Stanford University.
- Button, G. 1987. "Answers as Interactional Products: Two Sequential Practices Used in Interviews." *Social Psychology Quarterly* 50: 160-171.
- Cahn, J. E., and S. E. Brennan. 1999. "A Psychological Model of Grounding and Repair in Dialog." Pp. 25-33 in S. E. Brennan, A. Giboin, and D. Traum (Eds.), *Proceedings of the American Association for Artificial Intelligence Fall Symposium, Psychological Models of Communication in Collaborative Systems*. Menlo Park, CA: AAAI Press.
- Cannell, C. F., P. V. Miller, and L. Oksenberg. 1981. "Research on Interviewing Techniques." Pp. 389-437 in S. Leinhardt (Ed.), *Sociological Methodology*, Vol. 11, San Francisco: Jossey-Bass.
- Cicourel, A. V. 1973. *Cognitive Sociology: Language and Meaning in Social Interaction*. New York: Free Press.
- Clark, H. H. 1991. "Words, the World, and Their Possibilities." Pp. 263-277 in G. Lockhead and J. Pomerantz (Eds.), *The Perception of Structure*. Washington, DC: American Psychological Association.
- Clark, H. H. 1992. *Arenas of Language Use*. Chicago: University of Chicago Press.
- Clark, H. H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Clark, H. H., and S. E. Brennan. 1991. "Grounding in Communication." Pp. 127-149 in L. B. Resnick, J. M. Levine, and S. D. Teasley (Eds.), *Perspectives on Socially Shared Cognition*. Washington, DC: American Psychological Association.
- Clark, H. H., and R. J. Gerrig. 1983. "Understanding Old Words with New Meanings." *Journal of Verbal Learning and Verbal Behavior* 22: 591-608.
- Clark, H. H., and E. F. Schaefer. 1987. "Collaborating on Contributions to Conversations." *Language and Cognitive Processes* 2: 19-41.
- Clark, H. H., and E. F. Schaefer. 1989. "Contributing to Discourse." *Cognitive Science* 13: 259-294.
- Clark, H. H., and M. F. Schober. 1991. "Asking Questions and Influencing Answers." Pp. 15-48 in J. M. Tanur (Ed.), *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage.
- Clark, H. H., and D. Wilkes-Gibbs. 1986. "Referring as a Collaborative Process." *Cognition* 22: 1-39.
- Conrad, F. G., and M. F. Schober. 2000. "Clarifying Question Meaning in a Household Telephone Survey." *Public Opinion Quarterly* 64: 1-28.
- Dijkstra, W. 1987. "Interviewing Style and Respondent Behavior: An Experimental Study of the Survey-Interview." *Sociological Methods and Research* 16: 309-334.
- Drew, P. 1984. "Speakers' 'Reportings' in Invitation Sequences." Pp. 129-151 in J. M. Atkinson and J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.
- Fowler, F. J., and T. W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: Sage.

- Gibbs, R. W. 1994. *The Poetics of Mind*. Cambridge: Cambridge University Press.
- Goffman, E. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Goodwin, C. 1981. *Conversational Organization: Interaction between Speakers and Hearers*. New York: Academic.
- Grice, H. P. 1975. "Logic and Conversation." Pp. 225–242 in P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics*, Vol. 3: *Speech Acts*. New York: Academic.
- Groves, R. M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R. M. 1991. "Measurement Error across the Disciplines." Pp. 1–25 in P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. New York: Wiley.
- Gumperz, J. J. 1982. *Discourse Strategies*. Cambridge: Cambridge University Press.
- Heritage, J. 1984. *Garfinkel and Ethnomethodology*. New York: Basil Blackwell.
- Heritage, J., and M. L. Sorjonen. 1994. "Constituting and Maintaining Activities across Sequences: And—Prefacing as a Feature of Question Design." *Language in Society* 23: 1–29.
- Houtkoop-Steenstra, H. 2000. *Interaction and the Standardized Survey Interview: The Living Questionnaire*. Cambridge: Cambridge University Press.
- Krauss, R. M., and S. R. Fussell. 1996. "Social Psychological Models of Interpersonal Communication." Pp. 655–701 in E. T. Higgins and A. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles*. New York: Guilford.
- Marlaire, C. L., and D. W. Maynard. 1990. "Standardized Testing as an Interactional Phenomenon." *Sociology of Education* 63: 83–101.
- Maynard, D. W., and C. L. Marlaire. 1992. "Good Reasons for Bad Testing Performance: The Interactional Substrate of Educational Testing." *Qualitative Sociology* 15: 177–202.
- Maynard, D. W., and M. R. Whalen. 1995. "Language, Action, and Social Interaction." Pp. 149–175 in K. Cook, G. Fine, and J. House (Eds.), *Sociological Perspectives in Social Psychology*. Boston: Allyn & Bacon.
- Mishler, E. G. 1986. *Research Interviewing*. Cambridge, MA: Harvard University Press.
- Reddy, M. J. 1979/1993. "The Conduit Metaphor: A Case of Frame Conflict in Our Language about Language." Pp. 164–201 in A. Ortony (Ed.), *Metaphor and Thought*, 2nd ed. Cambridge: Cambridge University Press.
- Rogoff, B. 1990. *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Sacks, H., E. A. Schegloff, and G. Jefferson. 1974. "A Simplest Systematics for the Organization of Turn-Taking in Conversation." *Language* 50: 696–735.
- Schaeffer, N. C. 1991. "Conversation with a Purpose—or Conversation? Interaction in the Standardized Interview." Pp. 367–391 in P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. New York: Wiley.
- Schaeffer, N. C., and H.-W. Charng. 1991. "Two Experiments in Simplifying Response Categories: Intensity Ratings and Behavioral Frequencies." *Sociological Perspectives* 34: 165–182.
- Schaeffer, N. C., and D. W. Maynard. 1996. "From Paradigm to Prototype and Back Again: Interactive Aspects of Cognitive Processing in Standardized Survey Interviews." Pp. 65–88 in N. Schwarz and S. Sudman (Eds.), *Answering Questions:*

- Methodology for Determining Cognitive and Communicative Processes in Survey Research*. San Francisco, CA: Jossey-Bass.
- Schegloff, E. A. 1972. "Notes on a Conversational Practice: Formulating Place." Pp. 75–119 in D. Sudnow (Ed.), *Studies in Social Interaction*. New York: Free Press.
- Schiffrin, D. 1994. *Approaches to Discourse*. Cambridge, MA: Basil Blackwell.
- Schober, M. F. 1998. "Conversational Evidence for Rethinking Meaning." *Social Research* 65: 511–534.
- Schober, M. F. 1999. "Making Sense of Questions: An Interactional Approach." Pp. 77–93 in M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, and R. Tourangeau (Eds.), *Cognition and Survey Research*. New York: Wiley.
- Schober, M. F., and H. H. Clark. 1989. "Understanding by Addressees and Overhearers." *Cognitive Psychology* 21: 211–232.
- Schober, M. F., and F. G. Conrad. 1997. "Does Conversational Interviewing Reduce Survey Measurement Error?" *Public Opinion Quarterly* 61: 576–602.
- Schober, M. F., and F. G. Conrad. 1998. "Response Accuracy When Interviewers Stray from Standardization." Pp. 940–945 in *Proceedings of the American Statistical Association, Section on Survey Methods Research*. Alexandria, VA: American Statistical Association.
- Schober, M. F., F. G. Conrad, and J. E. Bloom. 2000. "Clarifying Word Meanings in Computer-Administered Survey Interviews." Pp. 447–452 in L. R. Gleitman and A. K. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.
- Schober, M. F., F. G. Conrad, and S. S. Fricker. 1999. "When and How Should Survey Interviewers Clarify Question Meaning?" In *Proceedings of the American Statistical Association, Section on Survey Methods Research*. Alexandria, VA: American Statistical Association.
- Smit, J. H., W. Dijkstra, and J. van der Zouwen. 1997. "Suggestive Interviewer Behaviour in Surveys: An Experimental Study." *Journal of Official Statistics* 13: 19–28.
- Strack, F., and N. Schwarz. 1992. "Communicative Influences in Standardized Question Situations: The Case of Implicit Collaboration." Pp. 173–193 in G. R. Semin and K. Fiedler (Eds.), *Language, Interaction and Social Cognition*. London: Sage.
- Suchman, L., and B. Jordan. 1990. "Interactional Troubles in Face-to-Face Survey Interviews." *Journal of the American Statistical Association* 85: 232–253.
- Suchman, L., and B. Jordan. 1991. "Validity and the Collaborative Construction of Meaning in Face-to-Face Surveys." Pp. 241–267 in J. M. Tanur (Ed.), *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage.
- Traum, D. R. 1994. "A Computational Theory of Grounding in Natural Language Conversation." Unpublished Ph.D. dissertation, Department of Computer Science, University of Rochester.
- van der Zouwen, J., W. Dijkstra, and J. H. Smit. 1991. "Studying Respondent-Interviewer Interaction: The Relationship Between Interviewing Style, Interviewer Behavior, and Response Behavior." Pp. 419–437 in P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman (Eds.), *Measurement Errors in Surveys*. New York: Wiley.
- Wentland, E. J. 1993. *Survey Responses: An Evaluation of Their Validity*. San Diego, CA: Academic.