CLARIFYING QUESTION MEANING IN A HOUSEHOLD TELEPHONE SURVEY

FREDERICK G. CONRAD
MICHAEL F. SCHOBER

Abstract This study contrasts two interviewing techniques that reflect different tacit assumptions about communication. In one, strictly standardized interviewing, interviewers leave the interpretation of questions up to respondents. In the other, conversational interviewing, interviewers say whatever it takes to make sure that questions are interpreted uniformly and as intended. Respondents from a national sample were interviewed twice. Each time they were asked the same factual questions from ongoing government surveys, five about housing and five about recent purchases. The first interview was strictly standardized; the second was standardized for half the respondents and conversational for the others. Respondents in a second conversational interview answered differently than in the first interview more often, and for reasons that conformed more closely to official definitions, than respondents in a second standardized interview. This suggests that conversational interviewing improved comprehension, although it also lengthened interviews. We conclude that respondents in a national sample may misinterpret certain questions frequently enough to compromise data quality and that such misunderstandings cannot easily be eliminated by pretesting and rewording questions alone. More standardized comprehension may require less standardized interviewer behavior.

Introduction

Imagine the following exchange between an interviewer (I) and a respondent (R) in a Current Population Survey (CPS) interview:

> I: Last week, did you have more than one job, including part-time, evening, or weekend work?
> R: Well, it depends . . . I baby-sit for different people—is that one job or more than one?

What should the interviewer do now? Different survey organizations that advocate standardized interviewing practices would recommend different approaches for the interviewer to take.

At one end of the spectrum, if the interviewer is administering a *strictly standardized interview* (see Fowler and Mangione 1990), she will not answer the respondent's query. Instead, by using "neutral probes" she will persuade the respondent to interpret the question by himself. Under the theory of standardization, the interviewer must maintain a neutral stance in order to minimize the possibility that she will bias responses. If all interviewers follow this procedure, the variability in the data due to interviewers should be low; low interviewer variance is a necessary component of high-quality data (see Beatty [1995] for a review of the history of standardized techniques).

An approach at the other end of the spectrum, advocated by Suchman and Jordan (1990, 1991) among others, is that the interviewer should answer the respondent's query using whatever words it takes to help the respondent understand the question from the survey designers' perspective, without unduly influencing the response. This can mean deviating from the script and probing in ways that would be forbidden in a strictly standardized interview. For example, our interviewer might explain that for this question the sponsoring organization counts baby-sitting for more than one employer as only one job (see U.S. Department of Commerce, 1994, pp. B1-4 to B1-6). We call this alternative practice *conversational interviewing*.

The rationale for conversational interviewing comes from theories of communication in various social sciences. The idea is that in ordinary unscripted communication, people can only be confident that they understand each other if they can *collaborate*, that is, discuss the current topic until they believe they have both taken it to mean the same thing (see, e.g., Clark 1992, 1996; Clark and Brennan 1991; Clark and Wilkes-Gibbs 1986; Schober and Clark 1989). Survey interviews share certain features with spontaneous conversations (Clark and Schober 1991; Schaeffer 1991; Schober 1999; Schober and Conrad, in press; Suchman and Jordan 1990, 1991), among them the fact that the same words can mean different things to different people; survey response accuracy may improve if participants collaborate to ensure that they understand questions in the same way.

In this article, we consider the costs and benefits of these alternative ap-

proaches. We recognize that few survey organizations consistently practice strict standardization or fully conversational interviewing. Some organizations that consider their practice to be standardized allow interviewers to provide scripted clarification whenever respondents ask for it (note that this practice is not strictly standardized, in that some respondents get clarification and others do not [Fowler and Mangione 1990, p. 21]). In fact, within an organization the policy can be inconsistent. For example, in one set of training materials CPS interviewers are taught that if respondents misunderstand a question, they must repeat the question verbatim; if the respondent still misunderstands the question, they must use only neutral probes to avoid influencing responses (*CPS Interviewer's Manual,* U.S. Department of Commerce 1994, pp. A2-6 to A2-8). Another set of materials, however, trains these same interviewers to "clear up misunderstandings before moving to the next question" (*Training Guide for New Interviewers*, U.S. Bureau of the Census 1997, p. D-10), which in practice includes providing definitions.

Also, interviewers within a single organization may vary in the techniques they practice. In a study of a small sample of interviewers at one organization (Schober and Conrad 1998), we found that interviewers ranged from behaving in a strictly standardized fashion to providing unscripted clarification when asked; on average, they deviated from strictly standardized technique 20 percent of the time. In another organization, interviewers deviated from strict standardization as little as 4 percent and as much as 36 percent of the time, depending on the question (Dykema, Lepkowski, and Blixt 1997).

This variation in what is considered standardized makes it important to evaluate a corresponding range of interviewing techniques. In this study we evaluate "pure" versions of standardized and conversational interviewing, both for practical experimental reasons and to highlight the contrasting philosophies these techniques embody. How the results apply to a particular real-world survey setting depends on the actual interviewing practices in each setting.

Both standardized and conversational interviewing are advocated to bring uniformity and comparability to survey data. However, they embody different assumptions about the nature of meaning and communication, even if these assumptions are not always articulated (Schober 1998; Schober and Conrad, in press). Their success is evaluated by different measures, and they rely on question pretesting in different ways. For rhetorical purposes we present these differences as if they were polar opposites, but we do not want to give the impression that all researchers explicitly subscribe to one position or the other; researchers' views can fall somewhere in between.

ASSUMPTIONS ABOUT MEANING

Strictly standardized interviewing is designed to present exactly the same stimulus to all respondents—that is, to standardize the *words* in a question—regardless of which interviewer reads the question. It thus embodies

the assumption that meaning is carried by words—what Akmajian et al. (1990) have called a "message model" of communication. Thus, as long as all respondents are read the same words, a survey will be standardized. If some respondents are read definitions of survey concepts and others are not, the survey is no longer strictly standardized. If definitions are to be presented to respondents, they must be presented to everyone; because definitions can be so long and complicated,[1] this is rarely done.

Conversational interviewing, in contrast, is designed to make sure that all respondents understand the question the same way—to standardize the *meaning* of that question—irrespective of who reads it to the respondent. It thus embodies the assumption that simply speaking words does not guarantee that the listener will grasp their intended meaning; speakers and addressees may engage in further dialogue in order to understand each other as well as they need to. Unlike some more radical proposals (e.g., Mishler 1986), conversational interviewing is based on the assumption that it is possible for meaning to be consistent across situations and respondents, as long as interviewers can converse with respondents to clarify those meanings. Conversational interviewing to clarify concepts should not be confused with earlier nonstandardized techniques (described in Beatty [1995]) designed to increase honest responding through better interviewer-respondent rapport. Similarly, the purpose of conversational interviewing should not be seen as motivating respondents to answer accurately (see Cannell, Miller, and Oksenberg [1981] for techniques to increase respondent commitment and motivation); the purpose is to clarify concepts in questions.

MEASURING THE TECHNIQUE'S SUCCESS

Strictly standardized interviewing is designed, in part, to make sure that interviewers do not bias responses. This can be measured by examining interviewer variance: if interviewer variance is low, then interviewers have probably not biased the responses. In contrast, the success of conversational interviewing can be measured by examining the extent to which respondents' answers fit the survey designers' concepts; if respondents can be shown to have interpreted questions as the survey designers intended, then response accuracy is high.

---

1. Consider the U.S. Department of Commerce's (1994, pp. B1-4 to B1-6) definition of a "job": "A job exists when there is a definite arrangement for regular work every week, or every month, for pay or other compensation (e.g., profits, anticipated profits, or pay in kind, such as room and board). A formal, definite arrangement with one or more employers to work on a continuing basis for a specified number of hours per week or days per month, but on an irregular schedule during the week or month, is also a job. . . . It is possible for individuals to have more than one employer, but only one job. If an individual does the same type of work for more than one employer in an occupation where it is common to have more than one employer, do not consider the individual a multiple jobholder. Examples include private household or domestic workers including babysitters, chauffeurs, gardeners, handypersons, cooks, and maids."

THE ROLE OF PRETESTING AND REFINING QUESTIONS

Both strictly standardized and conversational techniques assume that questions have been pretested, but they differ in the degree to which they rely on pretesting to promote uniform interpretation of questions. Under standardization, questions need to be pretested in order to unearth the misconceptions to which most respondents would likely fall prey; questions can then be rewritten so that most respondents will understand them as intended. Under conversational interviewing, pretesting is essential to rule out systematic misconceptions, but respondents may still interpret a pretested question idiosyncratically. This is because there is more to understanding a question than just "looking up" the meanings of the individual words and phrases in the mental dictionary. Respondents must also determine how the words in the question (e.g., "more than one job") correspond to their personal circumstances (e.g., baby-sitting for multiple employers). The idea is that no single question wording can, by itself, make this correspondence clear for all respondents. But if respondents and interviewers can converse about the question, they may be able to clarify the correspondence jointly.

Experiment

In the current study, we examine how conversational and strictly standardized interviewing techniques affect measurement error in a household telephone survey. This study builds on an earlier laboratory experiment (Schober and Conrad 1997) in which we compared response accuracy in conversational and strictly standardized telephone interviews. In that study respondents answered on the basis of fictional scenarios, rather than about their own lives, so that we could measure response accuracy directly. The finding was that conversational and strictly standardized techniques may be appropriate under different circumstances. When respondents' circumstances mapped onto questions in a straightforward way (e.g., a respondent who was asked, "Last week, did Pat have more than one job, including part-time, evening, or weekend work?" answered about a scenario in which Pat baby-sat for one family), responses were extremely accurate in both standardized and conversational interviews. When the circumstances mapped onto questions in a complicated way (e.g., Pat baby-sat for several families in one week), responses were quite inaccurate in strictly standardized interviews, but almost 60 percent more accurate in conversational interviews. These improvements in accuracy came at a substantial cost: conversational interviews took more than three times as long as standardized interviews.

Here we examine whether these results extend beyond the laboratory. It is entirely possible that they do not. First, mappings between question concepts and actual respondents' circumstances may be straightforward most of the time. If so, most respondents may understand questions as intended and re-

spond accurately, and strict standardization may be the more efficient technique. But if complicated mappings are relatively frequent, then respondents may not interpret questions comparably, and conversational interviewing may lead to more accurate responding. Such a result would suggest that conversational interviewing is worth exploring further. A second reason to test the generality of the Schober and Conrad (1997) laboratory findings is that it is unclear how conversational interviewing might affect response rates; actual respondents may be unwilling to engage in or complete lengthier conversational interviews even if these interviews help them understand questions better.

In our study, 227 respondents were interviewed at home by 20 experienced telephone interviewers calling from Westat, a social science research firm in Rockville, MD. Each respondent was asked the same survey questions twice, each time by a different interviewer, with about 1 week between interviews. The first interview was always standardized. For half the respondents (118), selected at random, the second interview was also standardized. For the other half (109), the second interview was conversational.

Ideally, one could directly measure response accuracy by comparing answers to verifiable independent sources, like direct observations or official records. In the current study we did not have access to either of these, and so we assessed accuracy of understanding with two different measures. The first measure was response change across interviews: if conversational interviewing improves comprehension beyond the laboratory, then responses that reflect misconceptions in a standardized interview should be corrected in a subsequent conversational interview. Thus respondents who participated in an initial standardized interview should change their responses more often in a subsequent conversational interview than they would in a subsequent standardized interview.[2]

The second measure involved respondents' explanations of their answers: if conversational interviewing improves comprehension, then respondents should be more likely to justify their answers using the same criteria as the survey designers. That is, if respondents who claim to have purchased an item (e.g., household furniture) in the last year are asked to list those purchases (e.g., table, sofa, lamp), more purchases should be "legal"—consistent with official definitions—in a conversational than in a standardized interview.

Considered independently, these measures allow alternative interpretations, but when considered together the alternatives are less likely. For the response change measure, one might expect, simply on the basis of test-retest reliability, that two different administrations (standardized-conversational) of a questionnaire should lead to greater response change than two identical administrations (standardized-standardized), even if comprehension has not im-

---

2. Note that this reverses the usual logic of reinterview studies, in which response change between interviews is typically considered to signal respondent problems.

proved. But if response change is accompanied by a higher percentage of legal explanations, then it is more likely that the response change results from improved comprehension. For the measure based on explanations, it is not possible to draw any conclusions about response accuracy. Respondents might have fewer misconceptions about a question in a conversational interview but nonetheless respond inaccurately because they still have *some* misconceptions about that question; thus improving comprehension might not necessarily improve response accuracy. But if respondents also change their answers when they provide more legal explanations, then it is more likely that improved comprehension actually leads to response change.

Note that unlike in the Schober and Conrad (1997) laboratory experiment, where half of the question-scenario pairings led to complicated mappings, here we did not control the frequency of complicated mappings and, therefore, the frequency of situations in which conversational interviewing might improve comprehension. If conversational interviewing really does produce more response change and more legal explanations, then complicated mappings must be frequent enough to worry about in actual survey settings, at least for these questions.

QUESTIONS

Respondents were asked the same 10 questions in both the first and second interviews. Five were about housing, adapted from the Consumer Price Index Housing (CPI-Housing) survey; these required numerical responses. For example, one question was, "How many bedrooms are there in your home?" The other five were about purchases, adapted from the Current Point of Purchase Survey (CPOPS); these required yes/no responses. For example, one question was, "During the past year, that is since July of 1995, have you purchased or had expenses for household furniture?" (see app. A for the complete list of questions). Respondents who answered yes to any purchase questions were also asked to list the purchases; conversational interviewers and respondents could discuss the rationale for listing or not listing a particular purchase. All questions had been pretested by the sponsoring agencies, who had published definitions for the key concepts (see app. B).

In each interview, question order varied: about half the respondents answered the five housing questions first and then the five purchasing questions, and the other half answered the five purchasing questions first and then the five housing questions. About half the respondents answered questions in the second interview in a different order than they had in the first interview. After answering the survey questions in the second interview, respondents answered several demographic questions.

Respondents were chosen from a nationally representative sample of residential households with telephones in the continental United States, generated through a random digit dialing process. Respondents were telephoned in their homes and asked to participate in two interviews to improve the data collection procedures at the Department of Labor. They were asked for permission to audio-record both interviews. They were not told anything beforehand about the interviews except that there would be two. When they were recontacted, they were reminded that they had recently completed the first of two interviews in which they had agreed to participate and were told that the purpose of the current call was to conduct the second interview. The two groups of respondents (standardized and conversational second interviews) were comparable demographically. The 118 respondents whose second interview was standardized were 44 men and 74 women; 88 were white, 17 black, 1 Asian/Pacific Islander, 1 Native American, and 10 other (1 refused to report); 9 had no high school diploma, 38 had completed high school only, 36 had some college but no degree, 19 had a college degree, and 15 had postgraduate degrees (1 other). The 109 respondents whose second interview was conversational were 36 men and 73 women; 88 were white, 16 black, 1 Asian/Pacific Islander, 1 Native American, and 2 other (1 refused to report); 8 had no high school diploma, 40 had completed high school only; 35 had some college but no degree; 14 had a college degree, and 11 had postgraduate degrees.

The 20 interviewers (10 in the first week, 10 in the second) were all professionals with no reliable differences in interviewing experience at Westat: standardized interviewers averaged 29.6 months and conversational interviewers averaged 34.4 months, $F(1,19) = 0.08$, n.s. Fifteen interviewers used strictly standardized technique, all 10 of the first-week interviewers and 5 of the second-week interviewers. The group of 15 standardized interviewers and the group of 5 conversational interviewers were comparable demographically. Of the 15 standardized interviewers, 13 were women and 2 were men; 5 were black, 7 were white, and 3 were Hispanic; 3 had additional professional interviewing experience elsewhere. The 5 conversational interviewers were 4 women and 1 man; 2 were black and 3 were white; 3 had additional professional interviewing experience elsewhere.

All of the interviewers conducted about the same number of interviews. For the 10 interviewers who conducted the initial standardized interviews, the median number of interviews was 26 (ranging from 8 to 27); for the five standardized interviewers who interviewed respondents the second time, the median number of interviews was 25 (ranging from 16 to 28); and for the five conversational interviewers who interviewed respondents the second time the median number of interviews was 24 (ranging from 16 to 27).

For the initial standardized interviews, 62.3 percent of the respondents contacted agreed to participate. For the second interview, conversational in-

terviewing did not reduce response rates. Response rates (proportion of those participating in the first interview) for the conversational (second) interviews (82.9 percent) were not reliably different from response rates for standardized (second) interviews (89.7 percent), $z(7)$, two-tailed $= 1.61$, n.s. But we should note that the respondents in this study had already agreed to participate in two tape-recorded interviews and had completed the first of these; we do not know whether less cooperative respondents might decline to participate in or complete conversational interviews more often than they would decline to participate in or complete standardized interviews.

INTERVIEWER TRAINING

All interviewers were trained on the key survey concepts (see app. B) for about 1 hour; this included a presentation about the concepts, a quiz, and group discussion about the concepts. Interviewers were trained in groups for an additional hour in either standardized or conversational interviewing technique. Interviewers were told that their respective techniques were designed to produce high-quality data. The 15 standardized interviewers were trained to conduct a strict version of standardized interviewing where they read questions exactly as worded and provided only nondirective probes but never provided definitions for the survey concepts. (Following Fowler and Mangione's [1990] logic, we explained that the earlier concepts training had been conducted so that they could judge when respondents had answered a question completely.) Probing techniques included rereading the question, providing the response alternatives again (as in "Is that a yes or a no?"), and otherwise probing neutrally (e.g., saying "We need your interpretation" or "Whatever it means to you").

The other five interviewers were trained to conduct conversational interviews. Like standardized interviewers, these interviewers were instructed to read the questions exactly as worded, but then they could say whatever they wanted to assure that the respondent had understood the question as the survey designer had intended. This included reading or paraphrasing all or part of a question, reading or paraphrasing all or part of a definition, and asking questions of the respondent to elicit information so that the interviewer and respondent could jointly reach a correct response. Interviewers could intervene at the respondent's request or voluntarily; that is, interviewers were licensed to intervene whenever they thought the respondent might have misunderstood the question. Interviewers seemed to find the technique easy to learn.

All interviewers were trained to explain to respondents that the study involved two interviews in order to help the Department of Labor improve its data collection procedures. Conversational interviewers were also trained to explain to respondents that this interview would be different from the first one and probably different from most interviews in which they had previously participated. These interviewers were to explain to respondents that official

definitions might differ from respondents' own ideas about what counts as, say, a room or a furniture purchase. And these interviewers were to encourage respondents to ask for clarification whenever respondents had the slightest doubt about how to interpret or answer questions. None of the interviewers told respondents anything about the likely duration of any interviews.

Note, therefore, that in this study two variables (interviewing technique and preinterview instructions to respondents) were manipulated simultaneously; both manipulations were necessary in order to compare pure versions of standardized and conversational interviewing techniques. If the manipulations do indeed affect response change and purchase reports, further studies could determine the relative contribution of each variable.

Results

IMPLEMENTATION OF INTERVIEWING TECHNIQUES

In order to interpret our results, we need to be sure that interviewers correctly implemented both interviewing techniques. Because doing this involved detailed and labor-intensive inspection of interviewer-respondent interaction, we sampled a subset of 35 pairs of interviews from the full set of 227 pairs of interviews, and we transcribed and coded the audiotaped interviews. We first sampled 40 pairs at random, with the constraint that at least two interviews by each second-week interviewer would be included. This would also ensure that about half the sampled pairs would have standardized second interviews, and about half conversational. We sampled four additional pairs of interviews so that there would be at least two interviews by each first-week interviewer as well. After discarding those pairs of interviews where one of the sessions had not been properly audiotaped, we were left with a subsample of 35 pairs of interviews (18 standardized both times and 17 with conversational second interviews), in which each interviewer had conducted at least two interviews.

In this subsample, we counted all *nonstandard interventions*—all those utterances by interviewers that would be "illegal" in pure standardized interviewing. These included providing definitions or clarifications of concepts in the survey questions (either verbatim from the official documents or improvised), requesting information from the respondent relevant to the definition, offering clarification, proposing potential purchases (mentioned in the definition) as examples, explaining how respondents' descriptions of their circumstances fit the response options, and rewording the survey question (after reading it verbatim initially). At least one nonstandard intervention occurred in 87 percent of the questions in conversational interviews, but in only 6 percent of the questions in standardized interviews. (The percentage was the same for second-week standardized interviews, 6 percent, as in all first-week standardized interviews.) Most (87 percent, 26 of 30) of these nonstandard

interventions in standardized interviews were partial but verbatim repetitions of the survey questions; these are illegal by only the strictest criteria of standardization. So clearly the two types of interviews were implemented in qualitatively different ways and much as we had intended.

MEASURE 1: RESPONSE CHANGE

In the full sample of 227 pairs of interviews, more responses changed when the second interview was conversational (21.8 percent change from first interview) than when it was standardized (11.0 percent change), $F1(1,225) = 45.61$, $p < .001$. This was true for all 10 survey questions, $F2(1,18) = 4.60, p < .05$.[3] Responses changed more in conversational second interviews regardless of respondents' gender (interaction of gender and second interview type, $F(1,224) = 0.93$, n.s.), race (interaction of race and second interview type, $F(3,224) = 0.83$, n.s.), or educational level (interaction of educational level and second interview type, $F(4,224) = 0.36$, n.s.).

The 11 percent rate of response change across the two standardized interviews is within the normal range for reinterviews in large government surveys like those sponsored by the Bureau of Labor Statistics[4] and probably reflects ordinary memory or reporting errors. The additional 11 percent change for conversational interviews must result from the difference in interviewing technique.

In the subsample of 35 transcribed pairs of interviews the pattern of response change was similar. Respondents in standardized second interviews changed 10 of 174 answers from the standardized first interviews, or 5.7 percent; respondents in conversational second interviews changed 35 of 165 answers from the standardized first interviews, or 21.2 percent. This gives us additional confidence that the subsample is representative of the full sample of 227.

MEASURE 2: "LEGAL" REPORTED PURCHASES

Whenever respondents answered yes to a purchase question, they were asked to list the purchases on which they based their response. We coded all the listed purchases recorded by interviewers in the full sample of 227 pairs of interviews, determining whether they were consistent with the official definitions ("legal"), inconsistent with the definitions ("illegal"), or uncodable. Codes were assigned only on the basis of what interviewers had recorded and independent of what was recorded for the other interview with the same respondent.

3. The statistic $F2$ indicates an analysis of variance with items (questions) as the random factor.
4. For example, in Current Population Survey reinterviews carried out between 1995 and 1997, 3.7 percent of respondents changed their answer to a question about full-time work status; 33.7 percent changed their answer to a question about being unemployed (McGovern and Bushery 1999).

Although the official definitions are quite explicit, we used the following additional conventions for borderline cases: (1) For the home maintenance and repair question, the official definition excludes work carried out by the respondent. For this question we accepted ambiguous responses like "carpeting" and "tiles" as legal unless there was explicit evidence that the respondent had done the work himself or herself, or had paid for materials only, or was remodeling (all of which are excluded). (2) For the telephone and furniture questions, we counted unquantified plurals ("phones" or "chairs") as two items. (3) We classified reported purchases like "living room furniture" as uncodable because they were too general to determine if they included "illegal" items.

In the first interview, 57 percent of respondents' purchases were consistent with the official definitions. The same proportion of reported purchases were legal (57 percent) when the second interview was standardized, but a far greater proportion (95 percent) were legal when the second interview was conversational, interaction $F(1,206) = 86.62$, $p < .0001$. This advantage for conversational interviewing was not because the respondents in the conversational interviews already happened to conceive of the purchases in the same way as the sponsoring organization: they were no more likely to have reported legal purchases in the first interview (58 percent) than were their standardized counterparts (56 percent). Thus, conversational interviewing did seem to reduce respondents' misconceptions.

The nature of respondents' misconceptions differed for different purchase questions. For some questions, most misconceptions clustered around one or two components of the official definitions. For others they were more widespread. To examine this we first broke each definition into its component parts. For example, the definition for "telephone or telephone accessories" consisted of six components: (*a*) telephone purchases, (*b*) include portable cellular phones, (*c*) include telephone answering devices, (*d*) include dialing devices, (*e*) include telephone accessories, (*f*) exclude payments for telephone service. Then we classified the 227 respondents' illegal or discrepant[5] purchase explanations according to which components of the definitions had been misunderstood. For example, if a respondent listed "paid my bills" to explain a "yes" response to the telephone purchase question, we classified it as a misunderstanding of component *f*. Similarly, if a respondent listed a cellular phone in only one interview, this was classified as a problem with component *b*. If a particular purchase explanation contained more than one illegal or discrepant justification it was classified as a problem with multiple components of the corresponding definition.

The results are displayed in table 1. For question 6, respondents mostly misunderstood two components. Sixty-eight (49 percent) of the 139 misconceptions concerned component 6*a*, mostly resulting from respondents' belief

---

5. Discrepant explanations are legal but are provided in only one of the two interviews.

Table 1. Misunderstandings of Question Concepts in Respondents' Listed Purchases

Concept and Definition Component	Number of Misunderstandings
6. Moving:	
a) moving	68
b) fees paid to	3
c) professional movers	62
d) include packing	0
e) include freight	0
f) include storage	5
g) include parcel delivery service	1
h) exclude U.S. Postal Service delivery	0
7. Telephone or telephone accessories:	
a) telephone purchases	55
b) include portable cellular phones	9
c) include telephone answering devices	31
d) include dialing devices	2
e) include telephone accessories	35
f) exclude payments for telephone service	81
8. Inside home maintenance or repair services:	
a) home	0
b) inside	35
c) work (maintenance or repair)	62
d) you paid	0
e) someone else to do, not work you did yourself	10
f) include inside painting	5
g) include plastering	8
h) include plumbing	36
i) include electrical	10
j) include insulation	5
k) include heating or air conditioning	27
l) include floor repair	5
m) include pest control	7
n) include service contracts	3
o) exclude home improvements	29
p) exclude new construction	8
q) exclude appliances	41
9. Household furniture:	
a) household	1
b) furniture	37
c) include tables	18
d) include chairs	24
e) include footstools	2
f) include sofas	22
g) include china cabinets	4

Concept and Definition Component	Number of Misunderstandings
h) include utility carts	1
i) include bars	0
j) include room dividers	0
k) include bookcases	5
l) include desks	10
m) include beds	21
n) include mattresses	17
o) include box springs	15
p) include chests of drawers	11
q) include night tables	7
r) include wardrobes	3
s) include unfinished furniture	0
t) exclude TV	12
u) exclude radio	1
v) exclude other sound equipment	6
w) exclude lamps and lighting fixtures	4
x) exclude outdoor furniture	2
y) exclude infants' furniture	1
z) exclude appliances	24
10. Whiskey or other liquors for home use:	
a) whiskey and other liquors	19
b) purchased	0
c) for consumption at home, rather than that purchased in a restaurant or bar for consumption there	0
d) exclude wine	44
e) exclude beer	57
f) exclude ale	0

that the category "moving expenses" included the cost of transporting themselves (gas, lodging, food, etc.). Sixty-two (45 percent) of the 139 misconceptions concerned component 6*c*, where respondents answered "yes" because they incurred do-it-yourself moving expenses like truck rental. Misconceptions for question 10, and to a lesser extent for question 7, also clustered around a few components of the definition. The misconceptions for questions 8 and 9 were far more variable. In question 8 respondents had misconceptions for 15 of 17 components of the definition, and for question 9 respondents had misconceptions for 23 of 26 components.

WHY DID RESPONDENTS CHANGE THEIR REPORTS OF PURCHASES?

By making sure that respondents understood the official definitions, conversational interviewers helped respondents to classify their purchases more accurately. Respondents in conversational interviews included purchases they should have included the first time, and they omitted purchases they had mistakenly included the first time. That is, in the 31 cases where respondents changed their answer to a purchase question in the second interview to "yes," the newly reported purchases were legal 90 percent of the time. In the 59 cases where respondents changed their answer in the second interview to "no," the previously included (now omitted) purchases were legal only 21 percent of the time.

In contrast, when respondents in standardized interviews changed their responses to a purchase question in the second interview, the second answer was no more likely to reflect a legal interpretation than an illegal one. When standardized respondents changed their answer to "yes" (22 cases), the newly reported purchases were legal 47 percent of the time. When respondents changed their answer in the second interview to "no" (25 cases), the previously included (now omitted) purchases were just as likely to be legal (52 percent).[6]

Closer analysis of the subsample of 35 transcribed interviews gives essentially the same picture as the coding of reported purchases in the full sample, this time for all 10 questions (i.e., for housing questions too). Consider this pair of excerpts from the same respondent. In the first interview (standardized), the interviewer leaves the interpretation of the question up to the respondent:[7]

(1) I: And how many people live in your home?
R: Five.
I: Five, okay.
R: Well.
I: Uh-huh.
R: We have a, I have a sister that's also in college.
I: Okay, so how so what would you consider.
R: She lives here during the summer and then, she she lives here.
I: Okay so how many would you *say live*.
R: *Five*.
I: Five. Okay.

6. Note that Measure 2 is flawed in one respect: respondents reported purchases only when they answered "yes" to the survey question, and so we cannot know respondents' level of accuracy in cases where they answered "no" both times.
7. In the transcribed excerpts, overlapping speech is enclosed in asterisks. A period between two spaces (.) represents a pause. A colon within a word indicates a lengthened sound. A hyphen at the end of a word ("that-") indicates that the word was cut off. Question marks indicate rising intonation, and utterance-final periods indicate falling or flat intonation, regardless of whether the utterance is a question or an assertion.

In the second (conversational) interview, the same respondent changes the answer because of the (different) interviewer's intervention:

(2) I: Okay, now, how many people live in your home?
R: Five.
I: Five people, okay. And uh uh we uh when I say how many people live in your home what I mean by this is, I don't want you to include uh people that are overnight lodgers, guests, visitors or people who are living away on business, in the armed forces or attending school.
R: Okay then four. We have one that's in school.
I: That's in school okay. So that's four people living in your home. Okay.

In the subsample, every time that respondents in conversational interviews provided different answers than in their first interviews (35 times), interviewers had provided at least one nonstandard intervention for that question in the second interview. In contrast, when respondents in standardized interviews provided different answers than in their first interviews (10 times), in only one of those cases had the second interviewer provided any nonstandard intervention.

One concern that proponents of strictly standardized interviewing have raised about conversational interviewing is that interviewers can mislead respondents (e.g., Fowler and Mangione 1990). That is, even if interviewers sometimes provide information that helps respondents to produce accurate answers, interviewers may just as often provide information that can lead respondents astray. Also, some interviewers may be more accurate than others, which might lead to greater interviewer-related variability in responses.

This does not seem to have been the case here, as seen in our subsample of 35 transcribed interviews. The five conversational interviewers provided accurate official definitions (verbatim or improvised) for 116 of the 121 questions where they presented definitions, an accuracy rate of 95 percent. No single interviewer was less accurate than the others: each interviewer presented one inaccurate definition. The one standardized second interviewer who (illegally) provided a definition presented inaccurate information.

This finding suggests that, as in the Schober and Conrad (1997) laboratory study, conversational interviewers can present highly accurate information to respondents, contrary to Fowler and Mangione's (1990) concerns. Of course, whether all interviewers could be trained to perform so well remains to be seen.

INTERVIEWER VARIABILITY

Another concern raised by proponents of standardized interviewing is that conversational interviewing will lead to undesirable levels of error due to interviewer variability. One way to test interviewer variability is to use rho-int, as described in Fowler (1991), which requires an experimental design in which each interviewer's respondents are representative of the larger sample. Unfortunately, this was not under our control in this experiment, and so we are unable to use rho-int. As a surrogate measure we carried out an analysis of variance with interviewer as the random factor and response as the dependent measure, for those questions with numerical answers (the housing questions). We found no interviewer effects in either conversational or standardized interviews; responses did not differ for different interviewers in standardized second-week interviews, $F(1,4) = 0.68$, n.s., nor did they differ in conversational second-week interviews, $F(1,4) = 1.24$, n.s.

Further testing with a larger sample and a design that allows the researchers to calculate rho-int is needed to determine more definitively how conversational interviewing influences interviewer effects. We propose that if conversational interviewing does, indeed, standardize meaning, then there is no reason that the more variable interviewer *behavior* that the technique entails will necessarily lead to greater variation between interviewers in the *responses* they collect. In theory, it should lead to reduced interviewer effects by promoting uniform interpretation of questions, regardless of who asks them.

WHEN DID CONVERSATIONAL INTERVIEWERS INTERVENE?

As in the Schober and Conrad (1997) laboratory experiment, conversational interviewers in this study were trained to clarify question meaning both when respondents explicitly asked for clarification and also whenever else the interviewers felt this might be needed. In the subsample of 35 transcribed interviews, conversational interviewers provided clarification in each of the very few cases (6 of 165 questions, 4 percent) that respondents explicitly asked for it. In the vast majority of the 143 cases where conversational interviewers clarified question meaning, they did so without an explicit request from the respondent (137 cases, or 96 percent of the time).

In some of these cases, conversational interviewers seemed to be responding to evidence that the respondent might be having trouble. For example, in 11 of the 165 cases (7 percent), interviewers provided clarification after respondents reported their circumstances instead of answering the question directly, as in this example where the question requires a "yes" or "no" answer and instead the respondent says "We've had some plumbing done":

> (3) I: uh and in the past five years that is since July of nineteen ninety one have you purchased or had expenses for inside home maintenance or repair services.

R: Um, we've had some plumbing work done.
I: Okay, now here again . uh like in the moving . it is- include only the inside work you paid someone else to do, not work you did yourself.
R: Oh, okay, someone else did.
I: Yea:h *(laughs)*
R: *Oh, oh, okay*
I: That's for the purposes of this survey.
R: Okay, when they asked me last time I think it was what we had done.
I: Uh huh.
R: Okay. No, no one else has ever worked inside the house.
I: Oka:y, all:right.
R: that we've
I: And that would include service contracts *for*
R: *Yes, that's right*
I: heating or air conditioning? Okay,
R: *no ma'am*
I: *so there's noth*ing, all right.

In 3 of the 165 cases (2 percent), interviewers provided clarification after respondents asked them to repeat the survey question, which might also have been evidence of respondent misunderstanding.

But in most cases interviewers clarified the meaning of the question even though respondents did not seem to have given any explicit evidence that they needed clarification. Interviewers tended to do this in two ways. They would immediately provide clarification after a respondent provided an answer (20 of 165 cases, 12 percent), as in example 2 earlier. Or they would present problematic parts of definitions immediately after they first read questions as a "preemptive strike" (73 of 165 cases, 44 percent):

(4) I: And how many other rooms are there other than bedrooms and bathrooms? And by that I mean um living room, dining room, kitchens, um lodgers' rooms, finished basements, or attic rooms, recreation rooms, or and or permanently enclosed sun porches, but um . these rooms would be um considered separate rooms th- as long as there's a permanent partition from floor to ceiling between the two areas? So, how many- with those definitions, how many um a how many rooms are there other than bedrooms and bathrooms?
R: One, is it kitchen and front room combined then, because there's no partition between them.
I: Okay, so the living room, dining room, kitchen is all one, is all in one.
R: Yes.
I: Okay, and um there are no other um sun porches, or attic rooms, or recreation room or
R: No.
I: Okay. Um, so that and there's . no permanent partition from floor to ceiling between the kitchen and the living room dining room.
R: No.
I: Okay, so I'm going to put one in.

This technique may have allowed interviewers to avoid lengthier clarification sequences later. It may also have reminded respondents that official definitions might differ from their own, even if the part of the definition presented was not relevant to the respondents' personal situations.

Note that although response change was always preceded by conversational interviewers' nonstandard interventions, nonstandard interventions did not always lead to changed responses. In 108 out of the 143 cases (76 percent) where interviewers deviated from strict standardization, respondents did not change their answers. This suggests that one cost of conversational interviewing may be that interviewers provide information that is not necessarily helpful (it does not change the answer) and that just lengthens the interview. How well respondents tolerate this is an open question.

Although they had the same training, conversational interviewers in this study intervened under different circumstances than those in the Schober and Conrad (1997) laboratory experiment. In that study interviewers often waited for evidence that respondents were having trouble interpreting questions. Here interviewers intervened without provocation, perhaps because respondents so rarely showed they had trouble with questions. It seems that conversational interviewers made their own judgments about the ways in which questions were likely to be misunderstood, and checked to see if respondents' circumstances were unusual. Better pretesting might allow researchers to design questions that include more such preemptive strikes ("and by furniture we mean . . ."). The problem with this approach, we argue, is that pretesting cannot rule out idiosyncratic misunderstandings. This follows from the idea that meaning does not reside exclusively in words but also in their correspondence to objects and concepts, in this case, the respondents' circumstances. As we propose later, even if pretesting could identify all possible misconceptions by all respondents, the "repairs" to the question would lead to extremely long interviews—perhaps longer than our conversational interviews.

INTERVIEW DURATION

Conversational interviews took 80 percent longer than standardized interviews. The median time for all respondents to complete the first interview (which was standardized for all respondents) was 5 minutes (timed in minutes by interviewers themselves). The median time to complete the second standardized interview was also 5 minutes, but it took 9 minutes to complete second conversational interviews, interaction of interview number (first or second) and second interview type (standardized or conversational), $F(1,224) = 60.88$, $p < .0001$.

Some of the increased duration for conversational interviews was the result of time that interviewers spent explaining the conversational interviewing

technique to respondents. Despite this, the 80 percent increase in duration in the current study is far less than the 300 percent increase in duration found in the Schober and Conrad (1997) laboratory experiment, where interviewers did not spend any time explaining conversational interviewing techniques. Why might this be? There are at least two possibilities. First, since conversational interviewers in this study were calling respondents in their own homes, they no doubt felt pressure to complete interviews quickly that may have been missing in the Schober and Conrad (1997) laboratory study. Second, time spent in conversational interviews presumably increases when complicated mappings are more frequent. In the Schober and Conrad (1997) experiment the mappings between the questions and the scenarios from which respondents were answering were complicated fully 50 percent of the time. In the current study, mappings were not controlled, and complicated mappings may well have been less frequent.

Discussion

Both strictly standardized and conversational interviewing techniques are intended to produce comparable responses. In one case this is achieved by holding the words constant, and in the other by permitting conversation to promote uniform interpretation. Our data show that conversational interviews actually led to more comparable responses than standardized interviews. Conversational interviewers helped respondents apply the concepts to their circumstances along the lines the survey designers intended, and this produced the intended understanding substantially more often. This leads to the seemingly paradoxical conclusion that comprehension can be made more consistent—and responses more comparable—when certain interviewer behaviors (discussions about the meaning of questions) are *less* consistent.

Our results also indicate that, surprisingly often, respondents can interpret seemingly straightforward questions differently than the survey designers intend. This happened less frequently in the present study than it did in the Schober and Conrad (1997) laboratory study, where, for purposes of experimental design, mappings between respondents' circumstances and questions were complicated 50 percent of the time. But in the present study complicated mappings *were* frequent enough to compromise comprehension, and thus response accuracy, at levels high enough to warrant concern. We have strong evidence for this in the fact that fewer than 60 percent of purchases listed by respondents in strictly standardized interviews should actually have been included, according to official definitions.

Conversational interviewing can reduce this form of measurement error, at least for some questions under some circumstances. But many aspects of its implementation remain untested. It remains to be seen whether all interviewers—or even a larger group of interviewers—would be as competent and

consistent as our conversational interviewers, especially for long question-naires with many complex definitions. It also remains to be seen whether respondents in a full-length real-world interview of 30 or 60 minutes would be as cooperative and motivated as those in this sample.

Is conversational interviewing the only—or the optimal—solution to the problem of respondent misconceptions? Obviously the increased interview duration that conversational interviewing seems to require is an obstacle to its implementation, along with the costs for better interviewer training and monitoring, increased investment in concept development, uncertain effects on response and completion rates, etc. Might there be a less costly solution? Might there be a more standardized solution?

Several alternatives each have advantages and disadvantages:

1. *Reword questions to include the parts of the definition that lead to misconceptions*. Perhaps the reason we saw so many misconceptions in this study was that the questions were poorly worded (their pretesting notwith-standing). Through more extensive pretesting one could determine which parts of the definition are most likely to cause misconceptions, and reword the questions accordingly. For example, using the information in table 1, one could reword our question 6 like this (additions in italics): "In the past 5 years, that is, since July of 1991, have you purchased or had expenses for *professional* moving, *not including your own transportation costs*?" With this wording, respondents in our study might well have understood the moving concept as intended on 130 more occasions than they did with the original wording.

This solution should work if (*a*) all (or most) respondents' misconceptions concern only a few components of the underlying concept and (*b*) miscon-ceptions about a particular component are similar from one respondent and one occasion to the next. Condition *a* is necessary in order to keep questions from getting too long. If respondents misunderstand many aspects of a concept then it would take a great deal of text to resolve all of these misunderstandings, which would lead to unwieldy, possibly incomprehensible questions. Con-dition *b* is necessary because one cannot write questions to prevent unknown misconceptions. If misconceptions are idiosyncratic or intermittent, then future misconceptions are likely to differ from those that have already been observed; standardized rewording is unlikely to resolve idiosyncratic misconceptions.

This solution would not work particularly well for a question like our question 8. As table 1 shows, respondents had trouble with almost every component of this definition. For such situations, one might try this stan-dardized solution:

2. *Ask about problematic parts of the definition in a series of questions*. This solution has the advantage of guaranteeing that all respondents consider all potentially problematic parts of a definition, although it has the disadvan-tage of increasing interview duration at least as much as conversational in-terviewing would—and for all respondents, whether or not they need the

clarification. Just as with conversational interviewing, inordinate time may be spent on unnecessary clarification. And although the situations are not perfectly comparable, preliminary evidence from a study in which respondents received complete definitions for each survey question (Bloom 1999; Schober, Conrad, and Bloom 1999) suggests that respondents may find such extensive probing aversive.

3. *Rethink the definitions*. Perhaps the reason we saw so many misconceptions in this study is that our questions involved unusually complicated or counterintuitive definitions. Perhaps redesigned questions or series of questions that rest on more intuitive concepts (as determined by pretesting or ethnographic methods—see Gerber [1999]; Turner and Martin [1984, chap. 7] for examples) could successfully be administered using standardized techniques.

However, the definitions we chose in this experiment were not unusually complicated for survey definitions.[8] And survey designers can have legitimate reasons for defining terms as they do. Although a concept like "other rooms, other than bedrooms and bathrooms" probably does not correspond to a preexisting mental category for most respondents, it is an important determinant of rental costs and thus is worth measuring. Similarly, "inside home maintenance and repair services" is probably not a natural class of expenditures for most respondents, but it has econometric coherence to the researchers conducting the survey and thus may warrant a dedicated question.

More important, it is unrealistic to expect that survey designers can always define terms in ways that match everyone's conceptions. Everyday notions about survey concepts can vary substantially, either across people or for the same people on different occasions (e.g., Belson 1981, 1986; Gerber, Keely, and Wellens 1996; Martin, Campanelli, and Fay 1991). When this is the case, this solution will not work.

Conclusions

As we see it, far more remains to be investigated about potential solutions to the problem of respondent misconceptions. Several important factors need to be disentangled. We need to distinguish, at least, (*a*) the quantity of information about a concept that an interviewer gives a respondent, (*b*) the form in which that content is presented (scripted or unscripted, as a series of questions or as a series of statements), and (*c*) who initiates clarification (the interviewer,

8. The definitions for both the housing and purchase questions are comparable in length (averaging 120.0 and 34.4 words, respectively) to the full set of CPI-Housing and CPOPS definitions from which they were drawn (averaging 102.7 and 31.1 words, respectively), and they are similar in length to definitions from other major federal surveys. For example, definitions in the Current Population Survey, the primary labor force survey in the United States, average 46.8 words in length, and definitions in the National Crime Victimization Survey, the primary source of information about crime incidents, victims, and trends in the United States, average 39.8 words.

by presenting unsolicited information, or the respondent, by asking for clarification). In the current study we have contrasted response accuracy in a situation where no clarification is given to a situation where an interviewer can present any amount of information in any form. We have begun to examine other kinds of situations in other experiments, for example, contrasting cases where interviewers respond only to explicit requests for clarification to cases where they can offer unsolicited clarification, and contrasting cases where interviewers can choose their wording to cases where they must read definitions verbatim (Schober, Conrad, and Fricker 1999). We have also contrasted cases where the respondent initiates clarification to cases where a computerized survey system also initiates clarification (Bloom 1999; Conrad and Schober 1999; Schober, Conrad, and Bloom 1999).

Beyond misinterpretation of survey concepts, strictly standardized question presentation may lead to misunderstanding in other arenas which might be addressed by conversational interviewing. For example, conversational techniques might help respondents correctly interpret the response task (Conrad 1999). Consider a respondent who is asked, "Last month how many meals did you buy in restaurants?" She may wonder whether to recall and count all such episodes, retrieve the rate at which she buys restaurant meals, or report a general impression of the number. If survey designers are clear about how the response task should be carried out, conversational interviewers might help this respondent select an approach.

It would be a mistake to argue that our current results extend beyond the sort of factual or behavioral questions typical of large government surveys; in particular, we do not intend our results to extend to opinion or attitude questions, nor to factual questions that concern sensitive topics like number of sexual partners or intravenous drug use. Although conceivably conversational interviewers could be trained to explain what survey designers mean by terms like "abortion," "the presidency," "often," or even "approve," the psychological mechanisms involved in answering attitude questions may differ fundamentally from those involved in answering behavior questions (e.g., Sudman, Bradburn, and Schwarz 1996). Nonetheless, such words are as open to interpretation as "bedroom" and "household furniture" are, and it is not clear that respondents all interpret these words the same way. But it is also unclear whether such questions could be clarified without influencing responses.

But in the case of nonsensitive factual or behavioral questions, our data suggest that conversational interviewing is an approach worth exploring further, particularly when the frequency of complicated mappings (and thus of potential misconceptions) is unknown. If complicated mappings are known to be rare, or if less precision is tolerable, presenting questions in a strictly standardized fashion, with no clarification, may suffice. If complicated mappings are known to be frequent, standardized solutions like always probing all problematic aspects of definitions may be optimal. In many cases, it is

impossible to know the frequency of complicated mappings; in those cases, conversational interviewing may turn out to be the more efficient technique for assuring accurate comprehension.

Appendix A

Questions (One of Two Orderings)

First I would like to ask you a few questions about housing.

1. How many bedrooms are there in your home?
2. How many full bathrooms are there in your home?
3. How many half bathrooms are there?
4. How many other rooms are there, other than bedrooms and bathrooms?
5. How many people live in your home?

The next set of questions is about purchases that you or anyone in your household may have made during different time periods. So when I ask if YOU have made a purchase, I mean you or anyone in your household.

6. In the past five years, that is, since July of 1991, have you purchased or had expenses for moving?
   IF YES: What purchases or expenses for moving did you have?
7. In the past five years, that is, since July of 1991, have you purchased or had expenses for telephone or telephone accessories?
   IF YES: What purchases or expenses for telephone and telephone accessories did you have?
8. In the past five years, that is, since July of 1991, have you purchased or had expenses for inside home maintenance or repair services?
   IF YES: What purchases or expenses for inside home maintenance or repair services did you have?
9. In the past year, that is, since July of 1995, have you purchased or had expenses for household furniture?
   IF YES: What purchases or expenses for household furniture did you have?
10. In the past year, that is, since July of 1995, have you purchased or had expenses for whiskey or other liquors for home use?
    IF YES: What purchases or expenses for whiskey or other liquors for home use did you have?

Appendix B

Official Definitions

CPI-HOUSING DEFINITIONS (FOR HOUSING QUESTIONS)

1. Bedrooms. A bedroom is a finished room specifically designed by the owner to be used for sleeping. A bedroom does not have to be used for sleeping in

order to qualify as a bedroom. For example, a bedroom that is being used as an office should be counted as a bedroom.

Do not count as a bedroom any room that was designed for another purpose but is being used as a bedroom. For example, a den being used as a bedroom is still a den and should not be counted as a bedroom.

Do not count as a bedroom any dens, living rooms, or other rooms that can be converted at night for sleeping.

Do not count any bedroom that the renter is denied access to or use of by the owner.

A one-room efficiency apartment does not have a bedroom.

2 and 3. Bathrooms. A full bathroom has (1) a flush toilet, (2) a bathtub or shower, and (3) a sink or washbasin with running water. Bathrooms that contain all of the above items, whether separated by a partition or door, are to be considered a full bathroom.

A half bathroom has any two of these three items: (1) a flush toilet, (2) a bathtub or shower, and (3) a sink or washbasin with running water.

If the only bathroom facilities do not meet the definition of a full or half bath, code zero. (For example, if there is only a flush toilet in a room.)

If a bathroom is shared by the occupants of more than one housing unit, the bathroom is included with the unit from which it is most easily reached.

4. Other Rooms, Other Than Bedrooms and Bathrooms. Include whole rooms such as living rooms, dining rooms, kitchens, lodger's rooms, finished basements or attic rooms, recreation rooms, and permanently enclosed sun porches. Rooms used for offices by a person living in the unit are also included in this survey. Rooms are counted even if they are not used.

Do not include bedrooms, bathrooms, unfinished attics or basements, halls, foyers or vestibules, balconies, closets, alcoves, pantries, strip or pullman kitchens, laundry or furnace rooms, open porches, and unfinished spaces used for storage.

A partially divided room, such as a dinette next to a kitchen or living room, is a separate room only if there is a permanent partition from floor to ceiling between the two areas. An L-shaped room, a "great" room, or a step-down is therefore counted as one room unless there is a permanent partition dividing the room into parts.

If a room is used by occupants of more than one unit, the room is included with the unit from which it is most easily reached.

Do not count any rooms that the renter is denied access to or use of by the owner. Do count rooms regardless of their year-round usability.

Bathrooms: exclude all bathrooms. While some rooms, such as a small room with only a wash basin, do not meet the definition of a bathroom, they are also to be excluded from the count of other rooms.

5. Living in a Housing Unit. A person is considered to be living in a housing unit even if the person is not present at the time of the survey. Live-in servants or other employees, lodgers, and members of the household temporarily away from the unit on business or vacation are included in the count.

Do not count any people who would normally consider this their (legal) address but who are living away on business, in the armed forces, or attending school (such as boarding school or college).

Do not count overnight lodgers, guests and visitors. Do not count day employees who live elsewhere.

cpops definitions (for purchase questions)

6. Moving. Fees paid to professional movers, including packing, freight and storage. Do not include the expenses involved if the respondent moves him/herself without professional help. Include parcel delivery service, except U.S. Postal Service.

7. Telephone or Telephone Accessories. Telephone purchases, including portable cellular phones; telephone answering devices; dialing devices; telephone accessories. Do not include payments for telephone service.

8. Inside Home Maintenance or Repair Services. Include only inside work you paid someone else to do, not work you did yourself. Include inside painting, plastering, plumbing, electrical, insulation, heating or air conditioning, floor repair, and pest control. Include service contracts. Do not include home improvements, new construction, or appliances.

9. Household Furniture. Tables, chairs, footstools, sofas, china cabinets, utility carts, bars, room dividers, bookcases, desks, beds, mattresses, box springs, chests of drawers, night tables, wardrobes, and unfinished furniture. Do not include TV, radio, and other sound equipment, lamps and lighting fixtures, outdoor furniture, infants' furniture, or appliances.

10. Whiskey or Other Liquors for Home Use. Whiskey and other liquors purchased for consumption at home, rather than that purchased in a restaurant or bar for consumption there. Do not include wine, beer, or ale in this consumer item.

References

Akmajian, Adrian, Richard A. Demers, Ann K. Farmer, and Robert M. Harnish. 1990. *Linguistics: An Introduction to Language and Communication*. 3d ed. Cambridge, MA: MIT Press.

Beatty, Paul. 1995. "Understanding the Standardized/Non-Standardized Interviewing Controversy." *Journal of Official Statistics* 11:147–60.

Belson, William A. 1981. *The Design and Understanding of Survey Questions*. Aldershot: Gower.

———. 1986. *Validity in Survey Research*. Aldershot: Gower.

Bloom, Jonathan E. 1999. "Linguistic Markers of Respondent Uncertainty during Computer-Administered Survey Interviews." Ph.D. dissertation, New School for Social Research.

Cannell, Charles F., Peter V. Miller, and Lois Oksenberg. 1981. "Research on Interviewing Techniques." In *Sociological Methodology*, ed. S. Leinhardt, pp. 389–437. San Francisco: Jossey-Bass.

Clark, Herbert H. 1992. *Arenas of Language Use*. Chicago: University of Chicago Press.

———. 1996. *Using Language*. Cambridge: Cambridge University Press.

Clark, Herbert H., and Susan E. Brennan. 1991. "Grounding in Communication." In *Perspectives on Socially Shared Cognition*, ed. Lauren B. Resnick, John M. Levine, and Stephanie D. Teasley, pp. 127–49. Washington, DC: American Psychological Association.

Clark, Herbert H., and Michael F. Schober. 1991. "Asking Questions and Influencing Answers." In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, ed. Judith M. Tanur, pp. 15–48. New York: Russell Sage Foundation.

Clark, Herbert H., and Deanna Wilkes-Gibbs. 1986. "Referring as a Collaborative Process." *Cognition* 22:1–39.

Conrad, Frederick G. 1999. "Customizing Survey Procedures to Reduce Measurement Error." In *Cognition and Survey Research*, ed. Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith M. Tanur, and Roger Tourangeau, pp. 301–17. New York: Wiley.

Conrad, Frederick G., and Michael F. Schober. 1999. "A Conversational Approach to Text-Based Computer-Administered Questionnaires." In *Proceedings of the Third ASC International Conference*, pp. 91–102. Chesham, UK: Association for Survey Computing.

Dykema, Jennifer, James M. Lepkowski, and Steven Blixt. 1997. "The Effect of Interviewer and Respondent Behavior on Data Quality: Analysis of Interaction Coding in a Validation Study." In *Survey Measurement and Process Quality*, ed. Lars Lyberg, Paul Biemer, Martin Collins, Edith de Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin, pp. 287–310. New York: Wiley.

Fowler, Floyd J. 1991. "Reducing Interviewer-Related Error through Interviewer Training, Supervision, and Other Means." In *Measurement Error in Surveys*, ed. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, pp. 259–78. New York: Wiley.

Fowler, Floyd J., and Thomas W. Mangione. 1990. *Standardized Survey Interviewing: Minimizing Interviewer-Related Error*. Newbury Park, CA: SAGE Publications.

Gerber, Eleanor. 1999. "The View from Anthropology: Ethnography and the Cognitive Interview." In *Cognition and Survey Research*, ed. Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith M. Tanur, and Roger Tourangeau, pp. 217–34. New York: Wiley.

Gerber, Eleanor, Catherine Keeley, and Tracy Wellens. 1996. "The Use of Vignettes in Evaluating Household Roster Information: Does Anybody Read the Rules?" In *Proceedings of the American Statistical Association, Section on Survey Methods Research*, pp. 1058–63. Alexandria, VA: American Statistical Association.

Martin, Elizabeth A., Pamela C. Campanelli, and Robert E. Fay. 1991. "An Application of Rasch Analysis to Questionnaire Design: Using Vignettes to Study the Meaning of 'Work' in the Current Population Survey." *Statistician* 40:265–76.

McGovern, Pamela, and John M. Bushery. 1999. "Data Mining the CPS Interview: Digging into Response Error." In *Proceedings of the Research Conference of the Federal Committee on Statistical Methodology, Monday B Sessions*, pp. 76–85. Washington, DC: Federal Committee on Statistical Methodology.

Mishler, Elliot G. 1986. *Research Interviewing*. Cambridge, MA: Harvard University Press.

Schaeffer, Nora Cate. 1991. "Conversation with a Purpose—or Conversation? Interaction in the Standardized Interview." In *Survey Measurement and Process Quality*, ed. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, pp. 367–91. New York: Wiley.

Schober, Michael F. 1998. "Conversational Evidence for Rethinking Meaning." *Social Research* 65:511–34.

———. 1999. "Making Sense of Questions: An Interactional Approach." In *Cognition and Survey*

Research, ed. Monroe G. Sirken, Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith M. Tanur, and Roger Tourangeau, pp. 77–93. New York: Wiley.

Schober, Michael F., and Herbert H. Clark. 1989. "Understanding by Addressees and Overhearers." *Cognitive Psychology* 21:211–32.

Schober, Michael F., and Frederick G. Conrad. 1997. "Does Conversational Interviewing Reduce Survey Measurement Error?" *Public Opinion Quarterly* 61:576–602.

———. 1998. "Response Accuracy When Interviewers Stray from Standardization." In *Proceedings of the American Statistical Association, Section on Survey Methods Research*, pp. 940–45. Alexandria, VA: American Statistical Association.

———. In press. "A Collaborative View of Standardized Survey Interviews." In *Standardization and Tacit Knowledge: Interaction and Practice in the Survey Interview*, ed. Douglas Maynard, Hanneke Houtkoop, Nora Cate Schaeffer, and Johannes van der Zouwen. New York: Wiley.

Schober, Michael F., Frederick G. Conrad, and Jonathan E. Bloom. 1999. "Enhancing Collaboration in Computer-Administered Survey Interviews." In *Proceedings of the American Association for Artificial Intelligence Fall Symposium "Psychological Models of Communication in Collaborative Systems,"* pp. 108–15. Menlo Park, CA: AAAI Press.

Schober, Michael F., Frederick G. Conrad, and Scott S. Fricker. 1999. "When and How Should Survey Interviewers Clarify Question Meaning?" In *Proceedings of the American Statistical Association, Section on Survey Methods Research*, pp. 986–991. Alexandria, VA: American Statistical Association.

Suchman, Lucy, and Brigitte Jordan. 1990. "Interactional Troubles in Face-to-Face Survey Interviews." *Journal of the American Statistical Association* 85(409):232–53.

———. 1991. "Validity and the Collaborative Construction of Meaning in Face-to-Face Surveys." In *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*, ed. Judith M. Tanur, pp. 241–67. New York: Russell Sage Foundation.

Sudman, Seymour, Norman Bradburn, and Norbert Schwarz. 1996. *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass.

Turner, Charles E., and Elizabeth Martin, eds. 1984. *Surveying Subjective Phenomena*. Vol. 1. New York: Russell Sage Foundation.

U.S. Bureau of the Census. 1997. *Introduction to CATI: Training Guide for New Interviewers*. Hagerstown, MD, Tucson, AZ, and Jeffersonville, ID Telephone Centers.

U.S. Department of Commerce. 1994. *Current Population Survey Interviewing Manual* (CPS-250). Washington, DC: Bureau of the Census.