When Do Respondent Misconceptions Lead to Survey Response Error?

Anna Suessbrick[1], Michael F. Schober[1], and Frederick G. Conrad[2]
New School for Social Research[1]
Institute of Social Research, University of Michigan[2]

Abstract

When interviewers clarify question concepts in a survey interview, respondents' conceptual alignment with the interviewer and answer reliability can improve (Conrad & Schober, 2000; Schober & Conrad 1997; Schober, Conrad, & Fricker, 2004). However, clarification may not always help, and it would be important to understand when and why it does. We analyzed 131 telephone survey interviews about tobacco use, tracking when and how interviewers clarified survey terms for respondents. We also had access to respondents' interpretations of the survey concepts as well as whether their responses changed in a re-interview in which standard definitions were provided. Results indicated that although respondents' conceptualizations were highly variable, matching official definitions only 44% of the time, respondents' answers were nonetheless often accurate, at a rate of 86%. While clarification sometimes improved responding, it did not guarantee accurate answers; nearly 12% of respondents' answers were unreliable despite their having received clarification, and sometime clarification could even be harmful. Clarification only improved responding when it was *helpful* to respondents in resolving their particular misconceptions. We describe alternative routes to accurate and inaccurate answers in survey interviews.

Keywords: question clarification, data quality, conversational interviewing, question interpretation, comprehension, measurement error, standardized interviewing

1. Introduction

Have you smoked at least one hundred cigarettes in your entire life?

At first glance, this survey question consists of unambiguous, non-technical concepts that should be easily and universally understood by respondents. Nevertheless, even straightforward question concepts are prone to variable interpretation. In one study, for example (Belson, 1981), 16% of respondents interpreted "you" in *How many hours of television do you watch each weekday?* to include other people, and 61% counted days other than the 5 weekdays. This finding illustrates what we call "conceptual variability"—speakers of the same language do not necessarily interpret concepts the same way.

Conceptual variability poses several problems for survey methodologists. If respondents interpret key question concepts differently than intended, then their question interpretation may vary such that they are not answering the questions we mean them to answer. Such variable interpretation may lead to responses that are not comparable to each other and that may be incorrect.

Conceptual variability in survey response has been examined in the lab (e.g., Schober & Conrad 1997; Schober, Conrad, & Fricker, 2004, Tourangeau, et al., in press) and in the field (e.g., Conrad & Schober, 2000). In the former studies, respondents answered survey questions on the basis of fictional scenarios. In the latter, a national telephone sample study, respondents answered questions on the basis of their own lives and experiences. In all of these, questions about facts and behaviors were excerpted or adapted from longer-length US government surveys. The evidence indicates that respondents do indeed vary substantially in their interpretations, that such variable interpretation leads to unreliable answers, and that better conceptual alignment improves answer reliability.

The results also suggested that conversational interviews, in which clarification of problematic concepts is provided, can reduce the rate of conceptual misalignment, leading to more accurate answers. Clarification can take many forms. Interviewers can give definitions of question concepts, fit the respondent's imprecise answer into one of the response alternatives, present a reduced set of response alternatives, etc. Clarification was found to be most effective at improving answer quality when the respondent's circumstances did not map in a straightforward way onto the question concepts. To illustrate, a government employee with a fixed schedule is likely to find the question *How many hours per week do you usually work?* comparatively more straightforward to interpret and answer than a freelancer whose hours vary substantially. Note that the findings in these studies suggest that the proscription against clarification in strictly standardized survey interviews (see, e.g., Fowler & Mangione, 1990), can lead to the un-

anticipated consequence that respondents sometimes misunderstand questions and answer inaccurately.

To assess the robustness and generalizability of our earlier results, we conducted a follow-up study using a full-length questionnaire from a national survey (Suessbrick, Schober, & Conrad, 2000). Respondents answered questions about attitudes as well as behaviors, based on their own life circumstances. The findings showed unexpectedly high rates of conceptual variability—not only between the interviewer and respondent but also across respondents. In fact, 10% of respondents took the wrong path through the questionnaire because of how they answered the initial filter question *Have you smoked at least 100 cigarettes in your entire life?*, as evidenced by their changing their answer to this question when presented with a standard definition.

The design of this study was intended to allow us to examine effects of clarification during interviews: some interviewers were trained to administer "conversational interviews" in which they were to provide clarification as needed. Despite the training, interviewers actually provided almost no clarification, making it impossible to evaluate the effects of clarification during telephone interviews.

The current study had two objectives. First, we wanted to administer conversational interviews that actually implemented the technique appropriately, in order to evaluate its effectiveness in a full-scale questionnaire. Second, we wanted to identify when misunderstandings compromise response accuracy and when clarification improves—or does not improve or harms—conceptual alignment and response accuracy.

The prototypical way that one might imagine clarification works in a survey interview is this:

Interviewer and respondent are not aligned
on an important survey concept
↓
Respondent receives clarification
from the interviewer
↓
Clarification resolves respondent's
misconception
↓
Respondent provides a more accurate answer

But this is not the only logical possibility. There are many different ways that survey participants can start and end a question-answer sequence (see Table 1) and this prototypical sequence is just one of several possible pathways (Path #5). For example:

- Path #1: Respondent's concept interpretation is aligned with the interviewer's (or, more accurately, the official definition), he receives no clarification, and respondent's final conceptualization remains aligned with the interviewer's.
- Path #6: Interviewer and respondent are misaligned on an important concept, respondent receives "clarification" that is not helpful at resolving his misconception, and respondent and interviewer remain misaligned.
- Path #3: Theoretically, respondent and interviewer could start the interview sequence conceptually aligned, unnecessary "clarification" is provided, which ends up harming the respondent's conceptualization, and respondent and interviewer end up misaligned.

Path	*I and R's Starting Conceptual Alignment*	*R Receives Clarification*	*Clarification Helpful?*	*Resulting Conceptual Alignment*
1.	Aligned	No	--	Aligned
2.	Aligned	Yes	Not Helpful	Aligned
3.	Aligned	Yes	Not Helpful	Misaligned
4.	Misaligned	No	--	Misaligned
5.	Misaligned	Yes	Helpful	Aligned
6.	Misaligned	Yes	Not Helpful	Misaligned

Table 1: Proposed conceptual coordination pathways. *I = Interviewer, R = Respondent*

On this view, simply receiving or not receiving clarification isn't enough. Instead, the effectiveness of clarification depends on how helpful it is, which depends on what the respondent's specific conceptual misalignment is (see also Lind, Schober, & Conrad, 2001). To be "helpful," clarification must pertain to the dimension of the question concept that is relevant to the respondent's circumstances. For example, even if a respondent conceives of work for pay differently than the survey designer, if the respondent has done nothing that could by any definition be considered work, this misalignment doesn't matter, and clarification about what counts as work won't be helpful. If a respondent interprets work differently than the survey designer (e.g., assumes that babysitting for different employers counts as having more than one job) but the interviewer provides clarification about an aspect of the survey concept not relevant to the respondent (e.g., about chauffeurs and gardeners), clarification will not be helpful.

2. Current Study

2.1 Methodology

We carried out new interviews using the same questionnaire and procedure as in Suessbrick, Schober & Conrad (2000), to build a corpus of 131 telephone interviews. Respondents in the full corpus were adults in New York City (61 females, 70 males) who were assigned to one of five experimental conditions (described below). Participants' mean age was 33.1 and they varied in ethnicity and education. In 4 of the 5 conditions, professional U.S. Census Bureau interviewers conducted the surveys. In the fifth condition, interviewers were graduate students given extra training to implement conversational interviews.

Respondents were brought into our laboratory and interviewers telephoned them from off-site, administering the interview via computer-assisted technology. The instrument was the Tobacco Supplement to the Current Population Survey. At the time when the data were collected, this supplement had been administered 1 or 2 times a year since 1992. Respondents answered 12–36 questions, depending on skip patterns, about their behaviors and attitudes.

2.1.1 Dependent Variables

The first of our two outcome measures was *Conceptual fit*, which we defined as the overlap between the respondent's conceptualization and the standard definitions from Suessbrick, et al. (2000). This was measured using a post-interview multiple-choice questionnaire. Each survey question contained anywhere from 1 to 7 concepts. For example:

Have you smoked at least 100 cigarettes in your entire life?

1. When you answered this question, did you interpret "smoking" to include:
   (a) Only puffs that you inhaled
   (b) Any puffs, whether or not you inhaled

2. How did you interpret "cigarettes"?
   (a) Cigarettes that you finished
   (b) Cigarettes that you partially smoked
   (c) Cigarettes that you only took a puff or two from

3. Did you interpret "cigarettes" to include:
   (a) Manufactured cigarettes
   (b) Hand-rolled cigarettes
   (c) Marijuana cigarettes
   (d) Cigars
   (e) Clove cigarettes
   (f) Something else. Specify: _____

By this measure 54% of respondents interpreted smoking in a way that matched the standard definition, which included <u>any puffs, whether or not you inhaled</u>.

The second of our measures was *Reliability of answers.* This was determined by the stability of respondents' answers when definitions were provided in a re-interview. More reliable answers (less response change) indicated that respondents and interviewers were more aligned during the initial telephone interview (that is, answers were accurate during the initial interview). Reliability of answers was measured with a re-interview, which was self-administered and contained complete definitions of the survey question concepts. Respondents were instructed to consider the official definitions when answering the question during this second administration. Following is an example from the self-administered re-interview:

Have you smoked at least 100 cigarettes in your entire life?

<u>Definition</u>
- We want you to include any puffs on any cigarettes, whether or not you inhaled AND whether or not you finished them.
- We want you to include hand-rolled cigarettes as well as manufactured ones, and tobacco cigarettes with additives like cloves.
- We DON'T want you to include cigars or non-tobacco cigarettes, like marijuana cigarettes.

Keeping this definition in mind, how would you answer this question?
1. Yes **2. No**

2.1.2 Independent Variables

The survey interviews varied as to type of clarification available, ranging from a strictly standardized interview (no clarification available), to clarification on-demand (if respondent explicitly requests it), to the most conversational (in which interviewers and respondents were free to initiate clarification sequences at any time). In two conditions respondents participated in strictly standardized interviews; in one case they were given a standard definition in the conceptualization questionnaire and in the other they were not. Although originally we had intended there to be only one clarification on- demand condition, one set of conversational interviews turned out to be effectively a clarification on- demand condition because clarification was almost never volunteered. So conditions 3 and 4 were both essentially clarification on-demand conditions. Condition 5 was intended to implement truly conversational interviews.

2.2 Results

2.2.1 Adherence to Interview Protocols

A manipulation check confirmed that novice interviewers conducting conversational interviews adhered to the appropriate protocol. Adherence was measured by analyzing the rate of interviewers' conversational interventions, such as mapping respondents' answers onto the available response alternatives, presenting a reduced set of response alternatives, and providing definitions. Interviewers provided some form of clarification or definition on 43% of all interview sequences, compared with 20% for the conversational condition using professional interviewers and 0–2% for the remaining three conditions, F (4, 122) = 61.07, $p < .0001$.

Additionally, with novice interviewers, respondents received clarification 56% of the time that they showed signs of needing clarification, compared with 38% of the time in Condition 4, and 10% of the time in Condition 3, F (4, 101) = 23.56, p < .0001. This suggests that conversational interviewing was indeed implemented well enough to be able to assess its effects in the current study.

2.2.2 Conceptual Fit

The results for our first dependent measure indicate that across all conditions, conceptual alignment between the respondent and interviewer was quite low. And as expected, more reliable answers had significantly better conceptual fit, F (1, 106) = 10.50, p = .002, with a rate of 37% for the answers that changed in the re-interview and 45% for those answers that did not (see Figure 1). In contrast to our expectations, however, there was no difference in rate of conceptual fit by condition, regardless of type of interview, F (4, 122) = 1.26, n,s

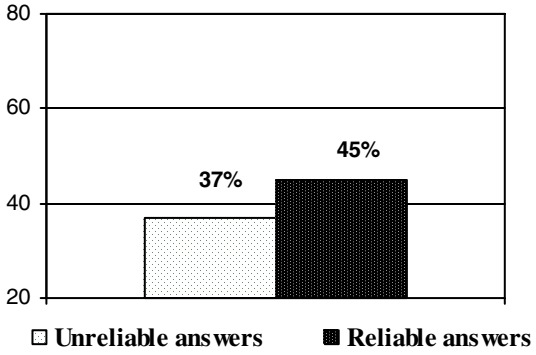

Figure 1: Rate of conceptual fit

2.2.3 Reliability of Answers

Respondents' answers were more reliable (that is, there was less response change upon re-interview), at 86%, than the rate of conceptual fit—which averaged 44%—would suggest. This is consistent with the notion that there are complex pathways to conceptual fit, and that sometimes respondents can answer reliably despite (non-critical) conceptual misalignment.

Further evidence that clarification doesn't always help is that answer reliability did not vary significantly among the different interviewing conditions. And it appeared that *how* clarification was provided —pre-emptively or in response to the respondent's cue—also had no effect on response accuracy.

Nonetheless, conceptual variability *can* lead to unreliable answers. About 11% of respondents' answers were unreliable (i.e., changed in the re-interview) when their conceptualizations were not aligned with the interviewers' and they received no clarification during the interview. This finding is surprisingly consistent with the rate of problematic answers in Conrad and Schober (2000), in which 11% of answers were unreliable when there was no clarification provided during the interview.

To emphasize this point, as in the subset of these data reported in Suessbrick et al. (2000), 10% of respondents in the current experiment changed their answer to the initial filter question, *Have you smoked at least 100 cigarettes in your entire life,* when presented with definitions in the re-interview. As a result, these respondents took the wrong route through the survey, answering some questions they should not have while not answering some questions they should have.

2.2.4 Coordination Pathways

We analyzed all the question-answer sequences in the corpus and found instances of all the logically possible coordination pathways (see Table 2):

	Number of sequences	Percent of sequences
Path 1	1435	75.0
Path 2	196	10.0
Path 3	32	1.7
Path 4	185	10.0
Path 5	46	2.4
Path 6	18	0.9

Table 2. Distribution of 1,916 interview sequences among the six coordination pathways from Table 1.

To make this more concrete consider the following examples of interactions from the various pathways.
Clarification helpful (Path #5): In the prototypical clarification sequence, clarification appeared to be helpful in resolving a respondent's probable misconception:

I: In the past year have you seen a medical doctor.
R: Regarding smoking? Or just in general.
I: Basically just in general. The reason you saw the doctor does not have to relate *to smoking.*
R: *okay* yes.

A review of this respondent's conceptualization questionnaire indicates that her concept interpretation matched the survey definition on the relevant dimension by the end of this question-answer sequence. Further, this respondent's answer did not change in the re-interview, suggesting that her answer was reliable.

Clarification unnecessary (Path #2): In some sequences clarification appeared to be unhelpful or unnecessary, as in this example:

I: In hospitals, do you think that smoking should be allowed in all areas, allowed in some areas, or not allowed at all.
R: Oh. Definitely: not allowed at all.
I: OK, and I'm just gonna clarify the definition. We want to make sure you consider all public areas like the waiting rooms, the cafeterias, and patient rooms, as well.
R: Yea. No.

Although we cannot know for certain whether the respondent's and interviewer's concepts of "hospital" were aligned at the beginning of this sequence, judging from the interaction it is likely to have been aligned. Not surprisingly, the respondent's answer did not change in the re-interview.

Clarification harmful (Path #3): In some sequences clarification seemed to sabotage the respondent's conceptualization and answer reliability, as in this example:

I: Have you smoked at least a hundred cigarettes in your entire life?
R: Um: in my entire life, no, not really.
I: Okay. And we want you to include any puffs on any cigarettes whether or not you inhaled and whether or not you finished them.
R: Okay.
I: And so have you smoked at least 100 cigarettes in your entire life?
R: Um: I would say yea! If you include that, yea.

Upon receiving a definition in the re-interview, this respondent changed her answer back to "no," suggesting that she and the interviewer were in all likelihood aligned in their concept interpretations before the clarification sequence but not after it.

But how frequently *is* clarification helpful? Our corpus provides one estimate. In interview sequences in which clarification occurred, respondents' conceptualizations ended up aligned with the survey definitions and their answers were found to be reliable (i.e., did not change in the re-interview) 80% of the time, as shown in Table 3.

R indicates need for clarification	*Conceptual alignment after interview*	*Effect of clarification*	*Percent of sequences (w/clarification)*
No	Aligned	?	56.1
Yes	Aligned	HELPFUL	23.6
No	Misaligned	NOT USEFUL, MAYBE HARMFUL?	10.2
Yes	Misaligned	NOT USEFUL	10.1

Table 3: Estimated effect of clarification on conceptual alignment in sequences in which clarification was given. *R = Respondent*

Looking more closely (second row of table), clarification was provided in response to a perceptible cue by the respondent, leading to improved conceptual fit and an accurate response, approximately 24% of the time. We can't know for sure what happened in the other cases (first row of table): presumably at least some of the time participants' ingoing conceptualizations were misaligned and clarification, although unprompted, improved alignment. About 20% of the time, however, the clarification was either not useful or may even have led respondents astray (third and fourth rows of table).

3. Discussion

This experiment supports findings in earlier research that conceptual misalignment between respondents and survey designers is far from rare. In the current study, respondents' conceptual fit averaged 44%, and their answers proved unreliable 14% of the time. More generally, believing that one has understood one's interlocutor can actually mask undetected conceptual misalignment (see also Schober, 2005). But the glass is also half full; sometimes misalignments are not critical to accurate responding, as evidenced by the 86% reliable answers.

Conceptual alignment does not appear to be an either/or phenomenon. Partial alignment is sufficient for successful communication if speakers' conceptualizations are aligned on the dimensions that matter. Alternatively, partial alignment may result in misunderstandings if speakers' conceptualizations are mis-

aligned on the important dimensions. In our experience, if a respondent interprets hospital to include both indoor (aligned) and outdoor (misaligned) areas and decides that smoking shouldn't be allowed at all in hospitals, the official definition probably wouldn't change her answer. However, if she decides that smoking should be allowed in some areas, and the outdoor area is the dimension that matters, the official definition probably would change her answer.

The data also support our proposal that there are multiple pathways to conceptual alignment (or misalignment)—not just the prototypical clarification sequence. In our current study, when respondents received clarification it helped or didn't harm conceptual fit and response accuracy 80% of the time. However, respondents' concept interpretations were still misaligned with the official definitions by the end of the question-answer sequence 20% of the time. Clarification was sometimes helpful, sometimes not useful, and occasionally harmful.

So when *do* respondent misconceptions lead to response error? Our results suggest that this occurs when the respondent and interviewer go down particular pathways: 1) When participants' concept interpretations are misaligned and the respondent does not receive clarification; and 2) When the clarification that the respondent receives is not relevant to her particular misconception (see also Lind et al., 2001, for a similar finding for web surveys). This suggests that simply empowering interviewers to clarify question concepts will not resolve all misalignments. But leaving interpretation entirely up to respondents is clearly also problematic, given the demonstrated frequency of conceptual misalignment and the risk it poses for answer reliability.

If clarification alone does not guarantee reliable answers, but rather must be *helpful* to the respondent in resolving her particular misconception, how and when should clarification be given? How should the interviewer diagnose what the respondent's particular misconception might be? It is possible that subtler displays of over- and under-certainty can improve judgment about when clarification is needed (e.g., Ehlen et al., 2005; Schober & Bloom, 2005). But the trick will be understanding what components of definitions are needed during that clarification, and making sure that clarification helps and does not hurt.

Acknowledgments

We thank Patrick Ehlen, Rene Holl, Alyssa Monnie, Maile O'Hara, and Gina Turner from the Psycholinguistics Laboratory at the New School for Social Research. This material is based upon work supported by the National Science Foundation under grant No. ITR-0081550.

References

Belson, W.A. (1981). *The design and understanding of survey questions*. Aldershot: Gower.

Conrad, F.G., & Schober, M.F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly, 64,* 1–28.

Conrad, F.G., Schober, M.F., & Dijkstra, W. (2004). Nonverbal cues of respondents' need for clarification in survey interviews. In *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.

Ehlen, P., Schober, M.F., & Conrad, F.G. (2005). Modeling speech disfluency to predict conceptual misalignment in speech survey interfaces. *Proceedings of the Symposium on Dialogue Modelling and Generation*, 15th Annual meeting of the Society for Text & Discourse, Vrije Universiteit, Amsterdam.

Fowler, F.J., & Mangione, T.W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error.* Newbury Park, CA: SAGE Publications.

Lind, L.H., Schober, M.F., & Conrad, F.G. (2001). Clarifying question meaning in a web-based survey. In *Proceedings of the American Statistical Association, Section on Survey Research Methods.* Alexandria, VA: American Statistical Association.

Schober, M.F. (2005). Conceptual alignment in conversation. In B.F. Malle & S.D. Hodges (Eds.), (pp. 239–252). New York: Guilford Press.

Schober, M.F., & Bloom, J.E. (2004). Discourse cues that respondents have misunderstood survey questions. *Discourse Processes, 38*, 287–308.

Schober, M.F., & Conrad, F.G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61,* 576–602.

Schober, M.F., Conrad, F.G., & Fricker, S.S. (2004). Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology, 18*, 169–188.

Suessbrick, A., Schober, M.F., & Conrad, F.G. (2000). Different respondents interpret ordinary questions quite differently. *Proceedings of the American Statistical Association, Section on Survey Research Methods.* Alexandria, VA: ASA.

Tourangeau, R., Conrad, F., Arens, Z., Fricker, S., Lee, S., & Smith, E. (in press). Everyday concepts and classification errors: Judgments of disability and residence. *Journal of Official Statistics.*