SURVEYS · QUESTIONNAIRE DESIGN

The Psychology of Answering a Survey Question: Comprehension, Retrieval, Judgement and Response

Answering a survey question takes four mental steps - understanding the question, recalling information, forming a judgement and reporting an answer. This article explains each step, the errors that arise at each one and the design choices that reduce them.

ModQu research article · 2,571 words · 24 references (APA 7th edition)

A survey question takes a few seconds to answer, and it is easy to assume that respondents simply read off a fact or an opinion stored in their heads. Research in psychology and survey methodology over the past four decades shows that the process is more complicated. To answer even a simple question, a respondent must understand what is being asked, search memory for relevant information, turn that information into a judgement and then express the judgement in the form the questionnaire requires. Each step can go wrong, and each step is influenced by how the question is written.

This way of thinking about survey answers grew out of a 1983 seminar that brought cognitive psychologists and survey researchers together, a movement now known as the cognitive aspects of survey methodology. One of its products was a four-stage model of the response process proposed by Roger Tourangeau (Tourangeau, 1984) and later developed at length (Tourangeau et al., 2000; Sudman et al., 1996). The model has become the standard framework for diagnosing problems with survey questions. This article explains the four stages, the errors associated with each and what questionnaire designers can do about them.

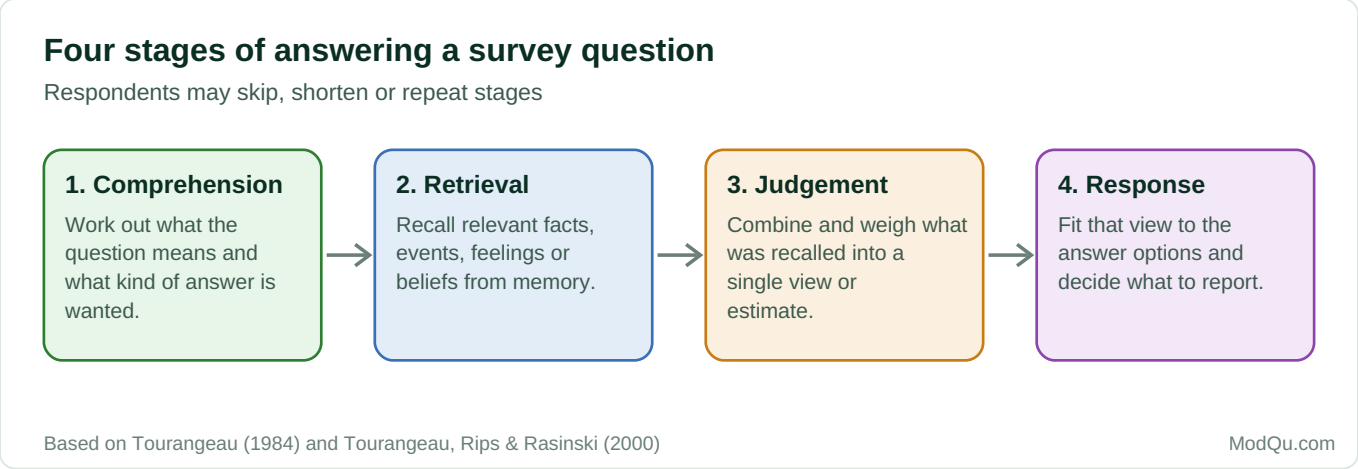


Figure 1. The four stages of the survey response process. Based on Tourangeau (1984) and Tourangeau, Rips and Rasinski (2000).

The stages are not always carried out in strict order, and respondents do not always complete all of them. Someone asked their age does not need to search memory or weigh evidence. Someone asked about their satisfaction with a public service may go back and forth between recalling experiences and forming a view. The model is best understood as a list of tasks that answering requires, not a fixed sequence.

Stage 1: Comprehension

The first task is to understand the question. This involves two kinds of understanding. The first is literal: what do the words mean? The second is pragmatic: what does the questioner want to know, and what kind of answer is expected?

Literal meaning

Many everyday words are less clear than they seem. In a series of studies in Britain, William Belson tested how respondents interpreted common survey questions and found that a large share of answers were based on interpretations that differed from the one intended. Words such as "you" (the respondent alone, or the household?), "weekday" and "regularly" were understood in different ways by different people (Belson, 1981). A question about how many people live in a household seems straightforward until the respondent thinks about a student child living away, a lodger or a partner who stays some nights.

Michael Schober and Frederick Conrad studied what happens when respondents' situations do not fit the question neatly. They asked respondents factual questions about scenarios, such as whether a particular purchase counted as buying household furniture, where the correct answer depended on official definitions. When the scenario was straightforward, respondents answered accurately whatever the interviewing style. When the mapping was complicated, accuracy was poor under strictly standardised interviewing and much higher when interviewers were allowed to explain the definitions (Schober & Conrad, 1997). A field study in a household telephone survey produced similar results (Conrad & Schober, 2000). Respondents often do not realise that their understanding differs from the researcher's, so they do not ask.

Pragmatic meaning

Respondents also interpret a question by assuming that the questioner is following the normal rules of conversation: that the question is relevant, that it asks for new information and that everything in it is there for a reason (Grice, 1975). Survey respondents apply these rules to every part of the questionnaire, including the response options, the examples, the title and the questions that came before (Schwarz, 1999).

This explains many apparently strange findings. When respondents were asked how many hours a day they watched television using a scale whose categories ran from "up to half an hour" to "more than two and a half hours," fewer reported heavy viewing than when the categories ran from "up to two and a half hours" to "more than four and a half hours" (Schwarz et al., 1985). Respondents took the middle of the scale as a signal of what typical viewing looked like. The scale was meant as a recording device; respondents read it as information.

Conversational rules also explain some question-order effects. When people are asked about their satisfaction with their marriage and then about their satisfaction with life as a whole, some exclude their marriage from the second answer, on the reasoning that they have already said how they feel about it. The effect changes when the two questions are introduced together as being about two separate topics, which signals that the second should be answered without repeating the first (Schwarz et al., 1991).

Stage 2: Retrieval

Once a respondent understands what is being asked, they must recall the information needed to answer. For factual questions, this means retrieving memories of events or behaviours. For attitude questions, it means bringing to mind relevant beliefs, feelings and experiences.

Remembering events

Memory for everyday events is limited and selective. Events that are recent, distinctive, emotionally important or repeated often in a consistent way are easier to recall. Routine events blur together. Forgotten events lead to under-reporting, especially over long reference periods.

A second error is **telescoping**, in which events are remembered as having happened more recently than they did, so they are wrongly included in the reference period. A classic study of household spending on home repairs found that respondents reported substantially more expenditure for the most recent month when there was no earlier interview to mark the start of the period, because jobs done before the period were being pulled forward into it (Neter & Waksberg, 1964). Bounding interviews, in which a previous interview marks the start of the period, reduce this error.

Memory can be helped. Shorter reference periods, cues that remind respondents of different kinds of events, and personal landmarks such as holidays or a change of job all improve recall. Event history calendars, which help respondents place events in relation to each other on a timeline, can improve the quality of retrospective reports (Belli, 1998). Asking respondents to consult records, such as bank statements, also helps when the information is available.

Retrieving attitudes

Recalling an attitude is also less straightforward than it might appear. For topics people care about and have thought about, a stable opinion may come to mind quickly. For many survey topics, however, respondents do not have a ready-made answer. Instead, they sample whatever considerations happen to be accessible at the moment and form an opinion on the spot (Zaller & Feldman, 1992). What is accessible depends partly on the questions that came before, which is why question order can change answers (Tourangeau & Rasinski, 1988).

Stage 3: Judgement

At the judgement stage, respondents combine what they have recalled into an answer. For factual questions, this often involves estimation. For attitude questions, it involves weighing considerations.

Counting versus estimating

Asked how many times they visited a doctor in the past year, a respondent might try to recall and count each visit. For a small number of distinctive events, this works. For frequent events, people more often estimate, for example by thinking "about once a month" and multiplying (Blair & Burton, 1987). Estimation is not necessarily worse than counting, since counting fails when events are forgotten, but it produces characteristic errors such as rounding to convenient numbers and answers heavily influenced by the response options. Bradburn and colleagues showed that many answers to autobiographical questions are based on inference and estimation rather than direct recall, and that respondents fill gaps in memory with what usually happens (Bradburn et al., 1987).

Context and comparison

Judgements also depend on standards of comparison. An earlier question can provide a standard: rating a very good restaurant first can make the next one seem worse by contrast. An earlier question can also bring certain considerations to mind, so that they are included in a later judgement and push it in the same direction. Which happens depends on whether respondents see the information as part of the target or as a separate comparison point (Strack & Martin, 1987; Tourangeau & Rasinski, 1988). These effects are the basis of most question-order effects.

Stage 4: Response

In the final stage, the respondent reports an answer. This requires mapping the judgement onto the response format and deciding whether to report it honestly.

Mapping

A respondent who has a clear view must still decide which point on a scale represents it. Is "quite satisfied" a 4 or a 5 out of 7? Labels on every point help, and so do scales with a clear midpoint and balanced options. When options are presented in a list, their order can matter: in self-administered surveys, options near the top tend to be chosen more often, while in spoken interviews, options heard last are favoured (Krosnick & Alwin, 1987).

Editing

Respondents may also edit their answers before reporting them. Questions about sensitive topics, such as drug use, voting, income or charitable giving, invite answers that present the respondent in a favourable light. This tendency, called **social desirability bias**, leads to under-reporting of disapproved behaviour and over-reporting of approved behaviour (Tourangeau & Yan, 2007). It is stronger when an interviewer is present. An experiment with university graduates, whose records could be checked, found that more respondents reported undesirable facts, such as failing a class, in a web survey than in a telephone interview (Kreuter et al., 2008).

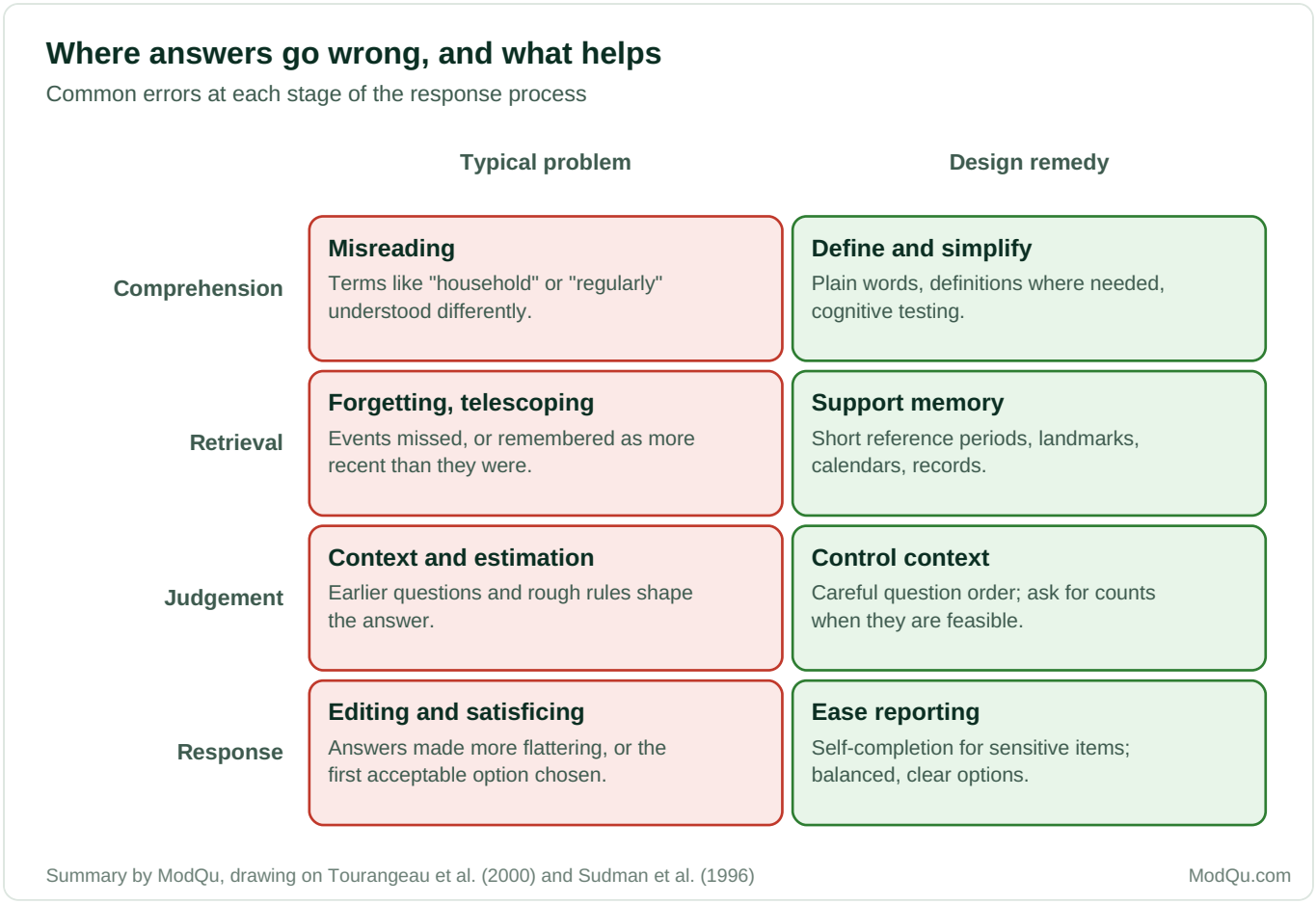


Figure 2. Typical errors at each stage of the response process and design remedies. Summary by ModQu, drawing on Tourangeau et al. (2000) and Sudman et al. (1996).

When respondents take shortcuts

The four-stage model describes what a careful respondent does. Not all respondents are careful all the time. Jon Krosnick argued that respondents often **satisfice**: rather than performing each stage thoroughly, they do just enough to give an answer that seems acceptable (Krosnick, 1991). Weak satisficing means doing each stage less carefully, for example choosing the first reasonable option. Strong satisficing means skipping stages altogether, for example saying "don't know," agreeing with every statement or choosing the same scale point for every item in a grid.

Satisficing is more likely when the task is difficult, when respondents have limited ability or motivation and when they are fatigued, for example late in a long questionnaire. It is not a moral failing but a predictable response to burden, and it can be reduced by making questions easier, keeping questionnaires short and explaining why the survey matters.

Early work at the University of Michigan made a related point about motivation. Charles Cannell and colleagues found that respondents gave more complete reports of health events when interviewers gave clear instructions, asked respondents to commit to answering carefully and gave feedback on good performance (Cannell et al., 1981). Respondents can do the cognitive work; they need a reason to do it.

Mode, interviewers and the response process

The way a question is delivered changes the demands at each stage. In a telephone interview, respondents hear the question once, at the interviewer's pace, and must hold the response options in memory while they decide. Long lists and scales with many points are therefore harder by telephone, and respondents rely more on the options they heard last (Krosnick & Alwin, 1987). On paper or on screen, respondents can reread the question and see all the options at once, but they also respond to visual features such as spacing, layout and the position of an option on the page.

Interviewers affect more than social desirability. A skilled interviewer can notice hesitation, repeat a question and encourage a respondent to take time over recall, all of which help the retrieval and judgement stages (Cannell et al., 1981). At the same time, strictly standardised interviewing, in which interviewers read questions exactly as written and give only neutral probes, prevents them from clarifying misunderstandings at the comprehension stage. The research on conversational interviewing suggests that allowing interviewers to explain official definitions improves accuracy for complicated cases, though it lengthens interviews (Schober & Conrad, 1997).

Self-administered web surveys remove the interviewer altogether. This reduces editing on sensitive questions but leaves respondents with no one to ask when a question is unclear, and with no one to keep them motivated. Definitions, examples and help text placed close to the question can partly replace the interviewer's role, but only if respondents read them.

Differences between respondents

The same question can impose very different burdens on different people. Respondents with less formal education, those answering in a second language and older respondents with weaker working memory may find long or abstract questions harder to process, which makes satisficing more likely among these groups (Krosnick, 1991). Respondents who know a topic well can retrieve and judge quickly; those for whom the topic is unfamiliar must construct an answer from scratch. This matters because differences in answers between groups may partly reflect differences in how hard the question was for each group, rather than real differences in what they think or do. Testing questions with a varied set of respondents is the best protection.

Putting the model to work

The main value of the four-stage model is diagnostic. When a question produces strange results, the model suggests where to look. Is the problem that respondents understand the question differently? That they cannot remember? That they are estimating rather than counting? That they are reluctant to give the true answer? Each diagnosis points to a different remedy, as Figure 2 shows.

The model is also the basis of **cognitive interviewing**, the most widely used method for pretesting questions. In a cognitive interview, a small number of people from the target population answer draft questions while thinking aloud or responding to follow-up probes such as "What does the term 'household' mean to you?" or "How did you arrive at that number?" The probes are designed to reveal problems at each stage of the response process (Willis, 2005; Presser et al., 2004).

For businesses that run surveys, the practical lessons are straightforward:

- **Test comprehension.** Words that are familiar inside a company, such as product names, internal categories or industry terms, are often unclear to customers. A few cognitive interviews will show where.
- **Ask about periods people can remember.** "In the past seven days" yields more accurate reports of routine behaviour than "in the past year." If annual figures are needed, it is often better to ask about a short period and scale up.
- **Respect the order of questions.** General evaluations, such as overall satisfaction, should usually come before specific ones, so that the specific questions do not narrow the general one.
- **Make reporting easy.** Label scale points, keep options balanced and allow respondents to complete sensitive questions privately.
- **Reduce burden.** Long, repetitive questionnaires invite satisficing. Every unnecessary question lowers the quality of the necessary ones.

Conclusion

A survey answer is the end product of a sequence of mental tasks: understanding the question, recalling relevant information, forming a judgement and reporting it in the required format. Errors can arise at each stage, from misunderstood words and forgotten events to estimation shortcuts, context effects and flattering edits. Some respondents also shorten the process when the task is demanding. The four-stage model gives questionnaire designers a way to anticipate these problems and to choose wording, formats and question orders that make accurate answering easier. It does not make survey measurement perfect, but it turns many apparently random errors into predictable ones, and predictable errors can be reduced.

References

Belli, R. F. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory, 6*(4), 383–406. https://doi.org/10.1080/741942610

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

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research, 14*(2), 280–288. https://doi.org/10.1086/209112

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science, 236*(4798), 157–161. https://doi.org/10.1126/science.3563494

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology, 12*, 389–437. https://doi.org/10.2307/270748

Conrad, F. G., & Schober, M. F. (2000). Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly, 64*(1), 1–28. https://doi.org/10.1086/316757

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). Academic Press.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly, 72*(5), 847–865. https://doi.org/10.1093/poq/nfn063

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*(2), 201–219. https://doi.org/10.1086/269029

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Neter, J., & Waksberg, J. (1964). A study of response errors in expenditures data from household interviews. *Journal of the American Statistical Association, 59*(305), 18–55. https://doi.org/10.1080/01621459.1964.10480699

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly, 68*(1), 109–130. https://doi.org/10.1093/poq/nfh008

Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly, 61*(4), 576–602. https://doi.org/10.1086/297818

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*(2), 93–105. https://doi.org/10.1037/0003-066X.54.2.93

Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly, 49*(3), 388–395. https://doi.org/10.1086/268936

Schwarz, N., Strack, F., & Mai, H.-P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly, 55*(1), 3–23. https://doi.org/10.1086/269239

Strack, F., & Martin, L. L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In H.-J. Hippler, N. Schwarz, & S. Sudman (Eds.), *Social information processing and survey methodology* (pp. 123–148). Springer.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. Jossey-Bass.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103*(3), 299–314. https://doi.org/10.1037/0033-2909.103.3.299

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Tourangeau, R. (1984). Cognitive science and survey methods. In T. B. Jabine, M. L. Straf, J. M. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). National Academy Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. SAGE. https://doi.org/10.4135/9781412983655

Zaller, J., & Feldman, S. (1992). A simple theory of the survey response: Answering questions versus revealing preferences. *American Journal of Political Science, 36*(3), 579–616. https://doi.org/10.2307/2111583