

Answering Questions in Surveys

Catherine Sears and Daniel M. Ennis

Background: Surveys are conducted to study three main categories: attitudes and beliefs, events and behaviors, and subjective experience. Product, concept, and brand testing can be viewed as types of surveys falling into the third category, subjective experience. This view encourages cross-disciplinary awareness of the survey and polling literature, as scientists in these fields grapple with many of the same problems that occur in sensory and consumer science. One idea that has developed in survey research, highlighted in a review of the main accomplishments of the field over the 20th century, is the idea of the satisficing-optimizing continuum arising from a process model of the way that people answer questions¹.

In this technical report, we consider the implications of these ideas in a scenario involving a comparison of answers to questions in a check-all-that-apply (CATA) format and an applicability scoring format².

Scenario: You are a consumer insights manager in a coffee company and are exploring the emotional responses of consumers to coffee drinking in order to develop an emotion lexicon. With a large list of emotion terms your typical approach would be to use a check-all-that-apply (CATA) list as shown in Figure 1. CATA lists are commonly used in survey research and have been increasing in popularity within sensory science in recent years².

Read the phrases and, for each phrase, mark the box if the phrase describes how drinking coffee makes you feel.

- I feel energized
- I feel guilty
- I feel elated

Figure 1. Three items from a set of one hundred in a CATA survey.

Yet as you review the results, you notice that respondents are tending to check mainly the earlier presented items and ignoring items toward the end of the survey. You also wonder what it means when a box is unchecked. Does the respondent mean to indicate that the item truly does not apply? Or was the item skipped unintentionally?

Answering Questions: When a respondent answers a question in a consumer survey, we would like to believe that all of the steps in Figure 2 are faithfully executed. Ideally, a respondent would first comprehend, interpreting the question and deducing its intent by settling on the meaning of each word and establishing relations among the concepts evoked by the words. Next, we assume that the respondent engages in retrieval, searching their memories for relevant information. The respondent then integrates this retrieved information into a judgment and finally makes carefully

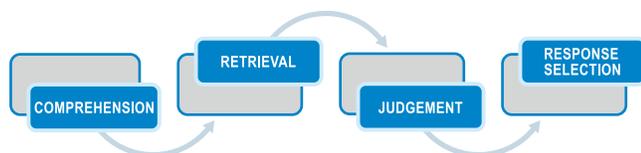


Figure 2. A process model for answering a survey question.

considered response selections, mapping their judgment onto response choices available to them, actively looking for a best fit^{3,4}.

Warwick *et al.*⁵ and Krosnick¹ have theorized that this model of survey response, *optimizing*, is what happens when respondents are internally satisfied by successful performance, believe survey results will help employers improve working conditions, want an intellectual challenge, feel altruistic, or are looking for emotional catharsis. Yet most people are unlikely to faithfully execute each of these four steps in response to each and every question and may expend less energy by *satisficing*. In fact, energy minimization may be a motivating factor in how people elect to answer questions. In considering the behavior of different subjects and even the same subject in a given test, we may observe a continuum from strong satisficing to optimizing as shown in Figure 3 corresponding to the degree to which the steps in Figure 2 are executed properly¹.



Figure 3. The satisficing – optimizing continuum.

A respondent engaged in *weak* satisficing executes all four steps in Figure 2 – comprehension, retrieval, judgment and response selection – but does so less thoroughly than a respondent who is optimizing. Instead of generating the most accurate answers, respondents settle for merely satisfactory ones. Respondents who are engaged in *strong* satisficing skip the retrieval and judgment steps altogether. They would interpret each question superficially and select what seems to be a reasonable answer without referring to any internal psychological cues relevant to the question¹. Teachers of unmotivated students know this behavior quite well.

CATA and Satisficing: Smyth *et al.*⁶ found that CATA lists in written format have a tendency to induce satisficing behavior in which respondents limit their processing effort by quickly selecting the first reasonable response in a CATA list. The respondents' weak satisficing behavior causes primacy effects, a disproportionate selection of items appearing early in the list of options. In Smyth *et al.*'s research, for CATA respondents who spent the mean response time or less, eight of ten questions were significantly more likely to be endorsed when they appeared in the first three positions in the list than when they appeared in the last three positions. These patterns of primacy were only shown by those who spent less time, suggesting that these patterns were from a lack of cognitive processing⁶. Interestingly, the satisficing-optimizing model predicts recency effects in oral interviews and these effects have been observed in practice. Ares *et al.*⁷ applied Smyth *et al.*'s work to sensory categories and showed that when consumers complete CATA questions for sensory characterization repeatedly, visual processing of

information is reduced and respondents may therefore leave items blank, even when relevant. A blank option in CATA does not necessarily mean “does not apply.” Respondents may leave an option blank for a number of reasons including that the option doesn’t apply, they are neutral or undecided, or they overlooked it.

Applicability Scoring: A related technique to CATA is *applicability scoring*, which requires respondents to mark what is both applicable and also *not* applicable². (This is sometimes called “forced-choice CATA,” a phrase avoided here to prevent confusion with forced-choice sensory testing methods such as *m-AFC*.) Access to both responses is needed to conduct statistical analyses to compare products or concepts using McNemar’s test. CATA, on the other hand, only requires a check when the item applies to the object being scored and is therefore not amenable to this type of analysis. Applicability scoring² may also lead respondents to more deeply process items and to score more options⁷. Respondents take longer to answer in the applicability scoring format than in the CATA format, perhaps because they need to commit to an answer for every item, and therefore are more likely to think of reasons that the options do or do not apply⁶.

Applicability scoring carries with it a risk of acquiescence bias, in which respondents who are truly neutral check the “does apply” option. Respondents may acquiesce for a number of reasons, including social desirability, or because they are following the rules of ordinary conversation, in which they feel they must contribute something in response to a question⁸. However, by including a test neutral category, Smyth *et al.*⁶ did not find strong evidence for acquiescence bias in applicability scoring.

Read the phrases and, for each phrase, mark the box on the left if the phrase describes how drinking coffee makes you feel. Mark the box in the middle if the phrase does not describe how drinking coffee makes you feel. Mark the box on the right if you do not have an opinion.

	Does Apply	Does Not Apply	Not Sure/ Cannot Decide
I feel energized	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I feel guilty	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
I feel elated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure 4. Three items from a set of one hundred in an applicability survey.

Applicability Scoring and the Emotion Survey: You conduct an applicability scoring experiment with your respondents using the same emotion terms used in the CATA experiment. Despite Smyth *et al.*’s demonstration that a neutral category does not draw out “yeses” in applicability scoring, you include a “Not Sure/ Cannot Decide” option, as shown in Figure 4, to ensure that respondents are not selecting “yes” because they are undecided⁹. The applicability scoring survey provides different results than the CATA methodology. Respondents take significantly longer to

answer, and, as shown in Figure 5 according to the scenario, the primacy effect has decreased, theoretically the result of reduced satisficing behavior. Figure 5 dramatizes the proportion of checked boxes in CATA or “yes” responses in applicability scoring that one would expect to see; all questions have been randomly rotated. You now have greater justification to conclude that a blank checkbox means that the respondent ignored the question.

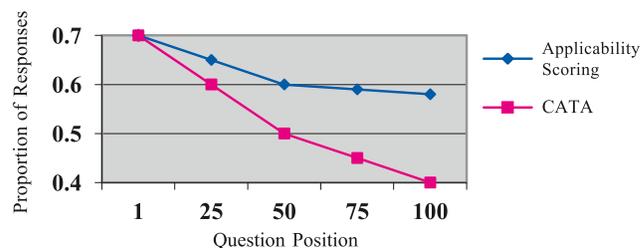


Figure 5. Primacy effects: Applicability vs. CATA scoring.

Conclusion: When the results of a poll or survey are reported, ideal conditions for respondent motivation are often not considered or they may even be assumed. Yet, respondents sometimes operate less than ideally and may be displaying satisficing behavior. While surveys, including product and concept tests, usually ensure against bias in the realms of the survey design, we must also consider respondent behavior. If respondents are “quick-clicking” without thoroughly executing the steps in Figure 2, the data is questionable at best, regardless of how representative those respondents are of the relevant population, or how carefully the questions are rotated. Avoiding the use of CATA for extensive surveys and considering applicability scoring instead is one potential way to discourage satisficing in favor of optimizing and thus reduce primacy effects.

References

- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, **50**(1), 537-567.
- Ennis, D. M. and Ennis, J. M. (2013). Analysis and Thurstonian scaling of applicability scores. *Journal of Sensory Studies*, **28**(3), 188-193.
- Cannell, C., Miller, P., and Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, **12**, 389-437.
- Tourangeau, R. and Rasinski, K.A. (1988). Cognitive process underlying context effects in attitude measurement. *Psychological Bulletin*, **103**, 299-314.
- Warwick, D. and Lininger, C. (1975). *The sample survey: Theory and practice*. New York: McGraw Hill.
- Smyth, J., Dillman, D., Christian, L., and Stern, M. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, **70**(1), 66-77.
- Ares, G., Etchemendy, E., Antúnez, L., Vidal, L., Giménez, A., and Jaeger, S. (2014). Visual attention by consumers to check-all-that-apply questions: Insights to support methodological development. *Food Quality and Preference*, **32**, 210-220.
- Grice, P. (1975). Logic and conversation. In P. Cole and J. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.
- Jaeger, S., Cadena, R., Torres-Moreno, M., Antúnez, L., Vidal, L., Giménez, A., Hunter, D., Beresford, M., Kam, K., Yin, D., Paisley A., Chheang, S., and Ares, G. (2014). Comparison of check-all-that-apply and forced-choice Yes/No question formats for sensory characterization. *Food Quality and Preference*, **35**, 32-40.