methods, data, analyses | 2025, pp. 1-45 DOI: 10.12758/mda.2025.02

Probing in Cognitive Interviews Can Promote Acquiescence

Frederick G. Conrad[1], Rachel E. Davis[2], Carolyn Lau[3], Melissa Armendáriz[1], Stephanie Morales[1], Timothy P. Johnson[4], Johnny Blair[5]

[1] *University of Michigan*
[2] *University of South Carolina*
[3] *Pew Research*
[4] *University of Illinois at Chicago*
[5] *Independent consultant*

Abstract

Cognitive interviewing is widely used to pretest survey questionnaires and is considered a best practice (e.g., Willis, 2005, 2018; Beatty & Willis, 2007). However, the method has been controversial because, among other concerns, it requires interviewers to probe respondents for more detail or clarity about their experience answering draft survey questions which may lead them to report "problems" they have not actually experienced (e.g., Conrad & Blair, 2009). The present study investigates this possibility from the perspective of Acquiescent Response Style (ARS) – the tendency for survey respondents to select positive responses such as "yes" or "strongly agree," irrespective of the question's content (e.g., Baumgartner & Steenkamp, 2001). For example, respondents in a cognitive interview might affirm experiencing a problem mentioned in or implied by an interviewer's probe even if they have not actually experienced it. We embedded a probing experiment in a pretest of a health survey in which respondents participated in cognitive interviews that used either *directive* probes (n=41) or *non-directive* probes (n=26). Directive probes explicitly queried respondents about a specific, intentionally unlikely interpretation of each question in a draft questionnaire; non-directive probes were open-ended. Directive probe (DP) respondents affirmed the interpretation queried in the probes over five times more often than respondents in the non-directive probe (NP) group volunteered these interpretations. This pattern was reversed for interpretations of the questions that were volunteered, i.e., about which DP respondents were not asked: NP respondents volunteered alternative interpretations over four times more than DP respondents. These effects were particularly pronounced for respondents with lower levels of education and who were younger. The findings suggest that directive probing in cognitive interviewing can promote responding that is reminiscent of ARS – an *affirmation bias* – and likely harmful for the quality of evidence produced in cognitive interviews.

Keywords: cognitive interviews, directive probes, acquiescent response style, affirmation bias, acquiescence, verbal probes, satisficing

Acknowledgements

The research reported here was supported, in part, by National Cancer Institute Grant # 1 R01 CA172283-01A1. We thank Francesca Terzoli and Torie Harmon for coding transcribed interviews. The article is dedicated to the memory of our colleague and co-author, Johnny Blair, whose commitment to improving how research is done inspired us.

Direct correspondence to

Frederick G. Conrad, University of Michigan, Ann Arbor, United States,
E-mail: fconrad@umich.edu

Cognitive interviewing is a widely used technique for pretesting questionnaires and considered essential for the creation of high-quality survey data (e.g., Miller, 2011; Ridolfo et al., 2020; Willis, 2005; Willis, 2015; Willis, 2018; Wright et al., 2021). There are many versions of cognitive interviewing, but the general approach is to conduct in-depth, semi-exploratory interviews about a questionnaire that is under development, during which respondents are asked to answer each question in a way that makes their thinking explicit, usually by thinking aloud and/or responding to verbal probes such as a request to paraphrase the question. The respondents' think-aloud protocols and responses to probes are examined for evidence that the respondent has encountered problems[1] answering the survey question – problems that would likely add measurement error to the resulting survey data if the questionnaire were to be used in production interviewing without further revision. For example, the respondent's verbal reports may make it evident that answering a question requires recalling events that are hard to distinguish from other similar events or that the respondent has misinterpreted the question relative to what the author intended. By rewording questions to resolve the problems uncovered in this way, researchers can minimize the chances that these same problems will introduce measurement error to the responses once production interviewing has begun, especially if the process is iterative.

The think-aloud procedure requires respondents to verbalize what is going through their minds while answering, but think-aloud protocols may not, by themselves, be interpretable by the cognitive interviewer, or whoever analyzes them (e.g., Ericsson & Simon, 1992). Thus, it is common for the cognitive interviewer to ask additional, often unscripted questions – what Willis (2005) calls "verbal probing" – to gain clarity about what the respondent may have meant. Several authors (e.g., Beatty, 2004; Conrad & Blair, 2009; Priede et al., 2014; Willis, 2005) have distinguished between different types of cognitive interviewing probes. The distinction most relevant to the current study is between probes that specifically refer to a problem that at least some respondents are anticipated to experience, e.g., "Did you think the question was asking about your expenses for prescription medicine?" and probes that do not refer explicitly to a possible problem, e.g., "What kind of expenses do you think that question was asking about?". The concern is that asking respondents whether they have experienced a particular problem, irrespective of evidence to that effect, may promote false

1 We use the term "problem" to cover both a respondent not understanding a question, i.e., they are confused by what they have been asked, and misunderstanding a question, i.e., the respondent interprets the question differently than its authors intended. In the former, they likely recognize their confusion and that they might be unable to respond, i.e., that they have encountered a problem. In the latter, they are likely unaware of the misalignment and thus unlikely to consider it a problem. For consistency with other research on cognitive interviewing which describes the evidence that a question is not functioning as intended as "problems," we use that term here.

reports of the problem (Conrad & Blair, 2009), which can lead to unnecessary changes to the question, wasting resources and potentially introducing new problems through the revision process.

Acquiescent Response Style. Respondents in production surveys – not necessarily pretests of questionnaires – sometimes endorse positive answer categories such as "Strongly Agree" irrespective of item content (e.g., Baumgartner & Steenkamp, 2001; Krosnick, 1991). This tendency, known as Acquiescent Response Style (ARS), is typically (although not always) observed when questions include bipolar scales (for example, a scale from "Strongly Agree" to "Strongly Disagree") used in questions about opinions and other subjective phenomena, primarily in interviewer-administered questionnaires (see Davis et al., 2024). There are several reasons why ARS may arise. First, the respondent may feel that selecting positive answers is more polite than selecting negative answers and may believe that being polite can facilitate the interaction with the interviewer. Second, in some cases, respondents may wish to avoid the effort required to carefully think through their answers and may want to do this without being conspicuous; selecting a response option because it is positive but without regard to what it means may satisfy both goals. Reducing the effort in this way can be seen as an instance of the more general tendency for some respondents to take mental shortcuts, referred to as *survey satisficing* (e.g., Krosnick, 1991; Roberts, *et al*., 2019), which also includes phenomena such as non-differentiation, rounded numerical answers, and primacy effects.

ARS is well known to be more common among Latinos than non-Latino whites (e.g., Aday et al., 1980; Liu et al., 2017; Ross & Mirowsky, 1984), possibly reflecting cultural factors such as *simpatía*, a Latino cultural value that promotes being pleasant, agreeable, likable, non-confrontational, and respectful in interpersonal interactions (e.g., Triandis et al., 1984; Davis et al., 2011). If so, this would be consistent with the politeness explanation for the phenomenon. We note that the prevalence of ARS has been found to differ between Latino subgroups (Davis, et al., 2019), presumably reflecting the considerable cultural, economic, and political diversity within the Latino population in the US (e.g., Zong, 2022).

ARS is also more common among respondents with less education (Liu et al., 2019; McClendon, 1991; Meisenberg & Williams, 2008; Messick & Frederiksen, 1958). Educational attainment is sometimes used as a proxy for "cognitive ability" or "cognitive sophistication" in studies of satisficing (e.g., Krosnick & Alwin, 1986); if that relationship extends to ARS (often considered a type of satisficing), it would be consistent with the effort reduction explanation as the response task is likely experienced as more difficult by those with lower levels of ability.

Also a possible indication of effort reduction, older respondents tend to show higher levels of ARS (e.g., Liu et al., 2019), potentially reflecting reduced aptitude due to cognitive aging and, thus, an impetus to simplify their task.

Whatever the origin of ARS, it is almost certainly a type of measurement error (Baumgartner & Steenkamp, 2001; Billiet & Davidov, 2008; Billiet & McClendon, 2000; Cheung & Rensvold, 2000; Hoffman et al., 2013; Weijters et al., 2008; Winkler et al., 1982). If respondents endorse statements they do not actually agree with or agree with less strongly than their response would indicate, this can distort survey estimates.

This article explores the possibility that an ARS-like process may be at play in cognitive interviews for pretesting questionnaires, contributing measurement error to the conclusions, much as ARS can distort the estimates and conclusions based on the data collected in production interviews. Because cognitive interviewers often probe respondents for more detail about their thinking than may be evident in their spontaneous verbalizations, respondents may affirm specific problems queried by probes much as they endorse positive response options irrespective of their actual opinions producing acquiescent survey responses.

Current Study

The current study investigates whether cognitive interview probes can lead respondents to agree with an interpretation of a question embodied in a probe, irrespective of whether they actually hold this interpretation. More specifically, the study asks whether and how often respondents in cognitive interviews agree with an interpretation mentioned in a probe even if that interpretation is implausible and unlikely to be arrived at spontaneously. Respondents were randomly assigned to one of two types of cognitive interview, either those in which interviewers administered *directive probes,* the Directive Probe (DP) group, or cognitive interviews in which the interviewer administered *non-directive probes,* the Non-directive Probe (NP) group. In the DP group, the questionnaire contained scripted probes, which asked the respondent if they interpreted the question in a specific – and unlikely – way; in the NP group, the scripted probes were open ended, asking respondents how they interpreted the question (see Table 1).

The critical aspect of directive probes as we define them here is that they ask the respondent to confirm or deny having experienced a specific problem, i.e., they are in effect Yes/No questions. In the current study the directive probes were scripted ahead of time, however cognitive interviewers as experts may – and, in our experience do – sometimes spontaneously ask the respondent to confirm that they have experienced a problem. We designed the question interpretations about which directive probes were administered to be highly implausible so that it would be unlikely for respondents to come up with these interpretations left to their own devices.

In contrast, the probes in the NP group did not mention any specific interpretations. For example, in Table 1 the NP probe asks, "who were you picturing doing this judging" (and if the respondent did not answer, they were provided an exhaustive set of options). This contrasts with the directive probe, which explicitly asks the respondent if they interpreted the judging to have been done by "strangers," chosen because in the authors' judgment respondents would be more likely to think of family members or other acquaintances than strangers. Thus, for a NP respondent to report the "strangers" interpretation, they would have to have volunteered, as opposed to being asked about, an unlikely interpretation. These NP probes are more like the kind of probes that are described in the cognitive interviewing literature. For example, Beatty and Willis (2007) propose a taxonomy of probes that are non-directive in that they do not refer to specific problems.

Further, requiring respondents in the NP group to volunteer their own interpretation provided evidence on whether the interpretations in the DP group were in fact unlikely. That is, if respondents in the NP group were to rarely volunteer an interpretation that was directly queried in the DP group, this would help confirm that the interpretation we designed into a directive probe was not the modal interpretation and so its affirmation by DP participants would raise concern about the veracity of their affirmation.

We note that the tasks that respondents in the DP and NP groups were asked to carry out were not identical. In the DP group, the task relied primarily on recognition[2], while the NP task relied primarily on recall: a DP respondent must determine whether the probed interpretation matches what is currently in mind while an NP respondent must articulate how they interpreted the question without any potential cues from the probe.

Hypotheses

H1a: Respondents in the DP group will be more likely to affirm the interpretation mentioned in the directive probes than will be NP respondents to volunteer that interpretaton. Thus, for a NP respondent to report interpreting the question in the same way described in the corresponding directive probe, the respondent would have had to reach the same unlikely interpretation explored in the directive probe, without it being mentioned.

2 The non-directive probes provided to interviewers included a version (in parentheses to indicate they were optional) that listed a relatively exhaustive set of response options. This was done so that if NP respondents were silent after being probed, they still had a chance to report their interpretation by selecting one of these options. When this option was exercised by the interviewer, it converted the respondent's task from primarily recall to primarily recognition. It still differed from the DP task in that the options were substantive, not "yes" or "no."

H1b: Respondents in the NP group will be more likely to provide an interpretation that is not mentioned in the corresponding directive probe than will DP respondents. This assumes that, without being asked about the interpretation described in the directive probes, NP respondents are unlikely to spontaneously arrive at that interpretation. If this is the case, then they will report an alternative interpretation.

Table 1 Examples of directive and non-directive probes. The material in parentheses after the non-directive probe was intended to be read only if the respondent was having trouble answering the probe.

Question	How important is it to you that people are judged by their own personal actions, and not by the actions of other people in their families? Is this not important, a little important, important, or extremely important?
Directive Probe	When you answered this question, did you think primarily about the judgments of strangers?
Non-directive Probe	When you answered this question, who were you picturing doing the judging? (Were you primarily thinking about close family and friends, acquaintances, strangers, some combination of these types of people, or someone else?)

If affirming the interpretation proposed in a directive probe is analogous to ARS, then this behavior should be more likely for the same subgroups who exhibit more ARS.

The evidence that Latinos tend to display high levels of acquiescence (e.g., Aday et al., 1980; Liu et al., 2017; Ross & Mirowsky, 1984) which varies between Latino subgroups (Davis et al., 2019), leads to the following hypotheses:

H2a: Latino respondents – in particular Cuban Americans, Puerto Ricans, and Mexican Americans – will affirm the interpretation proposed in directive probes more than will non-Latino White respondents.

H2b: Latino respondents – in particular Cuban Americans, Puerto Ricans, and Mexican Americans – will be less likely to offer an alternative response than will non-Latino Whites.

Further, the evidence that lower levels of education are associated with higher levels of ARS (Liu et al., 2019; McClendon, 1991; Messick & Frederiksen, 1958) leads to Hypotheses 3a and 3b:

H3a: Respondents with less formal education will affirm the interpretation proposed in directive probes more than will more highly educated respondents.

H3b: Respondents with less formal education will offer an alternative response less often than will more highly educated respondents.

Lastly, the evidence that older respondents tend to engage in ARS at higher rates than younger respondents (Lechner & Rammstedt, 2015; Lechner et al., 2019; Liu et al., 2019; Meisenberg & Williams, 2008) leads to the following hypotheses:

H4a: Older respondents will affirm the interpretation proposed in directive probes more than will younger respondents.

H4b: Older respondents will offer an alternative interpretation less often than will younger respondents.

Method

The experiment was embedded within a cognitive interview pretest of a questionnaire about Latino health conducted in the United States. Two versions of the questionnaire were tested over three rounds of cognitive interviews in English (n=86) and Spanish (n=37). A total of 45 closed-form questions covering a wide variety of topics including family relations, female gender roles, male gender roles, personal beliefs, assorted opinions, and cultural values, along with directive or non-directive probes, were included in versions of the cognitive interview guide for each probe condition. The questions for which probes were developed appear in Appendix A along with the probes for the corresponding probe group. In total, 123 in-person cognitive interviews were conducted at the University of Illinois Chicago Survey Research Laboratory. Respondents were randomly assigned to either the DP or NP group[3]. The resources available to the current project allowed us to transcribe and analyze 67 audio-recorded cognitive interviews, randomly selected from the larger pool.

Respondents. We recruited the respondents by placing ads on Craigslist and in local Spanish-language newspapers as well as by posting ads on listservs and flyers in neighborhoods with a high proportion of Latino residents. In addition, respondents recruited other potential respondents by word of mouth. Finally, staff called landline telephone numbers from samples believed to overrepresent Latino households, although in the end very few respondents were recruited from this sample source. Potential respondents completed a telephone screener in which they were asked about their ethnicity, gender, preferred language, and education, among other attributes. Additionally, respondents' level of acculturation (High Bicultural, Moderate Bicultural, Strong Latino, Latino-Leaning Bicultural, Anglo-Leaning Bicultural, Unclassified) was measured using the ARSMA-II (Bowman, 2005; Cuellar, I., et al., 1995), which was adapted slightly for use with an expanded set of Latino heritage groups. Only about 3% of respondents

3 Although each recruited sample member was randomly assigned to be interviewed following the DP or NP protocol, disproportionately more DP interviews were ultimately completed.

scored "Strong Latino," i.e., high Latino, low Anglo; other high Latino respondents were also high or moderate Anglo, thus placing them in the high or moderate Bicultural categories. Taken together this pattern of acculturation suggests that this sample was relatively acculturated (see Table 2).

Table 2 Respondent acculturation levels

Acculturation	Frequency	Percent
Strong Latino (high Latino, low Anglo)	2	2.99
Latino leaning bicultural (high Latino, moderate Anglo)	23	34.33
Moderate bicultural (moderate Latino, moderate Anglo)	14	20.90
High bicultural (high Latino, high Anglo)	6	8.96
Anglo leaning bicultural (moderate Latino, high Anglo)	12	17.91
Unclassified	10	14.93
Total	67	100

The cognitive interviews were conducted in the participant's preferred language: if they preferred "Only Spanish" or spoke "Spanish better than English" the interview was conducted in Spanish; if they reported preferring "Only English" or speaking "English better than Spanish" the interview was conducted in English; and if they answered "Both Spanish and English" the interview language was chosen at random. Eligible participants self-identified as a member of one of four groups: Mexican American, Puerto Rican, Cuban American, or non-Latino White (see Table B1).

Participants were each paid $50 upon completing the interview.

Of the 67 cognitive interviews analyzed in the current study, 41 (61%) were DP, and 26 (39%) were NP interviews[4]. Respondents were randomly assigned to type of cognitive interview so that ethnicity, gender, and interview language were roughly balanced between the two probe conditions (see description of these variables in "Analytic Approach" below). The distributions were indistinguishable between the two probe conditions: ethnicity, $\chi^2(1) = 0.39$, ns; gender, $\chi^2(1) = 0.27$, ns; interview language, $\chi^2(1) = 0.33$, ns.; and (while not deliberately balanced) education $\chi^2(1) = 0.63$, ns. The distributions of ethnicity, gender, interview language, and education across the two probe conditions in the 67 cognitive interviews are presented in Appendix B (Tables B1, B2, B3, and B4, respectively).

Both a directive and nondirective probe were constructed by the study team for each of the 45 questions in the draft questionnaire (see Appendix A).

4 The sample of 67 cognitive interviews was selected blind to the type of interview; thus the greater number of DP than NP interviews in the sample reflects the greater proportions of DP interviews in the total pool.

The directive probes were asked about a specific and, in the authors' view, unlikely interpretation. The authors' judgment about the likelihood of respondents interpreting the question in this way was confirmed by the low rate at which these interpretations were volunteered by the NP respondents (see Table 5). Many of the directive probes were designed to make the target interpretation unlikely by asking if it was the respondent's only interpretation, e.g.,

Q: How much do you believe that women should be comfortable voicing their opinions to men?

DP: When you answered this question, did you think only about when women have opinions about things that affect their families?

If respondents had interpreted the question to include the probed interpretation and others, they were free to indicate this, and this would have been coded as "Partial Affirm." This was rare for DP respondents (Table 5), suggesting that when they endorsed the probed interpretation, they were reporting it as their only interpretation.

The non-directive probes included an alternative version (in parentheses) that offered the respondent an exhaustive set of interpretations including the DP interpretation and, usually, an open option, e.g., "or something else?" In other words, while still non-directive the alternative version of each non-directive probe provides options from which respondents could choose. Interviewers were instructed to administer this version of the probe when NP respondents seemed unable to answer.

Interviewers. Seven interviewers conducted the cognitive interviews (see Table 3 for interviewer characteristics). Because the current study was embedded within an actual pretest for a production survey, the assignment of respondents to interviewers was driven largely by deadlines of the parent project and thus which interviewers were available when a respondent was recruited. No records were maintained of which interviewers conducted which cognitive interviews (as is the norm, in our experience, in actual pretests). As a result, we do not know whether an interviewer conducted DP, NP or both kinds of cognitive interviews.

The interviewers were trained in general interviewing techniques, cognitive interviewing techniques, and study-specific content. In the training sessions, the purpose of the main study (not the current cognitive interviewing study) was presented to the interviewers. Interviewers were reminded what cognitive interviewing is and the differences between cognitive interviewing and standardized field interviewing. Further, interviewers were instructed to read a "How To" guide (Willis, 1999) that covered cognitive interviewing techniques including probing, background theory, examples, and detection of problems. The interviewers were instructed to read each question as worded, to ask respondents to answer each question and report on their thinking, and after the respondent had both answered the question and reported on their thinking to read the scripted probe (whether directive or non-directive).

Table 3 Cognitive interviewer characteristics

Gender	6 female, 1 male
Native Language	5 bilingual, 2 English only
Latino ethnicity	5 Latina, 2 white
Profession	4 professional survey interviewers, 2 survey research supervisors, 1 PhD level social scientist

Questions. Twelve of the draft survey questions were designed to be asked only of males in the production interview and 15 were designed to be asked only of females in the production interviews: thus, in the cognitive interviews male respondents were asked 30 questions with probes and female respondents were asked 33 questions with probes. This created 1930 responses to the probes (1842 of which were codable) from the 67 cognitive interviews, all of which were analyzed in the current study.

Behavior Coding. All 67 cognitive interview audio recordings were transcribed and, if the interview was conducted in Spanish (n=28), first translated into English. Each transcribed interview was then coded by two independent judges for respondent and interviewer behaviors. A coding scheme, consisting of five behavior codes, was developed to classify respondents' answers to the probes based on the initial coding of 16 transcripts by one of the authors. The codes are presented in Table 4. The coding task was divided among two pairs of coders; one pair coded one set of arbitrarily selected interviews, and the other pair coded the remainder. Inter-rater reliability (κ) was computed across 66 interviews (one interview was used as a training case). The κ score was 0.83 indicating "strong" (McHugh, 2012) or "nearly perfect" (Everitt & Haye, 1992) agreement between the coders. After the κ score was calculated, the coders reconciled any differences in the codes they assigned so that one set of codes was available for analysis.

Note that NP respondents could not explicitly reject the DP interpretation: because their task was to report how they interpreted the question, not whether their understanding of the question matched one proposed by the researcher, they could offer an alternative interpretation, thus implicitly rejecting the DP interpretation; DP respondents could explicitly reject the probed interpretation by responding "no" when directly probed. Further, this means that DP respondents could, at their discretion, also offer an alternative; if they did volunteer an alternative interpretation, this was coded as "Provide Alternative" not "Reject." Thus, for NP respondents, offering an alternative and affirming or partially affirming the DP interpretation exhausted the possible responses to the (non-directive) probe but this was not the case for DP respondents because they could also explicitly reject it.

Table 4 Behavior Codes

Code	Probe Condition	Description	Example (hypothetical)
Reject	DP (only)	R answered "no" to the directive probe	I: When you answered this question, did you think primarily about the judgments of strangers? R: No
Affirm	DP	R affirmed interpretation questioned in the directive probe and did not provide additional interpretation	I: When you answered this question, did you think primarily about the judgments of strangers? R: Yes
	NP	R volunteered the interpretation about which DP Rs were explicitly asked	I: When you answered this question, who were you picturing doing the judging? R: I was imagining strangers
Partially affirm	DP	R affirmed interpretation questioned in the directive probe and volunteered additional interpretation(s)	I: When you answered this question, did you think primarily about the judgments of strangers? R: Yes, and I also thought of family members
	NP	R volunteered interpretation about which DP Rs were explicitly asked and volunteered additional interpretation(s)	I: When you answered this question, who were you picturing doing the judging? R: I was imagining strangers and I also thought of family members
Provide Alternative	DP	R rejected the interpretation in the directive probe and provided alternative interpretation	I: When you answered this question, did you think primarily about the judgments of strangers? R: No, I thought of family members
	NP	R provided interpretation about which DP Rs were not explicitly asked	I: When you answered this question, who were you picturing doing the judging? R: I thought of family members.
Not codable	Observed only in NP	R's answer to the probe did not make sense or was not responsive.	I: When you answered this question, did you think ONLY about times when a woman is in physical danger? R: Generally.

The coding categories further distinguished between affirming (or volunteering for NP respondents) the DP interpretation without volunteering another interpretation (Affirm) and affirming the DP interpretation as well as volunteer-

ing at least one other interpretation (Partially Affirm). We made this distinction because we believed that affirming both the DP interpretation and at least one other suggested a weaker endorsement of the former than if it were the only interpretation endorsed.

Analytic Approach

Dependent Variables

We modeled the number of times each respondent affirmed or partially affirmed the probe for each question (i.e., probe), treating the composite variable simply as "Affirmations" (see Statistical Analysis) The other dependent variable that we modeled was "Provide Alternative," which was the number of times (questions) that each respondent offered an interpretation other than the one queried in the directive probe (excluding the alternative interpretations mentioned in partial affirms[5]). Both dependent variables were calculated at the respondent level.

Independent Variables

Probe group was a binary variable, DP or NP. Ethnicity was treated as a categorical variable in the models: Mexican American, Puerto Rican, Cuban American, and non-Latino Whites[6] (distributions of respondents across ethnic subgroups within each probe group appear in Table B1, Appendix B). Respondents' ages, which ranged from 20 to 67 years, were recoded into a categorical variable (20 – 34; 35 – 54; 55 years and older). Educational attainment was represented as a binary variable that distinguished between those with less than a bachelor's degree and those with a bachelor's degree or higher.

Statistical Analysis

To test our hypotheses, we first fit Poisson regression models to the data for affirming the DP interpretation and for providing an alternative interpretation – both of which are counts – using the glm function in R. While appropriate for this type of count data, Poisson regression requires that the variance equal the mean (Ver Hoef & Boveng 2007), which was not the case for either dependent variable, i.e., the data were overdispersed, by the odTest function in the pscl

5 Because responses to the probes could only be assigned to one category, and affirming the directive probe is assumed to be measurement error, we prioritized its detection by treating partial affirmations as affirming the directive probe.

6 Note that the number of non-Latino Whites was small, n=7 in the DP group and n=3 in the NP group.

package in R. Thus, we fit negative binomial regression models, using the glm.nb function that is part of the MASS package in R. Negative binomial regressions are appropriate for count data and can be fit despite overdispersion. The models tested the effect of probe group, ethnicity, age, and education on the number of affirmations or alternative interpretations provided by each respondent. One set of models (3 and 4) includes the interaction of probe group and education.

Results

Table 5 displays the percent of probes which each respondent, on average, affirmed, partially affirmed, implicitly rejected by offering of an alternative and, for the DP group, explicitly rejected. It was possible that DP respondents could have overwhelmingly rejected the interpretations queried in the probes (i.e., by responding "no") given the relative implausibility of these interpretations. This was not the case. DP respondents, on average, rejected fewer than half (41.6%) of the probed interpretations. Instead, they affirmed or partially affirmed the probed interpretation roughly as often (45.2%) as they rejected it. If this level of affirmation reflects the rate of actual question interpretation, as opposed to ARS-like behavior, then NP respondents should have volunteered (implicitly affirmed) those interpretations in similar proportions. This was also not the case. NP respondents, on average, affirmed (i.e., volunteered) only 9.6% of the DP interpretations. Instead, they volunteered an alternative interpretation for 74.1% of the (non-directive) probes, that is, when asked how they interpreted the question, on average three-quarters of their interpretations differed from the NP interpretation. In contrast, DP respondents offered an alternative to only 13.3% of the (directive) probes. In other words, DP respondents affirmed the interpretation queried in the probes four times more than NP respondents volunteered these interpretations, and NP respondents volunteered alternative interpretations of the questions more than five times as often as did DP respondents.

Note that DP respondents exclusively affirmed the probe about six times as often as they partially affirmed it, i.e., also affirmed at least one other interpretation (38.6% vs. 6.6%). In contrast, NP respondents volunteered (affirmed) only the DP interpretation *less* than they partially affirmed that interpretation, i.e., endorsed the DP interpretation and also offered at least one other interpretation (9.6% vs. 13.2%). Thus, it appears that directive probes greatly restricted how DP respondents understood the questions or, perhaps more plausibly, their willingness to diverge from the probed interpretation.

Table 5 Mean percent of probes that each respondent affirmed, partially affirmed, implicitly rejected by providing an alternative and, for DP respondents, explicitly rejected. (Standard deviation in parentheses.)

Probe Group	Response to Probe			
	Affirm	Partially Affirm	Provide Alternative	Reject
Directive	38.6 (0.54)	6.6 (0.24)	13.3 (0.35)	41.6 (0.55)
Nondirective	9.6 (0.20)	13.2 (0.29)	74.1 (0.59)	

Note: 3.1% of responses to nondirective probes were uncodable.

Overall, the patterns in Table 5 are consistent with both H1a and H1b. We test H1a and H1b more directly in Models 1 and 2 (Table 6). In Model 1, the greater frequency of Affirmations (pooled Affirms and Partial Affirms) for the DP compared to the NP respondents is highly significant, confirming Hypothesis 1a, and serving as a check on the probe manipulation: the DP interpretations were rarely volunteered by NP respondents, i.e., respondents who were free to report how they understood the questions without being offered an interpretation by the researchers. Similarly, the greater frequency with which NP respondents offered an alternative to the probed interpretation than did DP respondents is highly significant in Model 2, confirming H1b.

To the extent that affirming an unlikely interpretation resembles ARS, it would follow that Latino subgroups might exhibit more affirmation than non-Latino Whites (H2a). Our data do not support this hypothesis. Mexican Americans and Puerto Ricans in the DP group affirmed the probe no more often than non-Latino Whites, and Cuban Americans affirmed the probe significantly *less* often than non-Latino Whites (Model 1), in a reversal of what we predicted and what would be expected based on the ARS literature in which Latinos generally exhibit more ARS than non-Latino Whites (e.g., Aday et al., 1980; Liu et al., 2017; Ross & Mirowsky, 1984). Regarding H2b, we proposed that Latinos would be less likely to offer an alternative response than would non-Latino Whites. There was no evidence in support of this prediction as shown in Model 2: the differences between the ethnic groups and non-Latino Whites were not significant.

Table 6 Negative binomial regression results for affirming the probe (Model 1) and offering an alternative interpretation (Model 2)

Variables	Model 1 (Affirming the Probe)			Model 2 (Offering an Alternative Interpretation)		
	B	*SE*	*p-value*	*B*	*SE*	*p-value*
Intercept	1.0189	0.2176	**<.0001**	3.0685	0.2131	**<.0001**
Directive	1.2111	0.1481	**<.0001**	-2.0189	0.1319	**<.0001**
Non-Directive (ref.)	.	.	.	.	.	.
Mexican/American	0.1252	0.2106	0.5523	-0.0722	0.2231	0.7460
Puerto Rican	-0.3467	0.2188	0.1130	0.2875	0.2192	0.1896
Cuban/American	-0.9268	0.2239	**<.0001**	0.0009	0.2149	0.9963
non-Latino White (ref.)	.	.	.	.	.	.
Age, years						
20-34 (ref.)	.	.	.	.	.	.
35-54	-0.2908	0.1441	**0.0436**	0.0960	0.1458	0.5104
≥ ≥ 55	0.0929	0.1666	0.5769	0.3875	0.1823	**0.0335**
Less than a bachelor's	0.6226	0.1599	**<.0001**	-0.2138	0.1423	0.1329
Bachelor's degree or higher (ref.)	.	.	.	.	.	.

More consistent with the ARS literature (e.g., Lechner et al., 2019; Meisenberg & Williams, 2008), education was a strong predictor of affirming the probe. Those without a bachelor's degree were significantly more likely to affirm the probe than those with a bachelor's degree or higher (Model 1), a finding that supports H3a. Given the strength of the education effect, we asked whether it was equally strong for both probe groups, i.e., was the tendency to affirm the probe in the DP group stronger for those with lower levels of education as the ARS literature would predict, without being similarly moderated by education in the NP group where the task was less likely to trigger ARS? Thus, we tested the interaction between probe group and educational attainment in Models 3 and 4. As shown in Table 7, the interaction is significant in Model 3: DP respondents with less than a bachelor's degree *were* significantly more likely to affirm the probe than those with more education but education made little difference in the NP group.

Figure 1 displays the average percentage of affirmations (affirms + partial affirms) for each respondent by probe group and education level. DP respondents with a bachelor's degree or higher affirmed the probe 31% of the time compared to those with less than a bachelor's degree who affirmed the probe 53% of the time. Differences are small and in the opposite direction in the NP

Table 7 Negative Binomial regression (including an interaction term) results for affirming the probe (Model 3) and offering an alternative interpretation (Model 4)

Variables	Model 3 (Affirming the Probe)			Model 4 (Offering an Alternative Interpretation)		
	B	*SE*	*p-value*	*B*	*SE*	*p-value*
Intercept	1.5682	0.2553	**<0.0001**	3.0242	0.2302	**<0.0001**
Directive Probe	0.5345	0.2435	**0.0281**	-1.9421	0.2132	**<0.0001**
Non-Directive (ref.)	.	.	.	.	.	.
Mexican/American	-0.0078	0.1989	0.9687	-0.0462	0.2275	0.839
Puerto Rican	-0.4362	0.2064	**0.0346**	0.3059	0.2200	0.1644
Cuban/American	-0.9506	0.2131	**<0.0001**	0.0101	0.2136	0.9624
non-Latino White (ref.)	.	.	.	.	.	.
Age, years						
20-34 (ref.)	.	.	.	.	.	.
35-54	-0.3078	0.1347	**0.0224**	0.0987	0.1451	0.4962
≥ 55	0.0692	0.2611	0.7718	0.3904	0.1816	**0.0316**
Less than a bachelor's degree	-0.7570	0.1599	**<.0001**	-0.1742	0.1678	0.2990
Bachelor's degree or higher (ref.)	.	.	.	.	.	.
Directive * Less than a bachelor's degree	0.9976	0.3038	**0.001**	-0.1218	0.2739	0.6566

group: those with a bachelor's degree volunteered the DP interpretation 26% of the time and those with less than a bachelor's degree volunteered that interpretation about as often, 21% of the time. Thus, those with less formal education are driving the increased number of affirms in the DP interviews[7].

Education did not affect the frequency of offering an alternative response, leading us to reject H3b (Model 2) and the interaction between education level and probe group was not significant for offering an alternative response (Model 4). No other interactions were significant and so none are included in the models.

7 When the significant education x probe type interaction is included in Model 3, the main effect of education is significant but reversed relative to its direction in Model 1. We attribute this to education moderating the main effect of probe type (the interaction of these two variables is significant) so that when the interaction is included in the model, the residual main effect of education is what "remains" after the interaction is removed, making it largely uninterpretable.

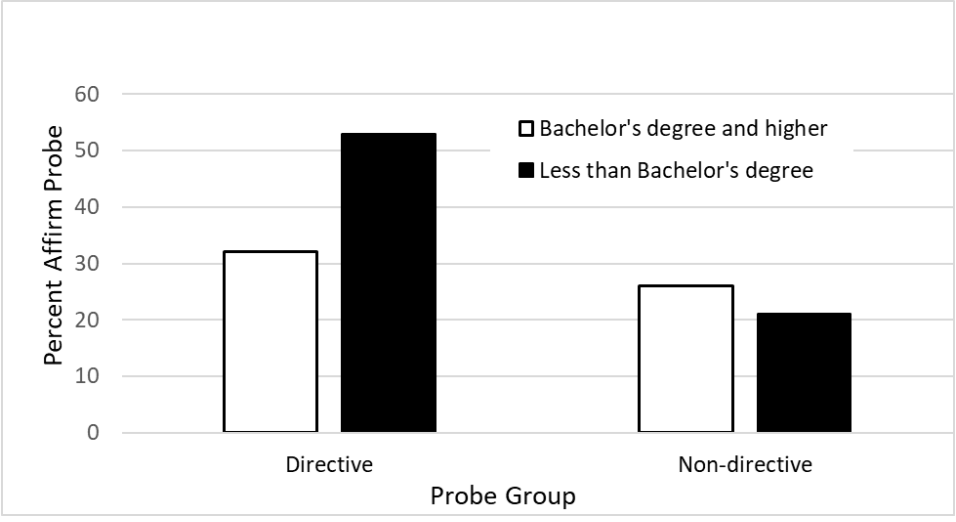

Figure 1 Percent of Affirmations (Affirm + Partial Affirm) for Probe by Education group.

Respondent age has been shown to increase ARS (e.g., Liu et al., 2019). Thus, we tested H4a (older respondents will affirm the probe more than younger respondents) in Models 1 and 3 and H4b (older respondents will offer an alternative less often than younger respondents) in Models 2 and 4. In all the models, the reference age was 20 - 34 years, the youngest group. Respondents between 35 and 54 years of age were *less* likely to affirm the probe than those under 34 years of age, reversing the typical finding in the ARS literature and contradicting H4a, significantly in Model 3 and marginally so in Model 1, while the differences between respondents 55 years and older and those 20 - 34 years of age were not statistically significant. Further, Models 2 and 4 show that those who are 55 years of age and older were significantly *more* likely to offer an alternative response in comparison to younger respondents, a reversal of the H4b prediction.

Finally, the reversal of Hypothesis 2a (more frequent affirmation of the probe among Latino subgroups) observed among Cuban Americans in Model 1 who affirmed the probe significantly less than did non-Latino Whites, is also observed among Puerto Ricans in Model 3. There is, as in Model 2, no support for Hypothesis 2b (more frequently offered alternative responses) in Model 4; none of the Latino subgroups offered alternatives at different rates than did non-Latino Whites.

Discussion

Respondents in the DP group were substantially more likely to affirm they had interpreted a set of survey questions in the rare and unintended ways queried in directive probes than were NP respondents to volunteer those interpretations without being explicitly asked about them. Similarly, NP respondents volunteered alternative interpretations substantially more often than did DP respondents. Taken together, these findings provide strong support for our original intuition that cognitive interview results are vulnerable to error when respondents are directly asked if they experienced a particular problem, especially compared to those for whom the problem was not explicitly presented.

It is possible that these differences could be related to differences in reporting tasks: for the DP group the task had the character of a *recognition* task while respondents in the NP group were, in effect, asked to *recall* their interpretation (or when the parenthesized alternative non-directive probe was administered, to choose from a set of alternatives). Recall is generally more prone to error than recognition (e.g., Anderson, 2020; Tulving & Thompson, 1973), so it is possible that NP respondents interpreted the questions in much the same way as DP respondents but simply forgot their interpretation. This seems unlikely because the interval between the question's delivery to respondents and when they were probed for their interpretation was very brief, presumably too brief for much forgetting to have occurred.

It is also possible that DP respondents' affirmations accurately reflected their interpretations, i.e., that they did in fact interpret the questions in the improbable ways queried in the directive probes. This, too, is unlikely given how rarely NP respondents volunteered the same interpretations. Moreover, the far greater frequency with which NP respondents volunteered alternative interpretations should have been mirrored by DP respondents, keeping in mind that respondents were randomly assigned to one probe group or the other. That this was not the case raises the question of why DP respondents might have affirmed understanding questions in a way that may not have been entirely accurate.

We have suggested that the patterns of results are due to ARS-like processes. To the extent that ARS is a type of survey satisficing (e.g., Krosnick, 1991; Roberts, et al., 2019), i.e., respondents simplifying the task, especially if their ability is limited, or reducing effort when their motivation is low, affirming the probe may perform much the same function. Respondents' education level serves as a proxy for cognitive ability in the survey satisficing literature (e.g., Krosnick & Alwin, 1987). In the current study, DP respondents with lower levels of education (less than a Bachelor's degree) were more likely to affirm the probe than those with more education, consistent with the satisficing view of ARS in the literature.

That the youngest respondents affirmed the probe more than the those in their middle years and offered an alternative less than respondents older than 55 years of age, reverses the general finding in the ARS literature and could argue against the ARS analogy we propose here. But it is consistent with greater survey satisficing by younger than older respondents (e.g., Anduiza & Galais, 2017; Liu et al., 2017; Zhang & Conrad, 2014), presumably reflecting younger respondents' reduced motivation.

To the extent that ARS is about getting along with conversational partners and lubricating the interaction involved in answering survey questions, we did not find evidence that affirming the probe served this purpose. Although the greater frequency of ARS among Latino than non-Latino White respondents has been attributed to *simpatía* – the cultural norm that promotes being pleasant, agreeable, likable, non-confrontational, and respectful in interpersonal interactions – two of the three Latino subgroups in the current study, Cuban Americans and, in one model, Puerto Ricans, affirmed the probe *less* than did non-Latino Whites, reversing the predicted effect; the remaining subgroup, Mexican Americans exhibited no more affirmation of the DP interpretation and volunteered alternative interpretations no less often than did non-Latino Whites. This result could reflect variation in cultural traditions between different groups of Latinos. Alternatively, it could be due to the relatively assimilated character of the Latino participants: most were at least moderately bicultural, with only two being classified as Strong Latino (Table 2). A less assimilated Latino sample might well have affirmed directly probed interpretations more than did the Latino respondents in the current study.

Related to ethnicity, the non-Latino Whites in the current study, having been recruited from sample sources believed to overrepresent Latinos, could have been more similar culturally to those who identified as Latino, e.g., they might have been highly assimilated Latinos who identify as White, than might non-Latino Whites from a more general sample source. But at least for our sample, the version of ARS we observed seems less related to *simpatía*, and more related to reducing effort when ability and motivation are lower.

Whatever the exact mechanism, directive probing appears to lead to an *affirmation bias*. It seems far easier to affirm a problem proposed by an interviewer than to generate a description of a different problem, potentially leading to false alarms (Conrad & Blair, 2009). This can certainly jeopardize the quality of information provided by cognitive interviews in which directive probes are administered, as well as the quality of survey data elicited after the questionnaire is revised – based on reported problems that include high levels of false alarms – and then administered in production research.

A clear practical implication of these findings is that interviewers should avoid directly probing specific problems in cognitive interviews. In some types of qualitative research, interviewers are authorized to confirm their understanding of the interview data with participants (e.g., Olson, 2016; Tracy, 2010), and as "detec-

tives" (Willis 1994) cognitive interviewers may wish to unambiguously confirm their hypotheses about potential problems by directly asking respondents. However, the current results seriously question the wisdom of this approach, at least as the primary method of exploring respondents' interpretations in cognitive interviews. If DP respondents were willing to agree with the unlikely interpretations suggested by the interviewers, directive probes might be affirmed even more often if the problems they mention are more plausible. This is not to suggest that directive probing is always ill-advised. For example, to clarify whether they have correctly understood something the respondent reported or implied, interviewers might directly ask respondents to confirm their understanding of the respondent's interpretation (what Conrad & Blair, 2009, called a "conditional probe") but in the absence of evidence that the respondent has interpreted the question in a particular way, the results of directive probing seem likely to mislead designers revising a questionnaire.

Instead, probing "around" the problems (interpretations) whose presence the interviewer wishes to confirm without mentioning them explicitly can corroborate their existence without introducing affirmation bias. The probes that were administered in the NP group requested open responses in most cases and so did not imply to respondents that a particular problem was under investigation. Similarly, the many example probes provided by Willis (2005) and Beatty and Willis (2007) also have an open, non-directive character. To be clear, the current results do not bear on the merits of scripted versus improvised probing: whether probes are planned or developed on the fly, it is possible to understand how respondents have understood a question without asking them to affirm they have interpreted the question in a specific, problematic way.

Next steps. We noted that respondents from some Latino subgroups affirmed directive probes less often than did non-Latino Whites, a reversal of the pattern predicted by the ARS literature, and we suggested that this pattern might have been due to the predominant bicultural orientations of the Latino participants. A future study might investigate whether less acculturated respondents, i.e., those for whom *simpatía* is presumably more prominent, might exhibit higher levels of affirming the probed interpretation in the current study. Similarly, a comparison group of non-Latino Whites who are recruited from general sample sources rather than sources in which Latino representation is expected to be high, could sharpen the comparison. If Latino subgroups and especially the least assimilated members of those subgroups exhibit more affirmation of probed interpretations than non-Latino Whites, it could begin to suggest that ARS – at least the version of it that seems to reflect a desire to avoid controversy and negativity – may be more prevalent in cognitive interviews than in those conducted for the current study.

Related to this, the sample of 67 cognitive interviews analyzed in the current study was large compared to typical pretests but not large enough for us to have full confidence that the differences in affirmations between Latino subgroups

(particularly between Cuban Americans) and non-Latino Whites would hold up with larger samples (see Blair & Conrad, 2011). Increasing the number (and typicality) of non-Latino Whites in a future study could reveal effects of subgroup membership, which were not detected in the current study due to insufficient power.

A follow-up study would not only collect data from a larger number of respondents in each subgroup but would (1) recruit a larger number of cognitive interviewers and, (2) randomly assign them to conduct either DP or NP interviews. This would make it possible to explore interviewer clustering of affirmations, alternative interpretations, and in DP interviews, rejections of the DP interpretation. It is possible that because of differences in how individual interviewers administer the probes or even deliver the draft questions, different interviewers might elicit different patterns of responses to the probes, analogous to interviewer effects in standardized, production interviews (e.g., Fowler, Lewis, & Magione, 1992; Groves & Magilavey, 1986; West & Blom 2017).

Davis et al. (2019) report no evidence that interviewers' characteristics affected ARS in production interviews, possibly suggesting that the way they conduct cognitive interviews is not related to the kind of ARS-like behavior observed in the current study. Nonetheless, it would be worth testing if cognitive interviewers differ from each other in whether and how often they administer directive probes. If interviewer effects of this type are small this would bolster the current findings and suggest that it is possible for cognitive interviewers to consistently explore question understanding without probing specific interpretations.

Conclusion

There is little doubt that revising survey questionnaires based on pretests is a low-cost way to help assure that the data collected in production research are as high quality as possible. The current study adds complexity to this view by providing evidence that sometimes pretest results are themselves subject to measurement error, in this case an affirmation bias that is triggered by interviewers probing specific misinterpretations of questions in cognitive interviews. Taking steps to reduce this source of error by, for example, training cognitive interviewers to avoid directive probing, seems likely to succeed. But it suggests that researchers may need to be more discriminating in how they interpret the results of cognitive interviews before revising questions based on those results and, more generally, to examine what other types of error in cognitive interview data may be attributable to how those interviews are conducted.

References

Aday, L. A., Chiu, G. Y., & Andersen, R. (1980). Methodological issues in health care surveys of the Spanish heritage population. *American Journal of Public Health, 70*(4), 367-374. https://doi.org/10.2105/AJPH.70.4.367

Anderson, J. R. (2020). *Cognitive psychology and its implications* (9th ed.). Worth Publishers.

Anduiza, E., & Galais, C. (2017). Answering without reading: IMCs and strong satisficing in online surveys. *International Journal of Public Opinion Research, 29*(3), 497-519. https://doi.org/10.1093/ijpor/edw007

Baumgartner, H., & Steenkamp, J. B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143 - 156. https://doi.org/10.1509/jmkr.38.2.143.18840

Beatty, P.C. (2004). The dynamics of cognitive interviewing. In Presser, S. Rothgeb, J.M., Couper, M.P., Lessler, J.Y., Martin, E., Martin, J., and Singer, E. (Eds.). *Methods for testing and evaluating survey questionnaires* (pp. 45-66). John Wiley and Sons.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287-311. https://doi.org/10.1093/poq/nfm006

Billiet, J. B., & Davidov, E. (2008). Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research, 36*(4), 542-562. https://doi.org/10.1177/0049124107313901

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling, 7*(4), 608-628. https://doi.org/10.1207/S15328007SEM0704_5

Bauman, S. (2005). The reliability and validity of the brief acculturation rating scale for Mexican Americans–II for children and adolescents. *Hispanic Journal of Behavioral Sciences, 27*(4), 426-441. https://doi.org/10.1177/0739986305281423

Blair, J., & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly, 75*(4), 636-658. https://doi.org/10.1093/poq/nfr035

Cuellar, I., Arnold, B., & Maldonado, R. (1995). Acculturation rating scale for Mexican Americans-II: A revision of the original ARSMA scale. *Hispanic Journal of Behavioral Sciences, 17*(3), 275-304. https://doi.org/10.1177/07399863950173001

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187-212. https://doi.org/10.1177/0022022100031002003

Conrad, F. G., & Blair, J. (2009). Sources of error in cognitive interviews. *Public Opinion Quarterly, 73*(1), 32-55. https://doi.org/10.1093/poq/nfp013

Davis, R. E., Conrad, F. G., Dong, S., Mesa, A., Lee, S., & Johnson, T. P. (2024). An ounce of prevention: using conversational interviewing and avoiding agreement response scales to prevent acquiescence. *Quality & Quantity, 58*(1), 471-495. https://doi.org/10.1007/s11135-023-01650-7

Davis, R. E., Johnson, T. P., Lee, S., & Werner, C. (2019). Why do Latino survey respondents acquiesce? Respondent and interviewer characteristics as determinants of cultural patterns of acquiescence among Latino survey respondents. *Cross-Cultural Research, 53*(1), 87-115. https://doi.org/10.1177/1069397118774504

Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (Revised Edition)*. MIT Press. https://doi.org/10.7551/mitpress/5657.001.0001

Everitt, B. S., & Hay, D. F. (1992). *Talking about statistics: A psychologist's guide to data analysis*. Halsted Press.

Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly, 50*(2), 251-266. https://doi.org/10.1086/268979

Hoffmann, S., Mai, R., & Cristescu, A. (2013). Do culture-dependent response styles distort substantial relationships? *International Business Review, 22*(5), 814-827. https://doi.org/10.1016/j.ibusrev.2013.01.008

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

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

Lechner, C. M., & Rammstedt, B. (2015). Cognitive ability, acquiescence, and the structure of personality in a sample of older adults. *Psychological Assessment, 27*(4), 1301–1311. https://doi.org/10.1037/pas0000151

Lechner, C. M., Partsch, M. V., Danner, D., & Rammstedt, B. (2019). Individual, situational, and cultural correlates of acquiescent responding: Towards a unified conceptual framework. *British Journal of Mathematical and Statistical Psychology, 72*(3), 426-446. https://doi.org/10.1111/bmsp.12164

Liu, M., Conrad, F. G., & Lee, S. (2017). Comparing acquiescent and extreme response styles in face-to-face and web surveys. *Quality & Quantity, 51,* 941-958. https://doi.org/10.1177/10731911211042932

Liu, M., Suzer-Gurtekin, Z., Keusch, F., & Lee, S. (2018). Response styles in cross-cultural surveys. In Johnson, T. P., Pennell, B. E., Stoop, I. A., & Dorer, B. (Eds.). *Advances in comparative survey methods: Multinational, multiregional, and multicultural contexts (3MC)* (477-499). *John Wiley & Sons.* https://doi.org/10.1002/9781118884997

Mangione, T. W., Fowler, F. J., & Louis, T. A. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics, 8*, 293-293. https://doi.org/10.1093/jssam/smz031

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences, 44*(7), 1539-1550. https://doi.org/10.1016/j.paid.2008.01.010

Messick, S., & Frederiksen, N. (1958). Ability, acquiescence, and "authoritarianism". *Psychological Reports, 4*(3), 687-697. https://doi.org/10.2466/pr0.1958.4.3.687

McClendon, M. J. (1991a). Acquiescence and recency response-order effects in interview surveys. *Sociological Methods & Research, 20(1), 60-103.* https://doi.org/10.1177/0049124191020001003

McClendon, M. J. (1991b). Acquiescence: Tests of cognitive limitations and question ambiguity hypotheses. *Journal of Official Statistics, 7(2), 153-166.* https://doi.org/10.1037/met0000631

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22(3), 276–282.* https://doi.org/10.11613/BM.2012.031

Miller, K. (2011). Cognitive interviewing. In Maddans, J., Miller, K., Maitland, A. & Willis, G. (eds.), *Question evaluation methods (pp. 51-75). Hoboken, NJ: John Wiley & Sons.*

Olson, K. (2016). *Essentials of qualitative interviewing.* Routledge. https://doi.org/10.4324/9781315429212

Priede, C., Jokinen, A., Ruuskanen, E., & Farrall, S. (2014). Which probes are most useful when undertaking cognitive interviews? *International Journal of Social Research Methodology, 17(5), 559-568.* https//doi.org/10.1080/13645579.2013.799795

Ridolfo, H., Ott, K., Beach, J., & McCarthy, J. S. (2020). Pre-testing establishment surveys: Moving beyond the lab. *Survey Practice, 13*(1), 11810. https://doi.org/10.29115/SP-2020-0003

Roberts, C., Gilbert, E., Allum, N., & Eisner, L. (2019). Research synthesis: Satisficing in surveys: A systematic review of the literature. *Public Opinion Quarterly, 83*(3), 598-626. https://doi.org/10.1093/poq/nfz035

Ross, C. E., & Mirowsky, J. (1984). Socially-desirable response and acquiescence in a cross-cultural survey of mental health. *Journal of Health and Social Behavior,* 189-197. https://doi.org/10.2307/2136668

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry, 16*(10), 837-851. https://doi.org/10.1177/1077800410383121

Triandis, H. C., Marin, G., Lisansky, J., & Betancourt, H. (1984). Simpatía as a cultural script of Hispanics. *Journal of Personality and Social Psychology, 47*(6), 1363. https://doi.org/10.1037/0022-3514.47.6.1363

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352-373. https://doi.org/10.1037/h0020071

Ver Hoef, J. M., & Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology, 88*(11), 2766-2772. https://doi.org/10.1890/07-0043.1

Weijters, B., Schillewaert, N., & Geuens, M. (2008). Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science, 36*(3), 409-422. https://doi.org/10.1007/s11747-007-0077-6

West, B. T., & Blom, A. G. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology, 5(2), 175-211.* https://doi.org/10.1093/jssam/smw024

Willis, G. B. (1999). Reducing Survey Error through Research on the Cognitive and Decision Processes in Surveys. https://www.hkr.se/contentassets/9ed7b1b3997e4bf4baa8d4eceed5cd87/gordonwillis.pdf.

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

Willis, G. B. (2015). The practice of cross-cultural cognitive interviewing. *Public Opinion Quarterly, 79*(S1), 359-395. https://doi.org/10.1093/poq/nfu092

Willis, G. B. (2018). Cognitive interviewing in survey design: State of the science and future directions. In D. Vannette and J. Krosnick (Eds.), *The Palgrave handbook of survey research,* 103-107. https://doi.org/10.1007/978-3-319-54395-6_14

Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology, 67*(5), 555-561. https://doi.org/10.1037/0021-9010.67.5.555

Wright, J., Moghaddam, N., & Dawson, D. L. (2021). Cognitive interviewing in patient-reported outcome measures: A systematic review of methodological processes. *Qualitative Psychology, 8*(1), 2–29. https://doi.org/10.1037/qup0000145

Zhang, C., & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods* 8(2),127-135.

Zong, J. (2022). A mosaic, not a monolith: A profile of the US Latino population, 2000 – 2022. Latino Policy & Politics Institute, UCLA. https://latino.ucla.edu/research/latino-population-2000-2020/

Appendix A

Questions with Probes.

Note that Non-Directive probes included in parentheses a version with an relatively exhaustive set of options so that respondents could choose one of several options if they were otherwise silent.

(How important is it to you that) teenage children be encouraged to develop their independence? (Is this…)

1 ☐ not important,
2 ☐ a little important,
3 ☐ important, or
4 ☐ extremely important?

[PROBE: DIRECTIVE] When you answered this question, did the word "encourage" make you think about providing rewards to children, such as sweets or present, when they show independence?
[PROBE: NON-DIRECTIVE] When you answered this question, in what ways were you thinking that teenage children would be encouraged to develop their independence? (Did the word "encourage" make you think about encouraging independence by providing opportunities, praise, rewards such as sweets or presents, some combination of these things, or something else?)

(How important is it to you that) people are judged by their own personal actions, and not by the actions of other people in their families? (Is this…)

1 ☐ not important,
2 ☐ a little important,
3 ☐ important, or
4 ☐ extremely important?

[PROBE: DIRECTIVE] When you answered this question, did you think primarily about the judgments of strangers?
[PROBE: NON-DIRECTIVE] When you answered this question, who were you picturing doing the judging? (Were you primarily thinking about close family and friends, acquaintances, strangers, some combination of these types of people, or someone else?)?

[TO BE ADMINISTERED TO FEMALE PARTICIPANTS ONLY]

How much do you believe that women are more responsible than men for taking care of the emotional needs of their families? Would you say you…

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, were you thinking primarily of those situations in which someone in the family is upset?
[PROBE: NON-DIRECTIVE] When you answered this question, what did "taking care of emotional needs" mean to you? (Did "taking care of emotional needs" make you think primarily about situations in which someone in the family was upset, making sure people are happy on a day-to-day basis, both of these situations, or something else?)

How much do you believe that a woman should think of others' needs before her own? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about what a woman thinks, regardless of how she acts?
[PROBE: NON-DIRECTIVE] When you answered this question, were you thinking only about how women think, or also how they act?

How much do you believe that women should be comfortable voicing their opinions to men? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think only about when women have opinions about things that affect their families?

[PROBE: NON-DIRECTIVE] When you answered this question, what kinds of opinions were you thinking about? (Were you thinking about women's opinions about things that affect their families or their opinions in general?)

How much do you believe that a woman has to be strong to be successful in life? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think that "strong" means that a woman insists on things being done her way?
[PROBE: NON-DIRECTIVE] What does the word "strong" mean to you in this question? (Did the word "strong" make you think about a woman insisting on getting her way, being physically strong, not showing fear, some combination of these things, or something else?)

How much do you believe that important decisions should be made by the man of the household? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, were you thinking that the man would only make important decisions after consulting his wife?
[PROBE: NON-DIRECTIVE] When you answered this question, were you thinking that the man would make important decisions all on his own, after talking with his wife, after talking with other family members, some combination of these actions, or something else?

How much do you believe that a woman should be free to make up her own mind? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think only about major decisions that a woman makes in her lifetime such as whether or not to get married or work outside the home?
[PROBE: NON-DIRECTIVE] What kinds of things were you thinking that a woman would make up her mind about when you answered this question? (Were you thinking about major decisions that a woman makes in her lifetime such as whether or not to get married or work outside the home, more minor decisions such as what clothes to buy, her opinions about things in general such as what she thinks about climate change, people, or movies, or some combination of these types of things?)

How much do you believe that a woman should never show fear? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about times when a woman is in physical danger?
[PROBE: NON-DIRECTIVE] What kinds of situations were you thinking about when you answered this question? (Were you thinking only about situations in which a woman simply feels nervous or uncomfortable such as when she is talking in front of a group of people, only situations in which she is in physical danger, both of these types of situations, or something else?)

How much do you believe that a woman should obey her husband's wishes? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about things that a husband feels very strongly about?
[PROBE: NON-DIRECTIVE] What kinds of "wishes" were you thinking about when you answered this question? (Were you thinking about a husband's wishes about small things, things that he feels very strongly about, both of these types of wishes, or something else?)

How much do you believe that although the man may not know it, the decisions are really made by the woman of the house? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about small decisions?
[PROBE: NON-DIRECTIVE] What kinds of decisions were you thinking about when you answered this question? (Were you thinking about only small decisions, only big decisions, or decisions in general?)

How much do you believe that women hold the most power within their households? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about a woman's power over the children in the household?
[PROBE: NON-DIRECTIVE] When you answered this question, which people were you thinking about when it comes to a woman's power? (Were you thinking mostly about how much power a woman has over her husband, her children, other people in the household, or some combination of these types of people?)

How much do you believe that a woman needs to be strong willed to gain the respect of others? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did the word "others" make you think ONLY about men?
[PROBE: NON-DIRECTIVE] Who did you think of as the "others" when you answered this question? (Were you thinking mostly about men, mostly about women, or both men and women?)

How much do you believe that women should be in charge of making their own decisions about their lives? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about minor decisions?
[PROBE: NON-DIRECTIVE] What kinds of decisions were you thinking about when you answered this question? (Were you thinking about minor decisions, only about major decisions, or all types of decisions?)

How much do you believe that it is the responsibility of the woman in the household to set a moral example for her family to follow? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think that "family" referred mainly to the woman's husband?
[PROBE: NON-DIRECTIVE] Which people in the family do you think this question is asking about? (Were you thinking mainly about the woman's husband, mainly about her children, someone else, or some combination of these people?

How much do you believe that child care should primarily be a woman's responsibility? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think it was asking who decides how children should be cared for, regardless of who actually provides the care?
[PROBE: NON-DIRECTIVE] When you answered this question, what did "child care" mean to you? (Did you think the question was asking who decides how

children should be cared for, who actually provides the care, both making decisions and providing the care, or something else?)

How much do you believe that a woman should not let others tell her what to do? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think that it was mostly asking about strangers?
[PROBE: NON-DIRECTIVE] When you answered this question, who were the "others" that you think this question was asking about? (Did you think that this question was mostly asking about a woman's husband, her children, her parents, her co-workers, strangers, someone else, or some combination of people?)

[TO BE ADMINISTERED TO MALE PARTICIPANTS ONLY]

How much do you believe that a man should be affectionate with his children? Would you say you…

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you compare how affectionate men and women are with their children?
[PROBE: NON-DIRECTIVE] When you answered this question, were you thinking only about men or were you comparing men and women?

How much do you believe that a man should not let others tell him what to do? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think that it was mostly asking about co-workers?
[PROBE: NON-DIRECTIVE] When you answered this question, who were the "others" that you think this question was asking about? (Did you think that this question was mostly asking about a man's wife, his children, his parents, his co-workers, strangers, someone else, or some combination of people?)

How much do you believe that a man should never show fear? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about times when a man is in physical danger?
[PROBE: NON-DIRECTIVE] What kinds of situations were you thinking about when you answered this question? (Were you thinking only about situations in which a man simply feels nervous or uncomfortable such as when he is talking in front of a group of people, only situations in which he is in physical danger, both of these types of situations, or something else?)

How much do you believe that it is necessary for a man to fight when challenged? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about older men?
[PROBE: NON-DIRECTIVE] When you answered this question, were you mostly thinking about younger men, middle-aged men, older men, some combination of ages, or men of all ages?)

How much do you believe that it is important for women to look good? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think primarily about whether women look clean and proper, as though they were going to church?
[PROBE: NON-DIRECTIVE] When you heard this question, what did "look good" mean to you? (Were you thinking whether women dress and do their hair in a sexy way, have a sexy figure, look clean and proper as though they were going to church, some combination of these things, or something else?)

How much do you believe that men cannot be expected to be as honorable as women? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about whether men could be as faithful to their wives and girlfriends as women are to their male partners?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of actions or qualities were you thinking about? (Were you thinking only about whether men could be as faithful to their wives and girlfriends as women are to their male partners, whether men are as religious as women, whether men are as moral as women, some combination of these things, or something else?)

How much do you believe that a man should be in control of his wife? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about times when the man and his wife are out in public?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of situations were you thinking about? (Were you thinking mostly about times when the man and his wife are in the privacy of their home, at the homes of family or friends, out in public, in some other type of situation, or some combination of situations?)

How much do you believe that if a woman is being insulted, a man should defend her? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about times when a woman is insulted by another woman?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of situations were you thinking about? (Were you thinking mostly about times when a woman is insulted by another man, times when a woman is being insulted by another woman, times when a woman is being insulted by someone else, or anytime a woman is being insulted?)

How much do you believe that men should be in charge of the finances in their households? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about major financial decisions?
[PROBE: NON-DIRECTIVE] What types of finances or financial decisions were you thinking about when you answered this question? (Were you thinking only about minor financial decisions, only about major financial decisions, both minor and major types of financial decisions, or something else?)

How much do you believe that a man obtains honor from treating other people with respect? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about how a man treats strangers?
[PROBE: NON-DIRECTIVE] What types of "other people" were you thinking about when you answered this question? (Were you thinking about how a man treats his wife, his children, close family and friends, acquaintances, strangers, someone else, or some combination of these types of people?

How much do you believe that men should not talk about their feelings? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about whether or not men should talk about their feelings with their wives?
[PROBE: NON-DIRECTIVE] When you answered this question, which people were you thinking about in terms of whom men should talk or not talk to about their feelings? (Were you thinking about whether or not men should talk about their feelings with their wives, their children, other close family or friends, acquaintances, strangers, someone else, or some combination of these types of people?)

How much do you believe that a man should respect a woman's opinion, regardless of her age? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about situations when the man is at work?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of settings were you thinking about? (Were you thinking mostly about private conversations between a man and his wife, social settings with friends and families, situations when the man is at work, something else, or some combination of these types of situations?)

How much do you believe that there are many things we have not discovered yet, so nobody can be absolutely certain that their beliefs are right? Would you say you…

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about the things that people learn through their personal experiences?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of discoveries were you thinking about? (Were you thinking mostly about the things that people learn through their personal experiences throughout their lives, mostly about discoveries made by scientists, both of these types of things, or something else?)

How much do you believe that it is good to be open-minded? (Would you say you…)

1 ☐ don't believe that at all,
2 ☐ believe that a little,
3 ☐ somewhat believe that, or
4 ☐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about your willingness to have new experiences?
[PROBE: NON-DIRECTIVE] When you answered this question, what kinds of things were you thinking about being open-minded about? (Were you thinking mostly about your willingness to have new experiences, to meet new people, to accept new ideas, something else, or some combination of these types of things?)

How much do you believe that you are certain that your ideas about the central issues in life are correct? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about having a happy family?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of "central issues in life" were you thinking about? (Were you thinking mostly about having a happy family, the way in which the world works, the meaning of life, the existence of God, whether human nature is essentially good or bad, something else, or some combination of these types of things?)

How much do you believe that it is better to risk saying too much than to risk being misunderstood? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about times when you are talking with someone who you know well?
[PROBE: NON-DIRECTIVE] When you answered this question, were you thinking mostly about times when you are talking with strangers, acquaintances, people who know each other well, or some combination of these types of people?

How much do you believe that how something is said generally communicates more information than the words used to say it? (Would you say you…)

1 ❐ don't believe that at all,
2 ❐ believe that a little,
3 ❐ somewhat believe that, or
4 ❐ believe that very much?

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about how the gestures that people use when they talk?

[PROBE: NON-DIRECTIVE] What does "how something is said" mean to you in this question? (Were you thinking mostly about how the gestures that people use when they talk, about how a person emphasizes certain words, how a person uses facial expressions to communicate meaning, something else, or some combination of things?)

In general, white Americans treat Latinos with respect.

1 ☐ 1 — STRONGLY DISAGREE
2 ☐ 2
3 ☐ 3
4 ☐ 4
5 ☐ 5
6 ☐ 6
7 ☐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you think ONLY about whether or not white Americans are polite, such as saying "please" or "thank you"?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of situations were you thinking about? (Were you thinking ONLY about whether or not white Americans are polite such as saying "please" or "thank you, whether they support laws and policies that help Latinos such as immigration reform and access to education and health care, whether they hire Latinos for jobs, something else, or some combination of things?)

Gay marriage should be illegal.

1 ☐ 1 — STRONGLY DISAGREE
2 ☐ 2
3 ☐ 3
4 ☐ 4
5 ☐ 5
6 ☐ 6
7 ☐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you think about men marrying men as well as women marrying women?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of couples were you thinking about? (Were you thinking mostly about men marrying men, women marrying women, both, or something else?)

If I were to choose a snack from a store, I would probably choose something sweet.

1 ❐ 1 — STRONGLY DISAGREE
2 ❐ 2
3 ❐ 3
4 ❐ 4
5 ❐ 5
6 ❐ 6
7 ❐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you think that "something sweet" included fruits such as bananas or pineapple?
[PROBE: NON-DIRECTIVE] When you answered this question, what kind of snacks were you thinking about? (Were you thinking about fruits such as bananas or pineapple, sugary snacks such as candy bars, both of these types of snacks, or something else?)

I enjoy watching reality TV shows.

1 ❐ 1 — STRONGLY DISAGREE
2 ❐ 2
3 ❐ 3
4 ❐ 4
5 ❐ 5
6 ❐ 6
7 ❐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you include nature shows?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of TV shows were you thinking about? (Were you including only reality TV shows, or did you also include news shows, nature shows, educational TV shows, something else, or some combination of shows?)

I enjoy cold weather.

1 ❐ 1 — STRONGLY DISAGREE
2 ❐ 2
3 ❐ 3
4 ❐ 4
5 ❐ 5
6 ❐ 6
7 ❐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you compare how much you enjoy cold weather versus hot weather?
[PROBE: NON-DIRECTIVE] What were you thinking about when you answered this question? (Were you thinking only about how much you enjoy hot weather or did it make you compare how much you enjoy cold weather versus hot weather?)

I always remain calm during a crisis.

1 ❐ 1 — STRONGLY DISAGREE
2 ❐ 2
3 ❐ 3
4 ❐ 4
5 ❐ 5
6 ❐ 6
7 ❐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] Did you include heated arguments between people as "crises" when you answered this question?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of crises were you thinking about? (Were you thinking that "crises" referred only to emergencies, or did you also include heated arguments between people or other types of situations?)

I enjoy listening to stories.

1 ☐ 1 — STRONGLY DISAGREE
2 ☐ 2
3 ☐ 3
4 ☐ 4
5 ☐ 5
6 ☐ 6
7 ☐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] Did you include listening to jokes when answering this question?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of stories were you thinking about? (Were you thinking only about stories that people tell, jokes, books on tape, some other type of stories, or some combination of these types of things?)

Gays should have the same marriage rights as straight men and women.

1 ☐ 1 — STRONGLY DISAGREE
2 ☐ 2
3 ☐ 3
4 ☐ 4
5 ☐ 5
6 ☐ 6
7 ☐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about the rights and privileges that come with being married, such as visiting a spouse in the hospital?
[PROBE: NON-DIRECTIVE] When you answered this question, what did "marriage rights" mean to you? (Were you thinking mostly about the right to get married, the rights and privileges that come with being married such as visiting a spouse in the hospital, both of these types of rights, or something else?)

When it comes to interacting with Latinos, most white Americans are racist.

1 ❐ 1 — STRONGLY DISAGREE
2 ❐ 2
3 ❐ 3
4 ❐ 4
5 ❐ 5
6 ❐ 6
7 ❐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about how white Americans react to hearing the opinions of Latinos on the TV, radio, or internet?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of interactions were you thinking about? (Were you thinking mostly about when white Americans are talking face-to-face with Latinos, mostly about how white Americans react to hearing the opinions of Latinos on the TV, radio, or internet, something else, or some combination of types of interaction?)

I like to spend time outdoors every day.

1 ❐ 1 — STRONGLY DISAGREE
2 ❐ 2
3 ❐ 3
4 ❐ 4
5 ❐ 5
6 ❐ 6
7 ❐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about active outdoor activities, such as walking?
[PROBE: NON-DIRECTIVE] When you answered this question, what types of activities were you thinking about? (Were you thinking of doing any particular activities – such as walking or gardening – or just sitting and enjoying the outdoors?)

Being around other people gives me energy.

1 ☐ 1 — STRONGLY DISAGREE
2 ☐ 2
3 ☐ 3
4 ☐ 4
5 ☐ 5
6 ☐ 6
7 ☐ 7 — STRONGLY AGREE

[PROBE: DIRECTIVE] When you answered this question, did you think mostly about feeling emotionally energized?
[PROBE: NON-DIRECTIVE] When you answered this question, what type of energy were you thinking about? (Were you thinking mostly about physical energy, mental energy, spiritual energy, something else, or some combination of types of energy?)

Appendix B

Distribution of Demographic Characteristics in the Two Probe Groups

Table B1 Latino ethnicity

Probe Group		Latino Ethnicity				
		Non-Latino White	Mexican-American	Puerto-Rican	Cuban American	Total
Directive	n	7	12	12	10	41
	%	17.1	29.3	29.3	24.4	100
Non-Directive	n	3	8	8	7	26
	%	11.5	30.8	30.8	26.9	100
Total	n	10	20	20	17	67
	%	14.9	29.9	29.9	25.4	100

Table B2 Gender

Probe Group		Gender		
		Male	Female	Total
Directive	n	21	20	41
	%	51.2	48.8	100
Non-Directive	n	15	11	26
	%	57.7	42.3	100
Total	n	36	31	67
	%	53.7	46.3	100

Table B3 Interview language

Probe Group		Interview Language		
		English	Spanish	Total
Directive	n	25	16	41
	%	60.9	39.0	100
Non-Directive	n	14	12	26
	%	53.9	46.2	100
Total	n	39	28	67
	%	58.2	41.8	100

Table B4 Education

Probe Group		Education					
		Less than High School	High School graduate	Some College	Bachelor's Degree	Graduate Degree	Total
Directive	n	5	9	10	9	6	39
	%	12.8	23.0	25.6	23.1	15.4	100
Nondirective	n	5	6	6	6	3	26
	%	19.2	23.1	23.1	23.1	11.5	100
Total	n	10	15	16	15	9	65
	%	15.4	23.1	24.6	23.1	13.9	100

Online First publication date: 25-02-2025