Question Order Effects in Cross-Cultural Web Probing: Pretesting Behavior and Attitude Questions

Hadler, Patricia

Mitglied der

Leibniz-Gemeinschaft

Question Order Effects in Cross-Cultural Web Probing: Pretesting Behavior and Attitude Questions

Patricia Hadler[1]

Abstract

Cognitive pretesting is an essential method of piloting questionnaires and ensuring quality of survey data. Web probing has emerged as an innovative method of cognitive pretesting, especially for cross-cultural and web surveys. The order of presenting questions in cognitive pretesting can differ from the order of presentation in the later survey. Yet empirical evidence is missing whether the order of presenting survey questions influences the answers to open-ended probing questions. The present study examines the effect of question order on web probing in the United States and Germany. Results indicate that probe responses are not strongly impacted by question order. However, both content and consistency of probe responses may differ cross-culturally. Implications for cognitive pretesting are discussed.

Pretesting questionnaires before fielding is considered indispensable by both elementary textbooks and experienced researchers in order to ensure data quality (Presser & Blair, 1994; Presser et al., 2004). Cognitive pretesting asks respondents to verbalize their thought processes while answering survey questions (Beatty & Willis, 2007). The method is used to evaluate respondents' answer process, identify problems they encounter while answering survey questions, and, based on these findings, suggest question revisions prior to data collection (Willis, 2015a).

Until recently, cognitive pretesting took place mainly in the form of face-to-face interviews (Willis, 2005). In a cognitive interview, a participant is presented a survey question by an interviewer and answers it. This is followed by one or several "specific questions or *probes* about how the participant set about answering the question being tested" (Collins, 2015, p. 14). For instance, a respondent may answer a survey question on their past behavior. Subsequent probing questions might ask how the respondent remembered their past behavior or what type of activities they

---

[1] GESIS - Leibniz Institute for the Social Sciences, Mannheim, Germany

Corresponding Author:
Patricia Hadler, GESIS - Leibniz Institute for the Social Sciences, P.O. Box 12 21 55, Mannheim 68072, Germany.
Email: patricia.hadler@gesis.org

included in their answer. Once the probes have been answered, the interviewer and participant move on to the next survey question and probes relating to this question. If the next question is an attitude question on the same topic, probing questions may be used to understand why the respondent agreed or disagreed with the statement or how they came to their answer.

In the past decade, web probing has developed as a complementary self-administered form of cognitive pretesting (Behr, Braun, et al., 2012; Behr, Kaczmirek, et al., 2012). Implementing probing techniques from cognitive interviewing into web surveys has proven especially useful to test web surveys in the same mode (Fowler & Willis, 2020, p. 466) and has gained popularity in the context of cross-national pretests (Behr et al., 2020; Braun et al., 2014). Online probes are open-ended questions that directly relate to a foregoing (usually closed-ended) survey question (Behr et al., 2017). Web probing encounters higher levels of nonresponse and generates shorter probe responses than cognitive interviews (Lenzner & Neuert, 2017; Meitinger & Behr, 2016). However, previous research indicates that both modes generate similar findings (Meitinger & Behr, 2016) and lead to similar revisions (Lenzner & Neuert, 2017).

Regardless of mode, the focus of cognitive pretesting lies on testing individual questions rather than the questionnaire as a whole (Lenzner et al., 2016). Cognitive pretests take place at an early stage in questionnaire development and usually examine only parts of questionnaires. In consequence, question sequence in cognitive pretesting must not necessarily be identical to that in the later survey. Ultimately, analysis of cognitive pretesting rests on the assumption that pretest results are independent of a tested question's position during pretesting.

Although cognitive pretesting has been used in the past to understand the causes of question order effects (Bishop, 1992; Bishop et al., 1985), the reverse effect of question order on cognitive pretesting has yet to be examined. To date, we do not know whether the order of presenting survey questions affects the responses given to probing questions and, if this is the case, whether this can impact revisions to survey questions. This research gap is all the more surprising, as the order of presenting questions can influence the response to survey questions (e.g., Schwarz & Sudman, 1992), and the response to survey questions is in turn known to influence respondents' likelihood of giving substantive probe responses (Zuell & Scholz, 2015). Should probe responses be impacted by the sequence of survey questions in a cognitive pretest, this endangers the validity of cognitive pretest results and, in consequence, the quality of the final questionnaire and survey data.

Question order effects "occur most when two questions are asked sequentially on the same topic or very similar topics" (Stark et al., 2018, p. 28; Tourangeau et al., 2003). One such case is the presentation of behavior and corresponding attitude questions. Questionnaires frequently include both question types, as many psychological frameworks use attitudinal measures to explain or predict behavior (most prominently in the theory of reasoned action; Ajzen & Fishbein, 1980; and the theory of planned behavior; Ajzen, 1985). Attitude–behavior consistency is generally moderate (Ajzen & Fishbein, 1977) and can be impacted by question order (Budd, 1987).

Even well-documented question order effects vary across countries. Reasons for differences in question order effects between countries are manifold, including substantive, topic-specific differences between countries but also differences in survey response styles (Stark et al., 2018). Moreover, past research on web probing has repeatedly pointed to cross-cultural differences in probe response quality irrespective of survey response behavior (e.g., Meitinger et al., 2019).

The research presents an experimental study on the effects of question order of a behavior and attitude survey question on probe responses in cross-cultural web probing. The following section discusses the potential influence of question order on probe response content and consistency. The topic of fare evasion is introduced as an example of related behavior and attitude questions. Next, the research design is presented, followed by results. The discussion outlines the implications for web probing methodology.

Question Order and Probe Responses

Examining the impact of question order on open-ended probes differs in several ways from analyzing closed-ended survey questions. For one, responses to open-ended questions are not limited by predefined answer categories (Reja et al., 2003). Moreover, examining order effects for probes is more complex as it involves not just two questions but at least four: two survey questions and two probes. On the other side, surveys and cognitive pretests have in common that they are founded on the dialogue structure of asking and answering and the reliance on contextual information to interpret questions and give meaningful answers (Conrad et al., 2014).

Probes can directly follow the survey question they pertain to (concurrent or embedded probing) or be administered after several or all survey questions (retrospective probing; Collins, 2015, p. 120). The rationale behind concurrent probing is that respondents are better able to recall their thoughts while answering the survey question and do not postrationalize their survey responses. Retrospective probing can be used to prevent probes from influencing thought processes during and responses to subsequent survey questions. To date, there is little research comparing the effects of concurrent and retrospective probing on probe responses (for a notable exception, see Fowler & Willis, 2020), nor in how far concurrent probing of related survey questions impacts responses to subsequent survey and probing questions. This study focuses on concurrent probing, where such question order effects are considered more likely to occur and endanger probe response quality.

The following sections apply two main perspectives of examining order effects to the context of cognitive pretesting. The first perspective focuses on the single probe and the occurrence of probe response content as a function of question order. The second analytical perspective examines the relation of responses to one another in the form of reported attitude–behavior consistency. The third section adds a cross-cultural perspective to question order effects in general and particularly web probing.

Probe Response Content

In its simplest form, testing for question order effects can be carried out by comparing responses as a function of question order. This main effect of question order on the response to the later shown question has been coined unconditional order effect (Rasinski et al., 2012). In contrast, conditional order effects do not automatically occur when a question is preceded by another question, but only when the respondent gives a certain answer to that preceding question. Thus, conditional order effects take into account a possible "interaction between question order and response to the antecedent question" (Smith, 1992, p. 164).

Imagine that a survey question on a behavior is asked, followed by a probe. This, in turn, is followed by a survey question on the attitude toward that behavior and another probe. The Gricean (1975) maxim of relation expects respondents to offer relevant information to these probes. For instance, a respondent is asked a behavior question on fare evasion, such as "Have you ever used public transport without having a valid ticket?" The respondent is asked to explain her answer in an open-ended probe. The answer to this probe should pertain to their behavior (i.e., "I have never avoided the fare and always buy a ticket"). A subsequent attitude question may ask how she feels toward people avoiding the fare. The response to a probe following this question should relate to her attitude (i.e., "avoiding the fare is absolutely wrong because other people pay the price").

The order of presenting the two survey and subsequent probing questions may impact probe responses. When respondents are first presented the behavior survey question and probe, they may elaborate on both their behavior and attitude. When respondents are subsequently presented the attitude survey question and probe, they might be less inclined to elaborate on their attitude a second time. Such an effect could be caused by another conversational maxim, known as the norm of

nonredundancy (Clark & Haviland, 1977), which requires information provided by each party to be new, not reiterating information the recipient already has (Schwarz, 1995). This situation describes a conditional order effect, in which the occurrence of probe response content is a function of both question sequence and whether this content has already been mentioned in answer to the first-shown probe. More precisely, in the abovementioned scenario, attitude-related probe response content should decrease when the behavior survey question is presented first and the respondent already mentioned their attitude in response to the anteceding probe.

However, researchers have also considered the notion of unconditional question order effects on probe responses, with pretest participants being inclined to "provide less information to later probes because they *believe* they have already provided relevant information" (Fowler & Willis, 2020, p. 463, italics by author). In this scenario, respondents would be less likely to mention content in answer to the second-shown probe regardless of which content they included in response to the first-shown probe. Such an effect would be particularly problematic if content indicating problems with a survey question are not uncovered, such as difficulties understanding a term used in the question, retrieving the information required, or finding a suitable answer category.

The applicability of the communicative principles underlying potential question order effects has not been tested in a web probing context. However, web probing implements the same or very similar probing techniques and sequence as cognitive interviewing, implicitly assuming similar communicative principles. Following this notion, the possibility of unconditional and conditional question order effects leads to two hypotheses on the occurrence of probe response content in answer to the second-shown probe in web probing:

> **Hypothesis 1:** When first one survey question and subsequent probe and then a second survey question and subsequent probe on a similar topic are presented, probe response content is less likely to be mentioned in answer to the second-shown probe (unconditional question order effect).
>
> **Hypothesis 2:** When first one survey question and subsequent probe and then a second survey question and subsequent probe on a similar topic are presented, probe response content is less likely to be mentioned in answer to the second-shown probe *if this content has already been mentioned in answer to a previous probe* (conditional question order effect).

Consistency of Probe Responses

Another important aspect when examining question order effects is how responses relate to one another. In studies examining general and specific questions, Schwarz et al. (1991) found that the correlation between marital and life satisfaction differed depending on question order when the two questions were assigned to the same conversational context. Studies examining the relation of two or more questions to one another have coined the term *associational* question order effect (Rasinski et al., 2012).

The relationship of attitude and behavior measures is often described in terms of consistency. Consistent responses are when attitudes and behavior correspond to each other, for instance when respondents hold a positive attitude toward behavior that they (intend to) engage in and negative attitudes toward behavior they do not (intend to) engage in. Reprising the example above, holding a lenient attitude toward fare evasion would be consistent for respondents who commit fare evasion.

Many factors influence the strength of relationship between behavior and attitude questions, such as strength of the attitude, visibility of the behavior, and psychological state of the respondent (Fazio et al., 1978; Liska, 1984). Also, respondents with past personal experience with a behavior are more likely to give consistent self-reports (Fazio & Zanna, 1981; Zanna et al., 1981).

Regarding question order, consistency between behavior and attitude self-reports increases when behavior-related questions are asked first (Budd, 1987). Respondents are likely to align their attitudinal responses with self-reports on behavior, especially when they establish a normative principle between two questions (Smith, 1992). An explanation for this effect is that attitudinal measurement instruments can prompt in situ attitude formation rather than assessing pre-existing attitudes. Preceding questions are a possible source of accessible in-formation for context-dependent questions (Bless & Schwarz, 2010). Respondents may base attitude judgments on information about their reported behavior when the survey context supports this (Schwarz & Bohner, 2001). Based on the notion of cognitive consistency (Heider, 1958), when attitudes are created or changed, this is mostly done in a way that is consistent with behavior, thus reducing cognitive dissonance (Festinger, 1957) and enhancing self-perception (Bem, 1967).

This leads to the following hypothesis concerning probe responses:

> **Hypothesis 3:** Consistency of probe responses is higher when the behavior question is asked first.

Cross-Cultural Question Order Effects in Web Probing

Much research on question order effects has examined one country only. However, even classic question order effects vary across countries (Stark et al., 2018). For instance, question order effects on the strength of relation between general and specific questions have been found across a range of topics such as marital and general satisfaction (Schwarz et al., 1991) and customer satisfaction (Schul & Schiff, 1993). A study on academic and general satisfaction reproduced the effect in a German sample but could not replicate it with Chinese participants (Haberstroh et al., 2002). In the aftermath, the applicability of the underlying conversational norms (Grice, 1975) to collectivist cultures has been questioned (Schwarz et al., 2010). In a comparative study involving 11 countries, Stark et al. (2018) find that both differences in survey response styles and topic-specific differences contribute to explaining differing question order effects across countries.

Topic-specific differences between countries may impact question order effects for attitude and behavior questions if, for instance, a certain behavior is more common or considered to be more acceptable in one country. Returning to the example of fare evasion, Germany and the United States have very different usage levels of public transport and dominant control systems for fare evasion (Buehler, 2011; Buehler & Pucher, 2012). In the United States, most public transport stations have paid areas that are physically secured in the form of ticket gates. In contrast, Germany relies on an honor-based proof-of-payment system in which passengers can board public transport without prior ticket control. Random ticket inspections and fines are used to enforce payment (Fürst & Herold, 2018). Research on public transport shows that revenue loss as a consequence of fare evasion is a financial risk in proof-of-payment systems, as boarding without a valid ticket is much easier, resulting in higher levels of fare evasion (Barabino et al., 2014). In the present case, cross-cultural differences in self-reported behavior of fare evasion may result in differences in the attitude–behavior relationship and impact question order effects.

Regardless of cross-cultural differences in survey responses, web probing experiments have repeatedly revealed differences in probe response quality even between postindustrial, individualist countries. Meitinger et al. (2019) found that American respondents generate higher levels of probe nonresponse, provide fewer themes, write shorter responses, and take less time to respond than German respondents. In another study, Meitinger et al. (2018) found that the sequence of asking several probes following one survey question impacted response quality, with respondents from different countries varying whether and in which way their response quality decreased. While U.S.

respondents were likely to react with probe nonresponse, German respondents were hardly affected by probe sequence. Thus, U.S. respondents generally show lower probe response quality than German respondents, and probe response quality seems to be impacted more strongly in the United States by contextual factors. However, these studies operationalized probe response quality using measures such as probe nonresponse or length of response, and it remains unclear how the prevalence of probe response content or consistency might differ between countries as a function of question order.

In summary, the effects of question order on probe response content and consistency are likely to differ across countries, both due to differences in response to the survey questions and due to differing probe response quality between countries. The present study defines cross-cultural differences as those differences between countries which cannot be explained by differences in survey response behavior. An undirected hypothesis is formulated to account for differences between countries:

> **Hypothesis 4:** Content and consistency of probe responses differ by country.

Procedure

Research Design

An experimental setup was chosen which randomized the order of a behavior and an attitude survey question and the subsequent probing questions in an online questionnaire. A general probe was used following each survey question. General probes ask respondents to explain their thoughts while answering; thus, the wording could be kept consistent across both survey questions. In the first experimental condition, respondents were first shown the behavior survey question (S_beh), directly followed by the probe (P_beh). They were then presented the attitude survey question (S_att) and subsequent probe (P_att). In the second condition, the order of the questions was reversed, with respondents first answering the attitude survey question (S_att) and probe (P_att), and thereafter the behavior survey question (S_beh) and probe (P_beh). Figure 1 illustrates the two experimental conditions.

The probing question was presented on a separate screen that repeated the question text as well as the respondent's survey response and asked them to state what they had thought about while answering the question in an open text field. The topic examined was fare evasion. The behavior question was drawn from the German General Social Survey (ALLBUS, 2000) and comprised four delinquent behaviors. Respondents were asked how often they had avoided the fare in the past. They could choose between six frequencies ranging from "never" to "more than 20 times" and a "don't remember" option. A general evaluative question on a range of behaviors from World Values Survey Wave 6 (2014) was chosen for the attitude question. Item sequence was adjusted to show fare evasion as the first topic. Respondents were asked to indicate on a 10-point scale in how far they felt that fare evasion could be justified. (Table A1 in the Online Appendix shows the wording of the survey and probing questions.)

A web probing study with respondents from Germany and the United States was conducted between July 25 and August 7, 2018. The main panel provider was Respondi AG, based in Germany, who cooperated with an international partner to recruit the U.S. sample. Equal quotas for gender, age (18–29, 30–49, and 50–64 years), and education (lower and higher) were used. Of the 1,947 panelists who responded to the survey invitation, 1,248 were screened out and 400 (Germany: $n = 192$; United States: $n = 208$) completed the survey, resulting in a completion rate of 57% (Callegaro & DiSogra, 2008). The web probing study contained a range of experiments. The questions analyzed in this study were from the middle of the survey and were shown directly after each other. Respondents
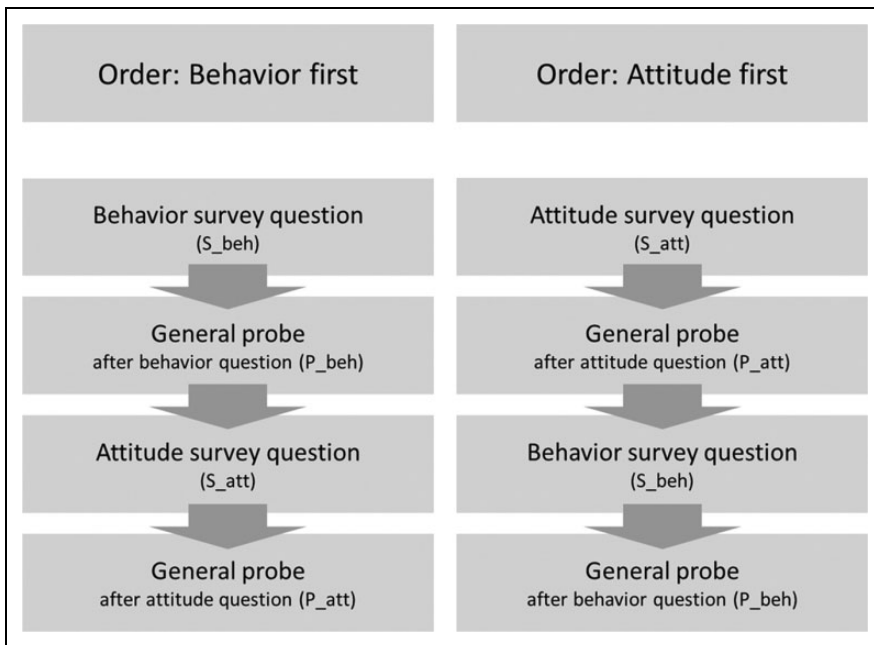
Figure 1. Experimental design.

who did not answer any of the open-ended probing questions in the course of the web survey (i.e., neither in the reported experiment nor in any part of the survey) were excluded from the sample, resulting in a final sample of $n = 333$ respondents (Germany: $n = 167$; United States: $n = 166$). A χ^2 test of independence revealed no significant association between the experimental condition and gender, age, or education in either country (Table A2 in the Online Appendix). The average survey completion time of the total survey was 27.6 min in Germany and 27.4 min in the United States for the final sample. Respondents were paid 2.00 Euros as an incentive for survey completion.

Coding Scheme

In order to quantify probe response content and consistency, probe responses were coded using three coding schemes. These codes served as dependent variables in the analyses. The first two coding schemes relate to probe response content (Hypotheses 1 and 2), the third to consistency (Hypothesis 3). The first scheme indicated whether and what type of behavior- and attitude-related content was contained in the probes. The second scheme was problem-based and indicated whether respondents reported a questionnaire understanding problem with the behavior survey question. The third scheme indicated whether responses to the two probes were consistent to one another.

Coding Scheme 1 was created by the author. After the initial coding scheme had been developed, the author and a research assistant coded one of the probes. Following an evaluation of differences, the scheme was refined to comprise more elaborate categorization rules. Then, both coders coded all probe responses. Both probes were coded once for behavior- and attitude-related content. Cohen's κ was strong, with values between .823 and .916 (Table A3 in the Online Appendix). Differences in coding were discussed and the final codes assigned mutually. Table 1 gives an overview of all codes.

Table 1. Coding Schemes.

Coding Scheme	Topic of Coding	Value	Value Label	Description and Examples
Coding Scheme 1: Behavior- and attitude-related content	Behavior	1	Explicitly admit	Explicit mention of avoiding the fare at least once in the past (deliberate or unintentional, also avoiding part of the fare): "The times that I got on without paying" "Once, when I had no money"
		2	Does not explicitly admit	Clear statement that respondent has never avoided the fare, neither deliberately nor by accident "I never did this" "I rarely use public transport, but when I did I always paid" Response refers to behavior, but remains unclear as to the action "I thought about what I did"
		0	No mention of behavior	Probe response does not mention the respondent's past or present personal behavior
	Attitude	1	Absolute condemnation	Respondent does not approve of behavior, indicating that there is no justification "If you can't pay, don't use public transport" "Avoiding the fare is stealing"
		2	More lenient attitude	Respondent finds the behavior justifiable/explainable, at least under certain conditions "It depends on . . ." "Sometimes you forget or don't have enough money"
		0	No mention of attitude	Probe response does not give any evaluative judgment on fare evasion
Coding Scheme 2: Problem-related content	Problem "missing answer category"	1	Named	Respondent clearly states that they have never used public transport before and are missing an answer option to indicate this "I've never had to take public transportation"
		0	Not named	Respondent does not (clearly) indicate that they have never used public transport before

(continued)

Table 1. (continued)

Coding Scheme	Topic of Coding	Value	Value Label	Description and Examples
Coding Scheme 3: Consistency of probe responses	Consistency	1	Consistent	Respondent admits to having avoided the fare in answer to P_beh (Theme Code 1, *behavior* = 1) and displays a more lenient attitude toward the topic of fare evasion in answer to P_att (Theme Code 1, *attitude* = 2) OR Respondent does not explicitly admit to having avoided the fare in the past (Theme Code 1, *behavior* = 2) in answer to P_beh and displays absolute condemnation toward this behavior in P_att (Theme Code 1, *attitude* = 1)
		2	Inconsistent	Respondent admits to having avoided the fare in answer to P_beh (Theme Code 1, *behavior* = 1) and displays absolute condemnation toward this behavior in P_att (Theme Code 1, *attitude* = 1) OR Respondent does not explicitly admit to having avoided the fare in the past in P_beh (Theme Code 1, *behavior* = 2) and displays a more lenient attitude toward the topic of fare evasion in P_att (Theme Code 1, *attitude* = 2)
		3	Unascertainable	Respondent does not relate to their behavior in response to P_beh OR does not relate to their attitude in response to P_att

Note. Coding Scheme 1 was carried out for each probe separately; Coding Scheme 2 was applied to the probe following the behavior question (P_beh).

Coding Scheme 1: Behavior- and attitude-related content. For behavior-related probe response content, the code indicated whether a respondent explicitly admitted to having avoided the fare in the past (1 = *admit*) or not (2 = *do not admit*), either by explicitly negating or remaining vague about their past behavior. If the probe response made no reference to the respondent's past or present personal behavior, it was coded as 0 = *no mention of behavior*. For attitude-related probe response content, the code distinguished between absolute condemnation (1 = *absolute condemnation*) and a more lenient attitude (2 = *lenient attitude*). If the probe response did not contain any information on the respondent's personal attitude toward fare evasion, it was coded as 0 = *no mention of attitude*. Responses that contained neither behavior- nor attitude-related content were coded as nonsubstantive. Answers in this category included nonresponse, single characters, off-topic remarks, and other noncodable content.

The codes from Coding Scheme 1 were further recoded into binary variables to mark the presence of behavior- and attitude-related probe response content. One code indicated whether a response mentioned the respondent's past or present personal behavior (1 = *yes*, *no*) and a second code whether it mentioned the respondent's attitude in any form (1 = *yes*, 0 = *no*).

Coding Scheme 2: Problem-related content. The second coding perspective focused on problems respondents reported while answering the survey questions and pointed to issues of question design. Problems were classified along the cognitive process of survey response (Tourangeau et al., 2000) for each survey question separately. Problems included misunderstanding of the term public transport (i.e., to include airplanes or taxis) and the concept of fare evasion (i.e., whether or not to include unintentional fare evasion).[1] The present analysis includes the only problem that emerged with sufficient size for quantitative analysis. This issue pertained to the behavior survey question (S_beh) and pointed to a missing answer category to indicate that a respondent had never used public transport before (1 = *'answer category missing' mentioned*, 0 = *not mentioned*). This response option is crucial to distinguish between honest customers and people to whom the question does not apply.

Coding Scheme 3: Consistency of probe responses. The third coding scheme coded probe responses as consistent (1 = *consistent*) if the respondent (a) admitted to having avoided the fare in the past and displayed a lenient attitude toward this behavior or (b) did not admit to the behavior and reported absolute condemnation. Probe responses were coded as inconsistent (2 = *inconsistent*) if the respondent (c) admitted to the behavior but displayed absolute condemnation as an attitude or (d) did not admit to the behavior but displayed a lenient attitude. If probe responses did not contain both behavior- and attitude-related content, consistency was coded as unascertainable (3 = *unascertainable*).

Data Analysis

Binary logistic regression models were carried out to examine the occurrence of probe response content and a multinomial logistic regression to examine the consistency of responses to one another. The dependent variables regarding the occurrence of probe response content were the binary behavior and attitude variables from Coding Scheme 1 and the problem-related content from Coding Scheme 2. The dependent variable for the consistency of probe responses was Coding Scheme 3.

Question order (1 = *behavior question first*, 0 = *attitude question first*) was used as a main predictor to test Hypotheses 1 and 3. To test whether previously mentioned probe response content is less likely to be mentioned in response to the second-shown probe (Hypothesis 2), two dummy variables were created that signified that the respective content had already been mentioned in

Table 2. Response Distributions for Survey and Probing Questions.

	Germany		United States	
	Behavior Question First % (n)	Attitude Question First % (n)	Behavior Question First % (n)	Attitude Question First % (n)
Behavior survey question (S_beh)				
Admit to fare evasion	54 (43)	51 (44)	16 (11)	15 (14)
Do not admit	46 (37)	49 (43)	84 (59)	85 (82)
Attitude survey question (S_att)				
Absolute condemnation	40 (32)	61 (53)	37 (26)	36 (35)
More lenient attitude	60 (48)	39 (34)	63 (44)	64 (61)
Probe following behavior question (P_beh)				
Behavior-related content (Coding Scheme 1)				
Admit	45 (36)	38 (33)	10 (7)	8 (8)
Do not admit	13 (10)	26 (23)	57 (40)	51 (49)
No behavior-related content	43 (34)	36 (31)	33 (23)	41 (39)
Attitude-related content (Coding Scheme 1)				
Absolute condemnation	5 (4)	9 (8)	3 (2)	5 (5)
More lenient attitude	34 (27)	28 (24)	7 (5)	3 (3)
No attitude-related content	61 (49)	63 (55)	90 (63)	92 (88)
Problem-related content (Coding Scheme 2)				
"Answer category missing" mentioned	0 (0)	0 (0)	33 (23)	17 (16)
Not mentioned	100 (80)	100 (87)	67 (47)	83 (80)
Probe following attitude question (P_att)				
Attitude-related content (Coding Scheme 1)				
Absolute condemnation	26 (21)	48 (42)	30 (21)	30 (29)
More lenient attitude	45 (36)	26 (23)	43 (30)	40 (38)
No attitude-related content	29 (23)	25 (22)	27 (19)	30 (29)
Behavior-related content (Coding Scheme 1)				
Admit	4 (3)	6 (5)	3 (2)	1 (1)
Do not admit	3 (2)	3 (3)	7 (5)	3 (3)
No behavior-related content	94 (75)	91 (79)	90 (63)	96 (92)

answer to the previously shown probe ($1 = $ *probe response content mentioned previously*, $0 = $ *probe response content not mentioned previously*). One dummy variable indicated this for behavior-related, the other for attitude-related content. As problem-related content was only coded for the behavior survey question, no dummy variable was created for the previous occurrence of problem-related probe response content. Country was inserted as a main predictor ($1 = $ *Germany*, $0 = $ *United States*) to examine cross-cultural differences (Hypothesis 4).

As past research has demonstrated the impact of survey response on responses to open-ended questions (Zuell & Scholz, 2015) and question order effects may be impacted by survey response behavior (Stark et al., 2018), models controlled for the response to the preceding survey question. To this end, survey responses were recoded into binary variables using the same logic as in Coding Scheme 1 (behavior question S_beh: $1 = $ *admit*, $0 = $ *do not admit*; attitude question S_att: $1 = $ *lenient attitude*, $0 = $ *absolute condemnation*). Gender ($1 = $ *women*, $0 = $ *men*) and age were

included as covariates as fare evasion has been associated with young men (Cools et al., 2018). Past studies have linked the strength of question order effects to lower education (Narayan & Krosnick, 1996), though this has been disputed in more recent studies (Stark et al., 2018; education: $1 = low$, $0 = high$). There is no previous research indicating whether the tendency toward social desirability responding impacts the likelihood of responding to probing questions. As the topic of fare evasion is potentially sensitive, the short scale for social desirability responding (KSE-G; Kemper et al., 2014) with the two dimensions "exaggerating positive traits" and "minimizing negative traits" were included as metric covariates in all models.

Prior to analysis, the prerequisites of logistic regression were tested. The multicollinearity diagnostic for metric and dichotomous predictors revealed good results, with all variance inflation factors (VIFs) slightly above 1. All models had few outliers. Data analysis was carried out using SPSS Version 24.

Results

The responses to the two survey questions confirm the reported difference in the prevalence of the fare evasion between Germany and the United States. While 52% of German respondents ($n = 87$) admitted to fare evasion, this was only the case for 15% of American respondents ($n = 25$). Interestingly, 61% ($n = 53$) of German respondents displayed absolute condemnation of fare evasion when they were first asked about their attitude; this value decreased to 40% ($n = 32$) when they were first asked about their personal fare evasion behavior. American respondents generally showed a more lenient attitude toward fare evasion (63%; $n = 105$).

Probe Response Content

The first focus of the analysis was the occurrence of probe response content. To examine this, probe responses were coded to indicate behavior-, attitude- and problem-related content. Table 2 shows the occurrence of probe response content and responses to the survey questions.

A clear majority of respondents mentioned their personal behavior in answer to the probe following the behavior question (P_beh: 62%; $n = 206$) and their attitude in response to the probe following the attitude question (P_att: 72%; $n = 240$). In no cases did respondents' probing answers contradict their survey responses (i.e., no respondent claimed to have committed fare evasion in answer to the survey question, but not the probing question or the other way around). Problem-related content was mentioned in answer to the probe following the behavior question and only by U.S. respondents (P_beh: 12%; $n = 39$).

Hypothesis 1 predicted that content was less likely to be mentioned in response to a probe when the respective probe was shown second. Hypothesis 2 specified that this would be the case when the respective content had already been mentioned in response to the first-shown probe. Hypothesis 4 predicted differences between countries regarding the occurrence of content. These hypotheses were tested using binary logistic regression models for behavior-related content to P_beh (Model 1), attitude-related content to P_att (Model 2), and problem-related content to P_beh (Model 3). To test Hypothesis 1, question order served as a main predictor. To test Hypothesis 2, the dummy variables on previous mention of probe response content were the main predictors for Models 1 and 2. To test Hypothesis 4, country was included as a main predictor in Models 1 and 2. Covariates were included as described under Data Analysis. The results of the binary logistic regressions are shown in Table 3.

Model 1 showed significant effects of response to the behavior question and gender on the occurrence of behavior-related probe response content in answer to the probe following the behavior question. Respondents who had committed fare evasion in the past were more likely to offer behavior-related probe response content, as were women. In contrast, respondents who reported

Table 3. Binary Logistic Regressions for Occurrence of Probe Response Content.

	(1) Behavior-Related Content			(2) Attitude-Related Content			(3) Problem-Related Content		
	B	SE	OR	B	SE	OR	B	SE	OR
Question order (1 = *behavior first*)	−.02	.24	0.98	−0.34	.27	0.71	0.82*	.40	2.27
Previous mention of content									
Behavior mentioned previously (1 = *yes*)	.72	.70	2.05	—	—	—	—	—	—
Attitude mentioned previously (1 = *yes*)	—	—	—	1.39*	.58	4.03	—	—	—
Country (1 = *Germany*)	−0.21	0.26	0.81	0.05	0.28	1.05	—	—	—
Survey response									
Behavior question (S_beh; 1 = *admit*)	.81**	.28	2.24	—	—	—	—	—	—
Attitude question (S_att; 1 = *lenient attitude*)	—	—	—	0.13	.27	1.14	—	—	—
Gender (1 = *women*)	.53*	.24	1.70	0.70**	.26	2.02	1.12*	.44	3.07
Age	.05	.04	1.05	0.07	.05	1.08	0.15*	.07	1.16
Education (1 = *low*)	−.30	.24	0.74	−0.39	.26	0.68	0.04	.40	1.04
Exaggerating positive qualities	.01	.13	1.01	−.04	.14	0.96	−0.11	.21	0.90
Minimizing negative qualities	−.28	.16	0.75	−.04	.17	0.96	−0.50	.35	0.60
Constant	.42	.67	1.52	0.52	.74	1.69	−2.07	1.11	0.13
Model parameters	$\chi^2 = 21.05$, $df = 9$, $p = .012$, $N = 333$, $R^2_{(Nagelkerke)} = .08$			$\chi^2 = 21.36$, $df = 9$, $p = .011$, $N = 333$, $R^2_{(Nagelkerke)} = .09$			$\chi^2 = 21.53$, $df = 6$, $p = .001$, $N = 166$, $R^2_{(Nagelkerke)} = .18$		

Note. B = logit coefficient; *SE* = standard error; *OR* = odds ratio.
*$p < .05$. **$p < .01$. ***$p < .001$.

that they had never avoided the fare and men were more likely to insert nonsubstantive responses (such as "fare evasion" or "nothing") or not respond to the probe at all. Contrary to Hypotheses 1 and 2, neither question order nor previous mention of behavior-related content influenced the occurrence of behavior-related content. However, the data basis for testing Hypothesis 2 was small, as only few respondents mentioned their personal behavior in answer to the probe following the preceding attitude question (see Table 2).

Model 2 showed a significant effect of having previously mentioned attitude-related probe response content ($B = 1.39$, standard error [SE] $= .58$, odds ratio [OR] $= 4.03$, $p < .05$) on the likelihood of doing so in answer to the probe following the attitude question. However, the direction of the effect was contrary to Hypothesis 2, with respondents who had previously volunteered attitude-related content being *more* likely to do this a second time. For instance, one respondent wrote in answer to the probe following the behavior question "It's not ok to not pay" and gave a very similar response in answer to the probe following the attitude question ("not paying for the bus is like stealing"). Further, there was an unexpected difference between countries regarding the independent variable. While almost 40% of German respondents offered attitude-related content in the probe response following the behavior question (P_beh), only 10% of U.S. respondents did this (see Table 2). In other words, the data basis for examining Hypothesis 2 was good for German respondents but again rather weak in the case of U.S. respondents. Gender of the respondent showed a significant effect in the model, with women more likely to mention their attitude than men. Question order had no significant effect. Thus, neither Hypothesis 1 nor Hypothesis 2 is supported for attitude-related content.

Model 3 examined the likelihood of respondents indicating a problem responding to the behavior survey question, namely, that a suitable response category was missing to indicate that they had never used public transport. When the behavior question was shown first, 33% ($n = 23$) of U.S. respondents mentioned this issue in answer to the probe; when the attitude question was shown first, the rate sunk to 17% ($n = 16$; see Table 2). Several independent variables had to be omitted from this model. First, the problem-related content referred to the behavior survey question and was only coded for the probe following this question (P_beh). Thus, the dummy variable regarding the previous mention of the same content was not coded, and Hypothesis 2 could not be tested in this model. Second, the problem code was only detected in the U.S. sample, certainly due to the stronger use of public transport in Germany. Thus, while the data demonstrate differences between countries, the analysis was only carried out using the U.S. sample and the variable country was omitted from the list of predictors. Finally, the problem was only coded for respondents who had answered the behavior survey question (S_beh) with "never." The variable response to the survey question (S_beh) was excluded from the model as it showed no variance.

Despite the low case number regarding the dependent variable and sample, Model 3 showed a good fit. It also showed significant effects of question order, gender, and age on the likelihood of reporting the problem. Respondents were more likely to mention this problem with the question when the behavior question was asked first ($B = 0.82$, $SE = .40$, $OR = 2.27$, $p < .05$), supporting Hypothesis 1. Women and older respondents were more likely to report the problem.

In summary, question order impacted the occurrence of probe response content for Model 3 only, lending limited support for Hypothesis 1. Hypothesis 2 could not be confirmed. Indeed, for attitude-related probe response content (Model 2), the likelihood of mentioning this content even increased among respondents who had already done so previously. Hypothesis 4 was partially confirmed. While the occurrence of behavior- and attitude-related content was independent of country, the problem-related content was specific to the U.S. context.

Table 4. Consistency of Probe Responses.

	Germany		United States	
	Behavior Question First	Attitude Question First	Behavior Question First	Attitude Question First
Probe consistency	% (n)	% (n)	% (n)	% (n)
Consistent	33 (26)	38 (33)	29 (20)	25 (24)
Inconsistent	14 (11)	15 (13)	19 (13)	24 (23)
Unascertainable	54 (43)	47 (41)	53 (37)	51 (49)

Consistency of Probe Responses

The second analytical focus was the consistency of probe responses to one another. Table 4 gives an overview of the consistency of probe responses as a function of question order. About one third (31%; $n = 103$) of probe responses were consistent to each other, with respondents either admitting to fare evasion ("I have done this on occasion") and displaying a lenient attitude (i.e., "It won't destroy the bus company") or denying having ever avoided the fare (i.e., "I always pay!") and showing a harsh attitude ("That's stealing"); 18% of responses were inconsistent ($n = 60$), with some respondents showing an awareness that their behavior and attitude did not match. One respondent wrote in answer to the probe following the behavior question: "That's about how many times I did this. Yeah, I'm a hypocrite and imperfect. I believe it's wrong to do this." The other 51% ($n = 170$) did not contain both behavior- and attitude-related content; thus, these respondents were coded as unascertainable. Consistent probe responses were slightly more frequent in Germany (35%) than in the United States (27%); however, in both countries, about half of all probe responses were unascertainable.

Hypothesis 3 predicted that the consistency of probe responses is higher when the behavior question is asked first. Hypothesis 4 predicted differences between countries regarding the consistency of probe responses. A multinomial logistic regression was carried out with *consistent* probe responses as the reference category. Main predictors were question order and country. Covariates were included as described under Data Analysis. Results are shown in Table 5. (Distributions of all predictors and covariates can be found in Table A4 in the Online Appendix.)

Regarding the comparison of consistent to inconsistent probe responses, there was a significant effect of country ($B = 0.80$, $SE = .39$, $OR = 0.04$, $p < .05$). U.S. respondents were more likely to give inconsistent than consistent answers than German respondents. For the comparison of consistent to unascertainable probe responses, there were significant effects of the response to the behavior survey question, gender, and age. Respondents admitting to the offense were more likely to give consistent and less likely to give unascertainable probe responses. Also, women and older respondents more likely to give consistent versus unascertainable probe responses. There was no significant effect of question order.

Hypothesis 3 could not be confirmed as there was no significant effect of question order on probe consistency. Hypothesis 4 could be confirmed, as there was a significant difference in probe consistency between countries, with German respondent more likely to give consistent responses than U.S. respondents.

Discussion and Conclusion

The present research set out to explore whether and how question order impacts probe responses to behavior and attitude questions in cross-cultural web probing. This was done by examining the

Table 5. Multinomial Logistic Regression for Consistency of Probe Responses.

	Inconsistent			Unascertainable		
Reference Category: Consistent Probe Responses	B	SE	p	B	SE	p
Question order (1 = behavior first)	.19	.34	.58	−.22	.26	.41
Country (1 = Germany)	.80*	.39	.04	.17	.29	.56
Survey response to behavior question (S_beh; 1 = admit)	−.23	.40	.56	.65*	.31	.03
Gender (1 = women)	−.21	.35	.55	.53*	.27	.04
Age						
1 = 18–29 years	.52	.42	.22	.78*	.33	.02
2 = 30–49 years	.73	.41	.07	.95***	.32	.00
3 = 50–65 years	—	—	—	—	—	—
Education (1 = low)	.10	.34	.78	−.31	.26	.24
Exaggerating positive qualities	.14	.19	.46	−.02	.15	.92
Minimizing negative qualities	−.01	.23	.97	.13	.18	.46
Constant	−1.73	0.99	0.08	−0.71	0.76	0.35
Model parameters	$\chi^2 = 34.332$, $df = 18$, $p = .011$, $N = 333$, $R^2_{(Nagelkerke)} = .113$					

Note. B = logit coefficient; SE = standard error.
*$p < .05$. **$p < .01$. ***$p < .001$.

impact of question order on the occurrence of probe content and the relation of the probe responses to one another in terms of self-reported attitude-behavior consistency in two countries. Content and consistency of probe responses were not strongly impacted by question order in the present study.

There was limited support for the first hypothesis. Question order did not impact the occurrence of broad themes such as mentioning one's behavior or attitude. The problem of the missing answer category for the behavior question was more likely to be coded when this question was shown first. However, the case numbers for this model were rather low as the problem was restricted to the U.S. context. Further research is recommended to examine the occurrence of problem-related probe response content.

The second hypothesis predicted that respondents would be less likely to mention probe response content if they had already mentioned this content in answer to a previous probe. This hypothesis could not be confirmed. The occurrence of behavior-related probe response content was irrespective of whether respondents had previously mentioned their behavior. The occurrence of attitude-related content even increased when respondents had mentioned their attitude previously. Thus, there is no support that the norm of nonredundancy applies in the context of web probing. Possibly, asking several questions on the topic of fare evasion made respondents' related attitudes more salient (for saliency as a question order effect, see Bradburn, 1983). However, most respondents refrained from mentioning their attitude in response to the behavior question and vice versa. More generally, the applicability of conversational norms in a self-administered web context must be questioned.

Contrary to the third hypothesis, consistency of probe responses was not affected by question order. It was, however, significantly impacted by the response to the behavior survey question. Respondents who did not admit to the behavior were less likely to give substantive probe responses in the subsequent probing question, confirming previous results that the response to survey questions determines the likelihood of giving substantive probe responses (Zuell & Scholz, 2015).

In line with the fourth hypothesis, significant differences in probe response content and consistency were found between countries. The results demonstrated both differences in survey response behavior between countries and differences that cannot be explained by country-specific survey response behavior. There was no significant impact of country on the likelihood of including

behavior- or attitude-related content; however, problem-related probe content only emerged in the U.S. sample. Second, German respondents were more likely to demonstrate attitude–behavior consistency in their probe responses, probably because German respondents were more likely to admit to fare evasion and align their attitude to be consistent with their self-reported behavior. Finally, U.S. respondents were far less likely than German respondents to mention both their behavior and attitude in one probe response. This result is in line with previous research that U.S. respondents generally offer less variety in themes (Meitinger et al., 2019). However, this finding can also be explained by the topic of public transport being less relevant to U.S. respondents. These findings underscore the importance of validating questionnaires in the language and country they are to be fielded in (see also Willis, 2015b). In summary, both predicting and explaining cross-cultural differences remain challenging, but including multiple countries in research designs is all the more crucial to make valid inferences about question order effects and to advance theorizing.

Finally, the results highlight that content-based coding makes the effects of question order and further contextual factors visible in ways that cannot be seen using standard measures of probe response quality, such as length of response or share of nonsubstantive answers. Future research in the area of web probing should combine quantitative and qualitative forms of analysis whenever possible.

The results are limited by several factors which light the way for future research. For one, the study examined question order effects by testing a behavior and attitude question. This is certainly a highly relevant case for questionnaire designers and a common setup in cognitive pretesting. However, examining different types of survey questions may help uncover the underlying mechanisms of question order effects in pretesting. For instance, two behavior or two attitude survey questions on the same topic might prove to be a better setting to examine the effects of redundancy and saliency.

Second, the current study employed only one probing technique. The use of general probes made it possible to use the same wording in both probing questions despite having different types of survey questions. However, other probing techniques, such as category selection probes or more specific probes, may lead to other effects (Demaio & Landreth, 2004). Also, potential differences between concurrent probing, as employed in this study, and retrospective probing require further research.

Third, while web probing offers the ideal setup to quantify findings, the effects found must also be examined in the context of face-to-face cognitive interviewing in order to make inferences about question order effects on cognitive pretesting in general. The application of communication principles may be more pronounced in interviewer-administered modes.

Finally, the present study was carried out in two Western, postindustrial countries. Systematic research on cross-cultural differences in pretesting is desirable using a larger sample of countries (Meitinger et al., 2019; Pan et al., 2010; Park et al., 2014) with different communicative principles (Grice, 1975).

Concluding, in the present study, probe response content and the consistency of probe responses were mostly not impacted by question order. This is good news regarding the stability and validity of cognitive pretest findings from web probing studies. At the same time, cross-cultural differences may impact the content and consistency of probe responses. In light of ever more complex survey and pretest designs, further methodological research is necessary to ensure that pretest results remain valid, reliable, and contribute to preventing measurement error in cross-cultural settings.

Data Availability

The quantitative data set of this study is available on request from the author. The answers to the open-ended questions are not publicly available due to them containing information that could compromise participant privacy.

Declaration of Conflicting Interests

Funding

Software Information

All analyses in the present study were conducted using IBM SPSS Statistics Version 24.0.

Supplemental Material

Supplemental material for this article is available online.

Note

1. The coding scheme is available from the author on request.

References

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Springer series in social psychology. Action control: From cognition to behavior* (pp. 11–39). Springer.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888–918. https://doi.org/10.1037/0033-2909.84.5.888

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice Hall.

ALLBUS. (2000). *Allgemeine Bevölkerungsumfrage der Sozialwissenschaften. ZA3452*. https://www.gesis.org/en/allbus/contents-search/questionnaires/

Barabino, B., Salis, S., & Useli, B. (2014). Fare evasion in proof-of-payment transit systems: Deriving the optimum inspection level. *Transportation Research Part B: Methodological*, *70*, 1–17. https://doi.org/10.1016/j.trb.2014.08.001

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

Behr, D., Braun, M., Kaczmirek, L., & Bandilla, W. (2012). Testing the validity of gender ideology items by implementing probing questions in web surveys. *Field Methods*, *25*(2), 124–141. https://doi.org/10.1177/1525822X12462525

Behr, D., Kaczmirek, L., Bandilla, W., & Braun, M. (2012). Asking probing questions in web surveys: Which factors have an impact on the quality of responses? *Social Science Computer Review*, *30*(4), 487–498. https://doi.org/10.1177/0894439311435305

Behr, D., Meitinger, K., Braun, M., & Kaczmirek, L. (2017). *Web probing—Implementing probing techniques from cognitive interviewing in web surveys with the goal to assess the validity of survey questions. (GESIS survey guidelines)*. GESIS. https://doi.org/10.15465/gesis-sg_en_023

Behr, D., Meitinger, K., Braun, M., & Kaczmirek, L. (2020). Cross-national web probing: An overview of its methodology and its use in cross-national studies. In P. C. Beatty, D. Collins, L. Kaye, J. -L. Padilla, G. B. Willis, & A. Wilmot (Eds.), *Advances in questionnaire design, development, evaluation and testing* (pp. 512–543). Wiley.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183–200. https://doi.org/10.1037/h0024835

Bishop, G. F. (1992). Qualitative analysis of question-order and context effects: The use of think-aloud responses. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 149–162). Springer.

Bishop, G. F., Oldendick, R. W., & Tuchfarber, A. J. (1985). The importance of replicating a failure to replicate: Order effects on abortion items. *Public Opinion Quarterly*, *49*(1), 105–114. https://doi.org/10.1086/268904

Bless, H., & Schwarz, N. (2010). Mental construal and the emergence of assimilation and contrast effects: The inclusion/exclusion model. *Advances in Experimental Social Psychology*, *42*, 319–373. https://doi.org/10.1016/S0065-2601(10)42006-7

Bradburn, N. M. (1983). Response effects. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 289–328). Academic Press.

Braun, M., Behr, D., Kaczmirek, L., & Bandilla, W. (2014). Evaluating cross-national item equivalence with probing questions in web surveys. In U. Engel, B. Jann, P. Lynn, A. C. Scherpenzeel, & P. Sturgis (Eds.), *European Association of Methodology. Improving survey methods: Lessons from recent research* (pp. 184–200). Routledge Taylor & Francis Group.

Budd, R. J. (1987). Response bias and the theory of reasoned action. *Social Cognition*, *5*(2), 95–107.

Buehler, R. (2011). Determinants of transport mode choice: A comparison of Germany and the USA. *Journal of Transport Geography*, *19*(4), 644–657. https://doi.org/10.1016/j.jtrangeo.2010.07.005

Buehler, R., & Pucher, J. (2012). Demand for public transport in Germany and the USA: An analysis of rider characteristics. *Transport Reviews*, *32*(5), 541–567. https://doi.org/10.1080/01441647.2012.707695

Callegaro, M., & DiSogra, C. (2008). Computing response metrics for online panels. *Public Opinion Quarterly*, *72*(5), 1008–1032. https://doi.org/10.1093/poq/nfn065

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Erlbaum.

Collins, D. (2015). *Cognitive interviewing practice*. Sage.

Conrad, F. G., Schober, M. F., & Schwarz, N. (2014). Pragmatic processes in survey interviewing. In T. M. Holtgraves (Ed.), *Oxford library of psychology. The Oxford handbook of language and social psychology* (pp. 420–437). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199838639.013.005

Cools, M., Fabbro, Y., & Bellemans, T. (2018). Identification of the determinants of fare evasion. *Case Studies on Transport Policy*, *6*(3), 348–352. https://doi.org/10.1016/j.cstp.2017.10.007

Demaio, T. J., & Landreth, A. (2004). Do different cognitive interview techniques produce different results? In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Wiley series in survey methodology. Methods for testing and evaluating survey questionnaires* (pp. 89–108). Wiley-Interscience.

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. *Advances in Experimental Social Psychology*, *14*, 161–202. https://doi.org/10.1016/S0065-2601(08)60372-X

Fazio, R. H., Zanna, M. P., & Cooper, J. (1978). Direct experience and attitude-behavior consistency: An information processing analysis. *Personality and Social Psychology Bulletin*, *4*(1), 48–51. https://doi.org/10.1177/014616727800400109

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fowler, S., & Willis, G. B. (2020). The practice of cognitive interviewing through web probing. In P. C. Beatty, D. Collins, L. Kaye, J. -L. Padilla, G. B. Willis, & A. Wilmot (Eds.), *Advances in questionnaire design, development, evaluation and testing* (pp. 451–469). Wiley.

Fürst, E., & Herold, D. (2018). Fare evasion and ticket forgery in public transport: Insights from Germany, Austria and Switzerland. *Societies*, *8*(4), 98. https://doi.org/10.3390/soc8040098

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgen (Eds.), *Syntax and semantics: Volume 3: Speech acts* (5th ed., pp. 41–58). Academic Press.

Haberstroh, S., Oyserman, D., Schwarz, N., Kühnen, U., & Ji, L. -J. (2002). Is the interdependent self more sensitive to question context than the independent self? Self-construal and the observation of conversational norms. *Journal of Experimental Social Psychology*, *38*(3), 323–329. https://doi.org/10.1006/jesp.2001.1513

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley & Sons.

Kemper, C. J., Beierlein, C., Bensch, D., Kovaleva, A., & Rammstedt, B. (2014). Soziale Erwünschtheit-Gamma (KSE-G). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/zis186

Lenzner, T., & Neuert, C. (2017). Pretesting survey questions via web probing—Does it produce similar results to face-to-face cognitive interviewing? *Survey Practice 10*(4), 1–11. http://surveypractice.org/index.php/SurveyPractice/article/view/359

Lenzner, T., Neuert, C., & Otto, W. (2016). *Cognitive pretesting. (GESIS survey guidelines)*. GESIS. https://doi.org/10.15465/gesis-sg_en_010

Liska, A. E. (1984). A critical examination of the causal structure of the Fishbein/Ajzen attitude-behavior model. *Social Psychology Quarterly*, *47*(1), 61–74. https://doi.org/10.2307/3033889

Meitinger, K., & Behr, D. (2016). Comparing cognitive interviewing and online probing: Do they find similar results? *Field Methods*, *28*(4), 363–380. https://doi.org/10.1177/1525822X15625866

Meitinger, K., Behr, D., & Braun, M. (2019). Using apples and oranges to judge quality? Selection of appropriate cross-national indicators of response quality in open-ended questions. *Social Science Computer Review*, *27*, 1–22. https://doi.org/10.1177/0894439319859848

Meitinger, K., Braun, M., & Behr, D. (2018). Sequence matters in online probing: The impact of the order of probes on response quality, motivation of respondents, and answer content. *Survey Research Methods*, *12*(2), 103–120. https://doi.org/10.18148/SRM/2018.V12I2.7219

Narayan, S., & Krosnick, J. A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly*, *60*(1), 58. https://doi.org/10.1086/297739

Pan, Y., Landreth, A., Park, H., Hinsdale-Shouse, M., & Schoua-Glusberg, A. (2010). Cognitive interviewing in non-English languages: A cross-cultural perspective. In J. A. Harness (Ed.), *Wiley series in survey methodology. Survey methods in multinational, multiregional and multicultural contexts* (pp. 91–113). Wiley & Sons.

Park, H., Sha, M. M., & Pan, Y. (2014). Investigating validity and effectiveness of cognitive interviewing as a pretesting method for non-English questionnaires: Findings from Korean cognitive interviews. *International Journal of Social Research Methodology*, *17*(6), 643–658. https://doi.org/10.1080/13645579.2013.823002

Presser, S., & Blair, J. (1994). Survey pretesting: Do different methods produce different results? *Sociological Methodology*, *24*, 73–104. https://doi.org/10.2307/270979

Presser, S., Rothgeb, J. M., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., & Singer, E. (Eds.). (2004). *Methods for testing and evaluating survey questionnaires. Wiley series in survey methodology*. Wiley-Interscience.

Rasinski, K. A., Lee, L., & Krishnamurty, P. (2012). Question order effects. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 1: Foundations, planning, measures, and psychometrics* (pp. 229–248). American Psychological Association. https://doi.org/10.1037/13619-014

Reja, U., Manfreda, K. L., Hlebec, V., & Vehovar, V. (2003). Open-ended vs. close-ended questions in web questionnaires. *Developments in Applied Statistics*, *19*, 159–177.

Schul, Y., & Schiff, M. (1993). Measuring satisfaction with organizations. Predictions from information accessibility. *Public Opinion Quarterly*, *57*(4), 536–551. https://doi.org/10.1086/269394

Schwarz, N. (1995). What respondents learn from questionnaires: The survey interview and the logic of conversation. *International Statistical Review/Revue Internationale de Statistique*, *63*(2), 153–168. https://doi.org/10.2307/1403610

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 436–457). Blackwell.

Schwarz, N., Oyserman, D., & Peytcheva, E. (2010). Cognition, communication, and culture: Implications for the survey response process. In J. A. Harness (Ed.), *Wiley series in survey methodology. Survey methods in multinational, multiregional and multicultural contexts* (pp. 177–190). Wiley & Sons.

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

Schwarz, N., & Sudman, S. (Eds.). (1992). *Context effects in social and psychological research*. Springer.

Smith, T. W. (1992). Thoughts on the nature of context effects. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 163–184). Springer.

Stark, T. H., Silber, H., Krosnick, J. A., Blom, A. G., Aoyagi, M., Belchior, A., Bosnjak, M., Clement, S. L., John, M., Jónsdóttir, G. A., Lawson, K., Lynn, P., Martinsson, J., Shamshiri-Petersen, D., Tvinnereim, E., & Yu, R. -r. (2018). Generalization of classic question order effects across cultures. *Sociological Methods & Research*, *23*, 1–18. https://doi.org/10.1177/0049124117747304

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Tourangeau, R., Singer, E., & Presser, S. (2003). Context effects in attitude surveys: Effects on remote items and impact on predictive validity. *Sociological Methods & Research*, *31*(4), 486–513. https://doi.org/10.1177/0049124103251950

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage.

Willis, G. B. (2015a). *Analysis of the cognitive interview in questionnaire design. Understanding qualitative research*. Oxford University Press.

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

World Values Survey. (2014). *World values survey. Wave six*. http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp

Zanna, M. P., Olson, J. M., & Fazio, R. H. (1981). Self-perception and attitude-behavior consistency. *Personality & Social Psychology Bulletin*, *7*(2), 252–256.

Zuell, C., & Scholz, E. (2015). Who is willing to answer open-ended questions on the meaning of left and right? *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *127*(1), 26–42. https://doi.org/10.1177/0759106315582199

Author Biography

Patricia Hadler is a researcher at the GESIS Pretest Lab, GESIS - Leibniz Institute for the Social Sciences and a doctoral candidate at the University of Mannheim. Her research interests focus on questionnaire design and evaluation, in particular cognitive testing and web probing. She can be contacted at patricia.hadler@gesis.org.