Context Effects in Internet Surveys: New Issues and Evidence

Jolene D. Smyth,
Don A. Dillman
Washington State University
Leah Melani Christian
University of Georgia

INTRODUCTION

When one asks another person, "Is it cold in here?" the answer is undoubtedly affected by the situation or context in which the question is posed. If it is winter and one is talking to a live-in partner, the question may refer to whether or not it is colder in the room than usual. If that same question is asked to a guest, however, it can take on a totally different meaning: "Is the room as warm as you would like or expect it to be?" The answer that the guest gives might be affected by other contextual factors such as norms of politeness, expectations of what actions the person asking might take given a certain answer, and whether the guest knows the questioner's preference for room temperature. If the question is posed while standing in a walk-in freezer, it takes on yet another meaning as the definition of cold now refers to temperatures sufficiently low to preserve food. In this context, the fact that one prefers rooms to be at a comfortable 22 degrees Celsius (71.6 degrees Fahrenheit) is largely irrelevant to the question at hand. In yet another context this question may not actually refer to temperature at all and, in fact, it may be more of a statement than a question. For example, if, following a particularly rude comment, one asks, "Is it cold in here?" the question, in this context, becomes an evaluation of the actions of another person in the room and can serve as an admonition for unacceptable behavior.

Context matters in normal conversation. Everything from the relationship between participants in the conversation to the physical setting, history of interaction, and presence or absence of others is combined to form the context of every-day conversations. This context

affects how simple statements are interpreted and responded to, oftentimes giving to statements meaning that would not be immediately obvious without it. Context also matters in how people interpret and respond to survey questions. During the latter part of the 20th century, evidence accumulated that a wide variety of contextual factors, such as previous questions and interviewer presence, influence the answers people choose to give to survey questions (e.g., Schuman & Presser, 1981; Schuman, 1992; Sudman, Bradburn & Schwarz, 1996). More recently, research has shown that factors such as visual layout of questions influence people's answers (Christian & Dillman, 2004; Smyth, Dillman, Christian, & Stern, 2006a; Tourangeau, Couper, & Conrad, 2004). Although our knowledge about the mechanisms through which context effects occur is limited, we now know that the interpretation and answering of survey questions is influenced by much more than the wording of individual queries.

It is becoming increasingly clear that many of the context effects found in paper and telephone modes extend to surveys conducted via the Internet as well. However, the substantial variability that is possible in Internet survey design means that the Internet does not provide a unified survey mode in the same way mail and telephone do. This variability in design possibilities corresponds to equally complex variability in possible context effects that result from the use and combination of specific design features. For example, under some circumstances Internet surveys closely resemble mail surveys, such as when they contain only visual stimulus, but under other circumstances, such as when they incorporate aural stimulus or interactive response feedback, they more closely resemble telephone surveys. In still another instance a combination of telephone- and Internet-like features may be included in a design, resulting in complex patterns of context effects. As a result of the myriad possibilities introduced by the Internet, it is becoming increasingly important to approach the examination of

context effects from the more specific vantage point of features of survey modes. From this vantage point we can examine when and how context effects in Internet surveys resemble those in mail and telephone surveys and also look at new sources of context effects found only in Internet surveys. The infancy of Internet surveying, seldom used prior to 1998, means that these effects are only now being studied and many important issues have yet to be systematically addressed.

Our purpose in this chapter is to first present a definition of context effects that eliminates from consideration factors beyond the control of survey researchers yet is sufficiently broad to incorporate diverse but related sources of survey context. We then examine four types of context effects that have been documented in mail and telephone surveys with an eye toward identifying new concerns that have arisen or may arise as a result of conducting Internet surveys. The four sources of context effects discussed are: the survey mode used to pose questions to respondents, the order in which questions are asked, the ordering of response options, and the choice of response scale. In addition to reviewing previous research, we report results of new context experiments in which we manipulate response scales across Internet and telephone modes.

WHAT ARE CONTEXT EFFECTS?

Survey researchers often use the term "context effects" narrowly to refer to the effects of prior questions on responses to subsequent questions (Schuman & Presser, 1981; Schuman, 1992; Sudman et al., 1996; Tourangeau, Rips, & Rasinski, 2000). For example, Schuman and Presser (1981) include context effects, which they define as transfers of meaning between questions, under a broader umbrella of question order effects. However, as Tourangeau et al. (2000) point out, the treatment of context as synonymous with question order is largely one of practicality.

> In principle, of course, question context could have a much broader meaning, reflecting the purposes of the survey (as these were presented to the respondent), the climate of opinion at the time of the survey, the presence of other people during the interview, the characteristics of the interviewer, or even the weather at the time of the interview. As a practical matter, though, most research on question context has focused on the effects of earlier questions on answers to later ones. (Tourangeau et al., 2000: p. 200)

Although they largely limit their discussion of context effects to question order effects, Tourangeau et al. (2000) point out that many aspects of the survey instrument as well as the respondents' surroundings, moods, and experiences can provide context that may affect how questions are interpreted and answered. The ubiquitous nature of context effects poses a dilemma for researchers of defining such effects so broadly that the term, context, loses its utility for describing such effects, or defining it so narrowly (e.g., only question order effects) that it captures only a small and incomplete aspect of an important phenomenon. With this dilemma in mind, we use the term "context effects" in this chapter to describe influences on question answers that occur because of information passed on to the respondent *from the survey environment* in which the question is answered. This definition allows us to eliminate superfluous influences that are beyond the control of the researcher (i.e., the weather, past experiences, and in many instances setting and mood) and would weaken the explanatory power of the concept, but to broaden the definition beyond question order effects.

The need to broaden the definition of context effects in this way is suggested by the work of Schwarz (1996) and Schwarz, Grayson, and Knäuper (1998) in which, using Grice's (1975) logic of conversation, they argue that respondents approach survey questionnaires as if they are engaged in cooperative conversations. Within this framework, respondents treat formal features

of the questionnaire, which represent the contribution of the researcher to the conversation and are assumed to be guided by Grice's (1975) maxims of communication,[1] as communicating the researcher's expectations. Thus, all formal features of the questionnaire or survey process can form context that is used by respondents to make sense of and answer questions, resulting in context effects.

CONTEXT EFFECTS AS A RESULT OF SURVEY MODE

To the extent that certain response options or questions are presented in differing, but seemingly mode-appropriate ways, different survey modes are bound to produce different response distributions (Dillman & Christian, 2005). For example, it is customary in telephone surveys to avoid presenting a "don't know" option to respondents, but to record don't know answers if they are volunteered (Dillman, 2000); to present scales with only the end-points labeled (Dillman & Christian, 2005 and Christian, 2003); and to use a forced-choice (yes/no) format for multiple-answer questions (Smyth, Dillman, Christian & Stern, 2006b and Smyth, Dillman, & Christian, 2006c). In contrast, those administering self-administered surveys often include don't know as an explicit option, present scales with all scale points labeled, and present multiple-answer questions in the check-all-that-apply format. With increasing use of Internet surveys, these practices raise questions about when Internet queries should be matched more closely to the telephone formats, when they should be matched more closely to the mail formats, and in what instances Internet-specific formats should be developed and deployed. Perhaps more important though, are the questions they raise about data quality in increasingly common mixed- or multi-mode research designs. The extent to which such changes across modes alter the survey context and results are empirical questions in need of additional research. Nonetheless, setting

[1] The Maxim of Quantity - contributions are meant to be as informative as possible without giving too much information; The Maxim of Quality - contributions are true and can be backed up by evidence; and The Maxim of Manner – contributions are to be clear, unambiguous, brief, and orderly. (Grice, 1975: pp. 45-46)

aside issues of question format differences, we can still expect context effects related to other features of various survey modes.

The Presence or Absence of Interviewers. Researchers have long recognized that the influence of social desirability on responses to sensitive or embarrassing questions (e.g., physical health or criminal activity and substance abuse—see Aquilino, 1994 and Phillips & Clancy, 1970) is higher in interviewer-administered than self-administered survey modes (Tourangeau & Smith, 1996); the presence of interviewers makes norms about interaction and concern with presentation of self much more salient (Dillman & Tarnai, 1991; de Leeuw, 1992; Krysan, Schuman, Scott & Beatty, 1994). To avoid this problem, some face-to-face interviews include self-administered forms of data collection for the sensitive questions (de Leeuw, 2005).

Given previous research, we can expect both mail and the Internet to produce fewer socially desirable responses because they are self-administered modes. For both mail and Internet surveys, however, this expectation is contingent on respondents' trust in the survey situation (e.g., Do they believe confidentiality statements and do they think their submitted answers are identifiable to the researchers?) and whether or not they are alone when they complete the questionnaire. In addition, for the Internet it is further contingent on respondents' perceptions of and trust in computer and Internet technology in general (i.e., Can information be stored on their computer without them knowing? Does the Internet provide a confidential vehicle for transmitting sensitive information?) (Couper, 2000). While more research is needed in this area, preliminary evidence indicates that Internet surveys do produce less social desirability than telephone surveys (Chang, 2001; Taylor, Krane & Thomas, 2005).

Another way that an interviewer-present survey context may lead to altered responses is through acquiescence or the tendency to give agreeable responses regardless of what the true

response might be; it is easier and more polite in most cultures to agree than to disagree (Javeline, 1999; Dillman, 2000). An example of acquiescence can be seen in findings showing that 60 percent of respondents in one group agreed with a statement blaming crime and lawlessness on individuals while 57 percent of respondents in another group agreed that social conditions were to blame from crime and lawlessness (Schuman & Presser, 1981). Five months later respondents were re-interviewed, but given the opposite question. Results indicated that over the two interviews about 25 percent of respondents agreed to both versions of the question.

As was the case with social desirability, since acquiescence is a norm-based phenomenon we expect to find higher degrees of it in interviewer-administered modes where social norms are invoked and lower degrees in self-administered modes where social norms can more easily be ignored. This expectation has been confirmed by some (Dillman & Tarnai, 1991), but disconfirmed by others (de Leeuw, 1992), leaving it an open question for further research, research that must now address the extent to which acquiescence is a problem in Internet surveys as well as telephone, mail, and other modes (see Smyth et al., 2006b for discussion of acquiescence in multiple-answer questions in Internet surveys).

Control of the Survey Process. Modal differences in who has control over the survey process can also lead to context effects in response distributions. In telephone and face-to-face surveys an interviewer has control over the survey process and can regulate both the order in which respondents receive questions and the speed with which the survey is completed. Telephone interviews tend to be conducted at relatively quick paces, thus limiting the amount of time that respondents have to process questions and formulate answers (Krosnick & Alwin, 1987) and leading to more use of pre-formed responses and previously accessed information (Dillman, Sangster, Tarnai, & Rockwood, 1996). In mail surveys, on the other hand, the

respondent maintains complete control over the speed of the survey and the order in which questions are processed (Dillman, 2000).

Internet surveys, however, can be designed to give respondents varying levels of control over survey speed and question order. Some Internet surveys, for example, impose time constraints by timing out if left idle for specified periods of time; others have no such constraints. In one comparison in which the Internet had no imposed time constraints, Christian, Dillman, and Smyth (forthcoming) found that respondents to the Internet mode took substantially longer to complete a 25 question survey (a mean of 21.4 minutes) than did respondents to the telephone mode (12.1 minutes). Varying levels of time pressure may result in respondents experiencing surveys in different contexts; however, empirical research into the exact mechanisms behind time pressure differences and the consequences of these differences remains scarce.

With respect to question order, scrollable Internet surveys where all the questions appear on one page give the respondent full control over what they process and when, making this type of Internet survey quite similar to mail surveys. In contrast, page-by-page survey construction allows researchers to impose various levels of control over the order in which questions are read, ranging from very little (like mail) to full control (like face-to-face and telephone). For example, Internet survey designers can control whether respondents are allowed to advance through the survey if they have not adequately answered a question and whether they can move backward through the survey to return to previous questions. Context effects associated with question order will be considered in more depth below. Suffice it to say here that having full control over the order in which questions are read and processed allows researchers to at least anticipate, and in some instances eliminate, context effects due to question order effects.

Aural vs. Visual Communication. A final fundamental mode difference that can affect responses is the basic form of communication that makes up the modes. Whereas interviewer-administered surveys are almost always (with the exception of show cards) based on aural communication, paper surveys are based entirely on visual communication. Internet surveys, however, can be designed entirely visually so that they emulate paper surveys or they can be designed using features such as sound and video clips, moving them closer to telephone and face-to-face interviews.

The distinction between these two types of communication is important because they provide different cues from which respondents infer information about the survey and the response task. For example, visual grouping and subgrouping in self administered questionnaires may communicate to respondents that questions (and/or response options) are related, leading to question order effects, a topic that will be discussed below. When the same questions are read over the telephone, respondents may not make the same connections between them. A number of recent studies have produced evidence that respondents to self-administered, visually-based surveys rely on visual cues in addition to the verbal messages on the page/screen to interpret and answer survey questions (Christian & Dillman, 2004; Couper, Tourangeau, & Kenyon, 2004b; Smyth, Dillman, Christian, & Stern, 2004; 2006a; Tourangeau et al., 2004).

New Mode Context Effects Introduced by Internet Surveys

In addition to the factors that they share in common with either mail or telephone surveys, Internet surveys are also complicated by a number of Internet-specific factors that can lead to alterations in context. The landscape orientation of most computer screens, for example, may encourage a horizontally oriented design, such as when response options are double or triple banked to take advantage of horizontal space. Moving response options into multiple columns

may interrupt the linear flow of scales or make some options appear more prominent than others, thus altering the interpretation and answering of the question. For example, Christian and Dillman (2004) report a tendency for respondents to focus on the top line of options when they are double or triple banked (also see Christian, 2003 and Dillman & Christian, 2005).

In addition, varying hardware and software configurations introduce the complication that what the designer creates may not actually be what the respondent receives. While steps can be taken to minimize alterations to the questionnaire as a result of hardware and software configurations, none are effective in all situations and some respondents will inevitably receive the questionnaire with altered visual effects (e.g., changes in font, color or spacing) that may then affect their interpretation and answering of questions. Survey compatibility with old or less common browsers and the likelihood that some respondents disable JavaScript (i.e., a scripting programming language commonly used in designing Internet surveys) also pose significant challenges for the Internet mode (see Schmidt, this volume). Perhaps more troublesome though is that different hardware and software configurations are likely to be associated with various demographic factors such as age and socioeconomic status, thus introducing systematic bias into results (Buchanan & Reips, 2001).

The Internet also introduces a number of technical innovations that may have context effects, some of which are as yet unknown. Researchers have virtually unlimited ability to use color, graphics, and, as mentioned previously, even sound and video clips in Internet surveys. The effects of such technical features on survey responses has barely begun to be explored although there is some evidence that such features may increase download time, promoting early termination (Couper, 2002). One such component that has received some research attention is the effect of including progress indicators in Internet surveys. Couper, Traugott, and Lamias

(2001) report only slight, and not significant, increases in completion rates when progress indicators are present in Internet surveys; however, they speculate that some of the gain in completion rates is offset by early terminations due to increased download times. Another innovative component that has received some research attention is the effects of using photos in Internet surveys (Couper, Conrad, & Tourangeau, 2005; Couper et al., 2004b; Witte, Pargas, Mobley, & Hawdon, 2004). The findings here generally indicate that respondents take into account the information in the photos when formulating their responses and that results are significantly altered by the presence of the photos. Finally, researchers are beginning to examine the effects of new question formats that are possible in Internet surveys. For example, one study indicates that when drop-down menus require respondents to scroll down to reveal all available answers, the options that appear initially are more prone to endorsement (Couper, Tourangeau, Conrad, & Crawford, 2004a).

CONTEXT EFFECTS AS A RESULT OF QUESTION ORDER

The most commonly studied type of survey context effects are question order effects whereby previous questions and the answers given to them affect how respondents interpret and answer subsequent questions. Question order effects become increasingly likely to occur the closer questions are to one another, in terms of both topic and location in the questionnaire. When they do occur, question order effects can be categorized as either assimilation effects, where responses to two or more question become more highly correlated, or contrast effects, where responses become less correlated. Since others (Schuman & Presser, 1981; Strack, 1992; Tourangeau, 1992; and Tourangeau et al., 2000) have reviewed question order effects in substantial depth, our discussion of them here is more limited and is meant only to provide a general framework for understanding potential question order effects in Internet surveys.

Oftentimes question order effects occur when a prior question brings to mind considerations that are then more accessible for use in interpreting and responding to a subsequent question—a priming function (see Schwarz & Bless, 1992; Schwarz & Clore, 1983; and Tourangeau, Rasinski, Bradburn, & D'Andrade, 1989). The result of priming is likely to be an assimilation effect when the information brought to mind is used to form a representation of the object being inquired about. In contrast, if the information brought to mind is used to form the referent or standard to which the respondent compares the object of inquiry, the result could be either a contrast or an assimilation effect depending on how the referent is formulated. For example, we would expect a contrast effect if the standard of comparison for a question asking about one's current quality of life was a particularly bad event in the past, such as a divorce process or a period of homelessness (both are extreme values on the dimension of judgment). If that same question were asked, in a context where the standard of comparison was an average of previous "measures" of quality of life, we would expect an assimilation effect (Tourangeau et al., 2000).

A previous question may also influence responses to subsequent questions by invoking a norm or cultural value. Oftentimes, but not always, the norms/values invoked are the norms of cooperative communication sketched out in Grice's maxims of communication. For this type of effect to occur, the respondent must perceive the two questions as being related topically, or, in other words, as belonging to the same conversational context. For example, respondents following Grice's maxims of communication might perceive questions as related because they expect them to be relevant to the conversation as it has been occurring. In this instance we might expect an assimilation effect (carryover effect) because seeing the questions as topically linked would lead to the use of similar considerations in formulating responses (for examples see

Schuman & Presser, 1981; Tourangeau & Rasinski, 1988; Strack, Schwarz, & Wänke, 1991). However, if the respondents invoke the norm of non-redundancy, they may assume that later answers should exclude information already used or reported in earlier answers. Following this conversational norm would result in a contrast effect, more specifically, a subtraction effect (for examples see Schuman & Presser, 1981; Schwarz, Strack & Mai, 1991b; Tourangeau, Rasinski, & Bradburn, 1991 and Mason, Carlson, & Tourangeau, 1994). This type of order effect is particularly common in part-whole question sequences (i.e., marital satisfaction and general satisfaction sequences).

Respondents may also invoke cultural norms and values that are not necessarily part of the logic of conversation. For example, norms of fairness (reciprocity or evenhandedness), consistency, and neutrality can be invoked in answering survey questions. The effects of the norm of evenhandedness on survey responses was demonstrated in a classic study by Hyman and Sheatsley (1950) and replicated by Schuman and Presser (1981). In both instances, researchers found that respondents were significantly more likely to answer a question asking whether communist reporters should be allowed to report on visits to the United States affirmatively after having answered affirmatively to a similar question about allowing U.S. reporters to report on visits to the Soviet Union. Similarly, research has shown that placing somewhat related items together in a questionnaire can result in assimilation effects because respondents don't want to appear inconsistent in their responses (Smith, 1983). Finally, fear of taking extreme positions on highly polarized topics may lead to contrast effects as respondents attempt to appear neutral and non-partisan by selecting some items but rejecting others (Tourangeau, 1992).

New Question Order Context Effects Introduced by Internet Surveys

Much of the above research and theorizing about question order effects predates the expansion of Internet surveying. While it is highly likely that the same types of order effects occur in Internet surveys for the same reasons as cited above, the Internet also introduces a number of factors not found in other modes that may impact question order effects. For example, in Internet surveys it is easy to vary how the question order is presented to respondents; this manipulation can range from simply flipping the order of two questions to complete randomization of all questions. However, the extent to which these are viable options depends greatly on the purpose and methodology (i.e., single vs. mixed mode; cross-sectional vs. longitudinal vs. panel; etc.) of the study.

As mentioned previously, researchers also have the option of presenting all of the questions on one screen, making the questionnaire more similar to a paper survey, or presenting questions on separate screens (page-by-page construction), making it more similar to a telephone survey. The choice that researchers make between these two options may have important implications for question order effects. Presenting all of the questions on one screen allows respondents to easily scroll back and forth throughout the survey to remind themselves of the context of the conversation and to reexamine relationships between questions, but doing so may increase the likelihood of context effects and the likelihood that the effects flow both forward and backward (i.e., early question affect later questions and visa versa). Using page-by-page construction, on the other hand, should reduce the likelihood of context effects by making the relationships between questions less obvious. As in telephone surveys, the extent to which early questions can affect later questions in a page-by-page design depends on respondents' memory capacity. However, the danger in page-by-page design is that the respondent may lose track of

the relationships between questions so much that they loose a sense of continuity in the survey and questions seem isolated.

The expectation of increased context effects when multiple questions appear on one screen has been confirmed in at least two published studies (also see Reips, 2002). In the first, Couper et al. (2001) found that item correlations were higher among items when they appeared together on a screen as opposed to on separate screens, but that the differences were not large enough to be statistically significant. In the second study Tourangeau et al. (2004) presented respondents with eight related items regarding diet all together on one screen, separated in groups of four on two screens, or each on its own screen. They found that the eight items were most highly correlated when all eight were on one screen followed by the two-screen (4 items per screen) treatment and then the treatment in which all items appeared on their own page. While the context effects reported in these studies are assimilation effects, it is likely that presenting part-whole question sequences together on a screen would produce contrast effects.

CONTEXT EFFECTS RESULTING FROM RESPONSE CATEGORY ORDER

Similarly to question order effects, the order in which response categories are presented can also influence the answers given to questions. Typically response order effects come in two types: primacy effects, in which options are more likely to be endorsed when they appear early in the list, and recency effects, in which options are more likely to be selected when they appear late in the list. Current research suggests several mechanisms that may underlie response order effects. The evidence seems to indicate that none of these factors can account for all of the nuances of response order effects, but instead multiple factors may oftentimes work in concert to produce or to obscure evidence of primacy and recency effects.

Memory Limitations and Response Order Effects

In early work researchers proposed memory limitations as an explanation for response order effects (Schuman & Presser, 1981; Krosnick & Alwin, 1987). The theory of memory limitations predicts recency effects, especially in questions with long or complex lists of options, because respondents have an easier time recalling response options that are stored in their short term memory, the last options read or heard, than those stored in their long term memory, the first options read or heard. Response options in the middle of lists are the least likely to be remembered. This theory, however, has been called into question based on findings of primacy effects (Mueller, 1970; Krosnick & Alwin, 1987) and response order effects in questions with only two or three response options (Schuman & Presser, 1981; Schwarz, Hippler, & Noelle-Neumann, 1992).

The Cognitive Elaboration Model of Response Order Effects

More recent work has led to the development of the cognitive elaboration perspective in which a three way interaction between response option order, presentation mode, and item plausibility is theorized (Schwarz et al., 1992; Sudman et al., 1996). This perspective relies on the underlying assumption that each response option is a short persuasive communication that brings to mind for the respondent either confirming or disconfirming cognitive elaborations. The more time respondents have to consider an option the more confirming or disconfirming information they will think of making them more likely to endorse or reject the option. The amount of time that respondents have to consider an option depends on both its serial position in the list and the survey presentation mode (visual vs. aural) (Krosnick & Alwin, 1987; Schwarz et al., 1992).

In visually-based survey modes respondents can deeply process items appearing early in the list but as they move through the list their minds become more cluttered and their ability to

process options becomes increasingly limited. Thus, respondents give options appearing later in the list less processing time and energy (Krosnick & Alwin, 1987). If early response options are plausible, bringing to mind confirmatory thoughts, they are more likely to be endorsed than those appearing later, resulting in primacy effects. However, if early items are implausible, bringing to mind disconfirming thoughts, they are less likely to be selected, resulting in recency effects.

In contrast, in aurally-based survey modes, interviewers control the pace of the survey and tend to move rapidly through the response options, not allowing long pauses until the end of the list. Respondents, therefore, have more time to process response options appearing later in lists. Processing-time limitations combined with memory limitations lead to deeper processing of later items. If the later options are plausible a recency effect is expected as they will be more likely to be selected but if they are implausible respondents will avoid selecting them, resulting in a primacy effect. Under typical survey conditions, where researchers try to avoid providing implausible response options, we would expect primacy effects in visually-based surveys and recency effects in aurally-based surveys.

Evidence supporting these expectations was found in research by Schwarz et al. (1992). In their experiments, the visual presentation format produced primacy effects while the aural presentation format produced recency effects. Based on the theory that response order effects depend on the confirming or disconfirming thoughts respondents have about the response options, Schwarz et al. (1992) hypothesized that they could eliminate response order effects by stimulating respondents to think about the specific content of response options in previous questions. In this way, they primed the respondent so that processing time limitations would have less influence; the respondent had already processed a lot of information about the topic by the time they got to the response options of interest. They found that including relevant context

questions in this way entirely eliminated response order effects, a result that supports the cognitive elaboration model of response order effects.

Satisficing as an Alternative Explanation for Response Order Effects

While the cognitive elaboration model seems to accurately account for response order effects, Tourangeau et al. (2000) point out that response order effects could be explained equally well by another more parsimonious perspective, Krosnick's (1991; 1999) theory of satisficing. According to this theory, respondents avoid expending the energy to provide an optimal answer by making a dichotomous judgment (i.e., yes/no) about each response option. They proceed through the options until they come to one they consider satisfactory and then endorse that option and discontinue their consideration of subsequent options. In visual modes, respondents can be expected to work from the top of the list down and in aural modes from the bottom of the list up for the same reasons discussed above.

A Complicating Factor: Anchoring Effects

Researchers also recognize that other mechanisms unrelated to memory, cognitive elaboration, or satisficing response strategies may underlie response order effects. For example, when respondents are asked to evaluate options on some dimension of judgment (i.e., degree of typicality, degree of appeal, etc.) one option may stand out as a standard or anchor against which all others are compared. If the option used as the standard sets a high expectation, then the moderate options that follow it will become less likely to be endorsed (primacy effect). If it sets a low expectation the moderate options that follow it become more likely to be endorsed (recency effect). In one example, respondents were asked which of the foods, potatoes, noodles, and rice, were "typically German." Significantly more respondents marked potatoes and noodles as "typically German" when rice was presented at the beginning of the list than when rice

appeared after these other options (Noelle-Neumann, 1970). Contrast effects such as these can greatly complicate predictions and interpretations of primacy and recency effects.

New Response Order Effects Introduced by Internet Surveys

To date very little research on the effects of response option order in Internet surveys has been done. The research that has been done, however, indicates that Internet surveys are prone to similar types of response option order effects as paper surveys. For example, Couper et al. (2004a) asked respondents to report which of twelve nutrients is the most important when selecting a breakfast cereal and found that options have significantly higher endorsement rates when they appear in the top five positions of the list. A second question asking about the most important feature in decisions about automobile purchases produced similar patterns of primacy. Perhaps more importantly though, their results showed that response option order effects were magnified under specific design conditions that are unique to Internet surveys. In particular, in what they refer to as the "visibility principle" response options that were immediately visible, in this case the first items in drop-down menus, were more likely to be selected than those that respondents had to scroll down to reveal.

While the Internet appears to present similar response option order effects as paper surveys it also provides innovative design features that might be used more easily and cost effectively to address such effects. Most importantly, Internet survey designers can vary the order that response options appear for each respondent up to the point of fully randomizing them in order to mitigate response option order effects. But again, the possibility of taking advantage of this ability greatly depends on the purpose of the study and its design.

RESPONSE OPTION SCALE EFFECTS

In addition to considering the possibility of response order effects, survey designers must attempt to avoid response scale effects. Scale effects occur when the choice of scale provided to answer a question affects the answers given. Tourangeau et al. (2000) offer a useful summary of five ways in which response scales can alter the way respondents answer questions: (1) positivity or leniency bias – respondents avoid the negative end of rating scales; (2) scale label effects – respondents avoid negative numbers because they imply more extreme judgments than low positive numbers (Schwarz, Knäuper, Hippler, Noelle-Neumann & Clark, 1991a); (3) response contraction bias – responses clump toward the middle of scales because respondents attempt to avoid extreme categories; (4) reference point effects – specific numbers (0, 10, 50, 100 etc.) have predetermined cultural connotations that either encourage or discourage their selection (i.e., the number zero implies absence of); and (5) scale range effects – the range of a scale either (a) changes definitions of the topic at hand (i.e., when asked about frequency of anger low scales may imply that only serious anger incidences should be counted and high scales may imply that less serious incidences are relevant) or (b) changes definitions of the distribution of the characteristic in the general public which are then used as a reference for the respondent in judging their own situation (i.e., low scales imply that average people don't get angry very often while high scales imply higher incidence of anger among average people) (Pp. 248). In addition the visual presentation of scales may influence and interact with these five scale effects. Research by Dillman and Christian (2005) and Christian and Dillman (2004) has shown that the answers respondents provide to both paper and Internet surveys are influenced by the visual presentation of scales.

The first two effects, positivity bias and scale label effects, are demonstrated in face-to-face and paper surveys by Schwarz et al. (1991a). When respondents were asked to rate how

successful they had been in life 34 percent of those who received a scale ranging from 0-10 rated themselves between 0 and 5 while only 13 percent of those who received a scale ranging from -5 to 5 (with the same endpoint labels) rated themselves on equivalent values (-5 to 0). Further experimentation revealed that respondents determined the meaning of the scale labels based on the numbers associated with them. The label "not at all successful" associated with a score of 0 can be interpreted to mean an absence of success, but the same label associated with a negative number (-5) may refer to the presence of explicit failure (also see Schwarz et al., 1998). This finding has recently been extended to Internet surveys (Tourangeau, Couper, & Conrad, 2006).

The third effect, the tendency for respondents to prefer middle categories, is confirmed in recent research on Internet surveys by Tourangeau et al. (2004). In fact, these researchers found that respondents are drawn toward the visual midpoint of scales, not necessarily the conceptual midpoint. In one experiment, respondents were presented with scales that included nonsubstantive response options ("no opinion" or "don't know"). These response options were visually separated from the scale for some respondents, but not for others. When they were not visually separated, responses tended to fall lower on the scale, toward the visual midpoint. When they were visually separated, responses tended to be higher, again toward the visual midpoint (which had moved up when the nonsubstantive responses were separated from the scale). In additional experiments, spacing between scale items was manipulated such that in some treatments the conceptual and visual midpoints of the scale were aligned and in others they were not. Results were similar in that respondents tended to be drawn toward the visual midpoint over the conceptual midpoint.

Several studies have addressed the fifth effect, scale range effects. This body of research is largely based on an early study by Schwarz, Hippler, Deutsch, and Strack (1985). They found

that when respondents were asked how many hours of television they watch per day only about 16 percent reported watching more than 2 ½ hours when they were presented with a scale ranging from "up to ½ hour" to "more than 2 ½ hours." In comparison, about 38 percent reported watching more than 2 ½ hours when they were presented with a scale ranging from "up to 2 ½ hours" to "more than 4 ½ hours" (Schwarz et al., 1985). In addition, respondents given the low scale range estimated that others watch television significantly less than did respondents given the high scale range did (2.7 hours vs. 3.2 hours). These findings indicate that respondents use information in scales to inform their answers; they may assume that the scale represents the most common answers or even that the mid-point of the scale represents the response of a "typical" respondent. Given such information, respondents may then formulate their answer based on how typical they think they are.

Schwarz and colleagues' (1985) findings were replicated and extended by Rockwood, Sangster, and Dillman (1997). These researchers found that the range provided in response options has significant effects for questions about frequent and mundane behaviors (hours students spend studying or watching television), but does not have significant effects for questions about more rare topics (such as grade point average), suggesting that the scale range is more influential when respondents have to estimate frequencies rather than simply recall factual information. In addition, they established that response option effects such as these occur in both mail and telephone surveys and that in telephone surveys they can be confounded by other factors such as social desirability and pace of the interview (see the research example below for an extension to Internet surveys).

New Response Scale Effects Introduced by Internet Surveys

Because Internet surveys are generally visually based they are not plagued by the same memory limitations as telephone surveys. As a result, unlike telephone surveyors, Internet surveyors can use fully labeled scales, a practice that has the potential to eliminate vagueness in some scales, thus reducing the amount of interpreting and inferring that the respondent has to do and thereby reducing context effects. This practice seems preferable, but again, may be limited by any mixed-mode considerations.

The Internet also introduces the opportunity to utilize new and innovative ways of displaying response options, the effects of which are largely unknown as of yet. We have already discussed the use of drop-down menus. Another example is the sliding scale in which the respondent uses the mouse to drag a marker along a scale and drop it at the point on the scale that they feel best represents them or their opinion. Bayer and Thomas (2004) compared seven point sliding scales to conventional scales constructed using radio buttons. They found that the sliding scales took significantly more time to complete and that they resulted in higher mean values because respondents with sliding scales were more likely to endorse the highest point on the scale than were those with radio button scales. While Bayer and Thomas' (2004) sliding scales limited respondents to seven possible scale points, others have examined visual analogue scales in which the respondent can mark any point on the scale resulting in interval level scalar data (Funke & Reips, 2006). Results indicate that visual analogue scales (when linearly transformed) also produce more extreme responses than categorical scales. As more and more innovative scalar designs are developed for Internet surveys it will be important to systematically study the effects they have on respondent answers.

A RESEARCH EXAMPLE: HOW THE ESTIMATION OF ROUTINE BEHAVIORS IS INFLUENCED BY MODE AND RESPONSE SCALE CONTEXTS

A survey conducted in the fall of 2004 emulated the Schwarz et al. (1985) and Rockwood et al. (1997) studies by asking a randomly selected sample of Washington State University Undergraduate students how many hours per day they typically study, spend on a computer, and watch television. Students were randomly assigned to receive a low estimate scale (½ hour or less to more than 2 ½ hours), a high estimate scale (2 ½ hours or less to more than 4 ½ hours) or an open-ended answer box in either the Internet or telephone mode (see Table 1 for response rates and Figure 1 for treatments). This study extends past research by examining scale range effects in the Internet mode and by considering the extent to which such questions are affected by the Internet context relative to the telephone context.

Results

Response distributions and significance tests for the low and high range scales are shown in Table 1. These results indicate that Internet surveys are prone to the same type of scale effects reported by Schwarz et al. (1985) and Rockwood et al. (1997). When asked how many hours per day they study 29.8 percent of Internet respondents to the low range scale and 71.1 percent of Internet respondents to the high range scale reported studying over 2.5 hours per day, a difference of 41.3 percentage points! Similarly, only 29.2 percent of low range scale respondents reported over 2.5 hours of computing per day while 66.0 percent of high range scale respondents reported this amount of computing and only 8.4 percent of low range scale respondents reported watching over 2.5 hours of television per day compared to 21.3 percent of high range scale respondents. Results are similar within the telephone mode. Of the respondents who received the low scale 34.2 percent reported spending over 2.5 hours studying, 25.4 percent reported spending over 2.5 hours on a computer, and 6.4 percent reported spending over 2.5 hours watching television. In comparison, among those receiving the high scale 73.6 percent

reported studying, 51.1 percent reported computing, and 17.7 percent reported watching television over 2.5 hours per day. All of these differences across scale ranges for both the Internet and telephone modes are highly significant ($p \leq .000$).

These results indicate that the scales used provided different contexts resulting in different responses; however, they do not address whether or not survey mode context effects occurred either independently or in conjunction with the scale effects. For the three questions analyzed here, logistic regressions were conducted to estimate the likelihood of reporting greater than 2.5 hours of activity based on mode and question form. The logistic regression results allow us to do three things that the chi-square analyses reported above do not: (1) test for the significance of response scale effects while controlling for mode effects and visa versa, (2) examine whether response scale effects and mode effects depend on each other (interaction effects), and (3) estimate the magnitude of the scale and mode effects.

The findings of the logistic regression equations (Table 2) confirm the findings reported above. Respondents were 5.35 times more likely to report studying, 3.07 times more likely to report computing, and 3.13 times more likely to report watching TV over 2.5 hours per day if they received the high range scale. Respondents to all three questions were also more likely to report spending over 2.5 hours per day on these activities if they received the open-ended response format, although this format did not have as large of effects as the high range scale (Studying = 1.99 times, computing = 1.60 times, and TV = 2.82 times).

With regard to mode effects, the findings of the logistic regressions indicate that mode did not have an independent effect on the results of any of the three questions analyzed. However, significant interaction effects were found between mode and the open-ended response format for both the study time and computing time questions, indicating that the effect of the

open-ended response format depends on survey mode. In particular, Internet respondents to the open-ended question format (compared to the low scale range respondents) were 3.62 times more likely to report over 2.5 hours of studying per day while their telephone counterparts were only 1.99 times more likely to report this much studying. Similarly, Internet respondents to the open-ended format were 3.22 times more likely to report over 2.5 hours of computing per day while their telephone counterparts were only 1.60 times more likely to report this much computing. Additionally, although it did not reach significance, the interaction effect of the high range scale and survey mode did approach significance for the computing time question.

It is possible that the apparent mode effects in these models are actually due to mode-related non-response if respondents who spend a lot of time studying or using computers were more likely to respond to the Internet mode than to the telephone mode. In analyses not shown here this alternative explanation was explored. At the final phase of data collection 600 telephone non-responders were re-contacted and asked to complete the survey via the Internet mode; 144 (24%) did so. If high study or computing time respondents are more likely to respond to the Internet than the telephone mode these "converted" respondents should report higher study and/or computing time than those who initially responded to the telephone mode. However, this does not appear to be the case as the percentages of respondents reporting over 2.5 hours of each activity per day do not significantly differ between these two groups. These results suggest that the mode differences in these questions are not due to mode-related nonresponse bias.

Overall then, the results of the present experiment confirm the findings of Schwarz et al. (1985) and Rockwood et al. (1997) with respect to low and high range scales and extend those findings to the Internet mode. Internet respondents also rely on the information they gather from the context of the survey process, in this case from both the range of the scales provided and the

mode in which the survey is conducted, to help them formulate their responses to questions. As a result, it seems advisable to minimize context in an effort to get unbiased responses from respondents. To that end, Schwarz et al. (1985) recommended the use of open-ended response formats instead of scalar formats for behavior frequency questions. While this strategy may reduce response option context effects, the data presented here indicate that the open-ended format was more prone to mode context effects than the other formats. In fact, the most important finding of this experiment might be that the effects of different contexts do not always happen independently of each other. The complex relationships and potential tradeoffs between different types of survey context is a topic in need of much further study.

CONCLUSION

In everyday conversation, we encounter countless seemingly simple questions such as: What time is it? What should I do next? Are we headed in the correct direction? Do I have enough money? We oftentimes fail to explicitly realize that our understanding of these questions and the answers we provide to them are based on far more than the meaning of the words that form them; they also depend on the context of our everyday conversations. The task of answering a survey question shares common elements with the task of answering any one of the questions we face everyday. Context matters in our everyday and survey conversations; whether the question is, "Is it cold in here?" or "How many hours do you typically study per day?"

Within survey methodology the term context effects has long been used in a limited way to refer to question order effects. However, more recent work has begun to recognize that other survey characteristics, such as mode and presentation of response options, also play an important role in setting the survey context. We argue that the definition of context effects within survey methodology should be broadened to include multiple factors that influence the survey context;

however, this expanded definition should exclude factors external to the survey process itself that are out of the researcher's control to maintain the usefulness of the concept to survey methodologists.

In this chapter, we have reviewed four major types of context effects (survey mode, question order, response option order, and response scales). We focused on how these four types of context effects manifest themselves in Internet surveys and also discussed relevant research on paper and telephone surveys that enhances our understanding of the causes and consequences of context effects in Internet surveys.

One of the most important themes to emerge from the developing body of research using Internet surveys is that context effects can vary widely depending on the specific design features employed in the survey. The flexibility and ability to combine features from various other survey modes when designing Internet surveys means that it is not accurate to assume Internet surveys will perform similarly to paper or telephone surveys. Whether the context of an Internet survey more closely resembles a paper survey or a telephone survey, depends on what specific features are employed in its design and implementation. To understand context effects in Internet surveys it is important to isolate which features they share with other survey modes as well as new features and combinations that are unique to Internet surveys.

Even though the four types of context effects are summarized separately, the research example at the end of the chapter reveals how multiple types of context effects work together to influence the question-answer process. The experiment highlights the fact that different types of context effects oftentimes occur in conjunction with one another; sometimes reducing one type of context effect (e.g., using open-ended answer spaces to avoid the suggestive information in conventional scales) may increase another (e.g., response differences due to the contexts of web

and telephone modes). As surveys become more and more complex, understanding how different types of context effects occur independently and/or in conjunction with each other becomes important in reducing measurement error.

REFERENCES

Aquilino, W. S. (1994). Interview mode effects in surveys of drug and alcohol use: a field experiment. *Public Opinion Quarterly, 58, 210-240.*

Bayer, L. R. and Thomas, R. K. (2004). A Comparison of sliding scales with other scale types in online surveys. Paper presented at *RC33 6th International Conference on Social Methodology*. August, 2004.

Buchanan, T., & Reips, U.-D. (2001). Platform-dependent biases in online research: do mac users really think different? In K.J. Jonas, P. Breuer, B. Schauenburg, & M. Boos (Eds.), *Perspectives on internet research: concepts and methods.* Retrieved June 8, 2006 from http://www.psych.uni-goettingen.de/congress/gor-2001/contrib/contrib/buchanan-tom/buchanan-tom.

Chang, L. (2001). A comparison of samples and response quality obtained from RDD telephone survey methodology and internet survey methodology. Doctoral Dissertation. Department of Psychology. The Ohio State University, Columbus, OH.

Christian, L. M. (2003). The influence of visual layout on scalar questions in web surveys. Master's Thesis: Washington State University, Pullman, WA.

Christian, L. M. & Dillman, D. A. (2004). The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly, 68, 1, 57-80.*

Christian, L. M., Dillman, D. A., & Smyth, J. D. (Forthcoming). The effects of mode and format on answers to scalar questions in telephone and web surveys. In C. Tucker, M. Brick, E. DeLeeuw, B. Harris-Kojetin, L. Japec, P. Lavrakas, et al. (Eds.), *Telephone surveys.* New York: Wiley-Interscience.

Couper, M. P. (2000). Web surveys: a review of issues and approaches. *Public Opinion Quarterly, 64, 464-494.*

Couper, M. P. (2002). New technologies and survey data collection: challenges and opportunities. Presented at the International Conference on Improving Surveys, Copenhagen, Denmark. August, 2002.

Couper, M.P., Conrad, F. G., & Tourangeau, R. (2005). Visual effects in web surveys. Paper presented at the American Association of Public Opinion Research. Miami, Florida, May 2005.

Couper, M. P., Traugott, M. W., & Lamias, M. J. (2001). Web survey design and administration. *Public Opinion Quarterly, 65, 230-254.*

Couper, M. P., Tourangeau, R., Conrad, F. G., & Crawford, S. D. (2004a). What they see is what we get: response options for web surveys. *Social Science Computer Review, 22, 1, 111-127.*

Couper, M. P., Tourangeau, R., & Kenyon, K. (2004b). Picture this! exploring visual effects in web surveys. *Public Opinion Quarterly, 68, 255-266.*

De Leeuw, E. D. (1992). *Data quality in mail, telephone, and face to face surveys*. Amsterdam: TT Publications.

De Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21, 233-255.*

Dillman, D. A. (2000). *Mail and internet surveys: the tailored design method*. New York: John Wiley & Sons.

Dillman, D. A. & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. *Field Methods, 17, 1, 30-52.*

Dillman, D. A., Sangster, R. L., Tarnai, J., & Rockwood, T. H. (1996). Understanding differences in people's answers to telephone and mail surveys. In M. T. Braverman & J. K. Slater (Eds.), *New directions for evaluation series*, 70 (Advnces in survey research), (pp. 45-62). San Francisco: Jossey-Bass.

Dillman, D. A., & Tarnai, J. (1991). Mode effects of cognitively-designed recall questions: a comparison of answers to telephone and mail surveys. In P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, & S. Sudman (Eds.), *Measurement Errors in Surveys* (pp. 73-93). New York: Wiley.

Funke, F. & Reips U.-D. (2006) Visual analogue scales in online surveys: non-linear data categorization by transformation with reduced extremes. Presented at the 8th International Conference GOR06, March 21-22, 2006, Bielefeld, Germany. Retrieved 6/20/06 at http://www.frederikfunke.de/papers/gor2006.htm.

Grice, H. P. (1975). Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, 3: speech acts* (pp. 41-58). New York: Academic Press.

Hyman, H. H., & Sheatsley, P. B. (1950). The current status of American public opinion. In J. C. Payne (Ed.), *The teaching of contemporary affairs: twenty-first yearbook of the National Council for the Social Studies* (pp. 11-34). New York: National Education Association.

Javeline, D. (1999). Response effects in polite cultures: a test of acquiescence in Kazakhstan. *Public Opinion Quarterly, 63, 1-28.*

Krosnick, Jon A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5: 213-236.*

Krosnick, Jon A. (1999). Survey research. *Annual Review of Psychology*, *50: 537-567*.

Krosnick, J. A. & Alwin, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly*, *51, 201-219.*

Krysan, M., Schuman, H., Scott, L. J., & Beatty, P. (1994). Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly, 58, 381-399.*

Mason, R., Carlson, J.E., & Tourangeau, R. (1994). Contrast effects and subtraction in part-whole questions. *Public Opinion Quarterly, 58, 569-578.*

Mueller, J. E. (1970). Choosing among 133 candidates. *Public Opinion Quarterly, 34, 3, 395-402.*

Noelle-Neumann, E. (1970). Wanted: rules for wording structured questionnaires. *Public Opinion Quarterly, 34, 2, 191-201.*

Phillips, D. L., & Clancy, K. J. (1970). Response biases in field studies of mental illness. *American Sociological Review, 36, 512-514.*

Reips, U.-D. (2002) Context effects in web surveys. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online Social Sciences.* (pp. 69-79). Seattle, WA: Hogrefe & Huber.

Rockwood, T. H., Sangster, R. L., & Dillman, D. A. (1997). The effect of response categories in questionnaire answers: context and mode effects. *Sociological Methods and Research. 26, 1, 118-140.*

Schmidt, W. C. (forthcoming). Technical considerations when implementing online research. In A. Joinson, K. McKenna, T. Postmes, and U.-D. Reips. (Eds.) *The oxford handbook of internet psychology.* Oxford University Press.

Schuman, H. (1992). Context effects: state of the past/state of the art. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 5-20). New York: Springer-Verlag.

Schuman, H. & Presser, S. (1981). *Questions and answers in attitude surveys: experiments on question form, wording, and context.* New York: Academic Press.

Schwarz, N. (1996). *Cognition and communication: judgmental biases, research methods, and the logic of conversation.* Mahwah, NJ: Lawrence Erlbaum.

Schwarz, N., & Bless, H. (1992). Scandals and public trust in politicians: assimilation and contrast effects. *Personality and Social Psychology Bulletin, 18, 574-579.*

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45, 513-523.*

Schwarz, N., Grayson, C. E., Knäuper, B. (1998). Formal features of rating scales and the interpretation of question meaning. *International Journal of Public Opinion Research, 10, 2, 177-183.*

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

Schwarz, N., Hippler, H. J., Noelle-Neumann, E. (1992). A cognitive model of response order effects in survey measurement. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research* (pp. 187-201). New York: Springer-Verlag.

Schwarz, N., Knäuper, B. Hippler, H. J., Noelle-Neumann, E. & Clark, L. (1991a). Rating scales: numeric values may change the meaning of scale labels. *Public Opinion Quarterly, 55, 4, 570-582.*

Schwarz, N., Strack, F., & Mai, H. (1991b). Assimiltion and contrast effects in part-whole question sequences: a conversational logic analysis. *Public Opinion Quarterly, 55, 3-23.*

Smith, T. W. (1983). An experimental comparison between clustered and scattered scale items. *Social Psychology Quarterly,* 46, 163-168.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2004). How visual grouping influences answers to internet surveys. Social and Economic Sciences Research Center Technical Report #04-023, Washington State University, Pullman, WA 99164.

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006a). Effects of using visual design principles to group response options in web surveys. *International Journal of Internet Science, 1, 6-16.*

Smyth, J. D., Dillman, D. A., Christian, L. M., & Stern, M. J. (2006b). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly, 70, 66-77.*

Smyth, J.D., Dillman, D.A., & Christian, L.M. (2006c). Does 'yes or no' on the telephone mean the same as 'check-all-that-apply' on the web?" Paper presented at the Second International Conference on Telephone Survey Methodology. Miami, Florida, January 12-15, 2006.

Strack, F. (1992). "Order effects" in survey research: activation and information functions of preceding questions. In N. Schwarz & S. Sudman (Eds.), *Context Effects in social and psychological research* (pp. 23-34). New York: Springer-Verlag.

Strack, F., Schwarz, N., & Wänke, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition, 9, 111-125.*

Sudman, S., Bradburn, N., & Schwarz, N. (1996). *Thinking about answers*. San Francisco, CA: Josey-Bass.

Taylor, H., Krane, D., and Thomas, R. K., (2005). How does social desirability affect responses?: differences in telephone and online surveys. Paper presented at the American Association for Public Opinion Research. Miami, Florida, May 2005.

Tourangeau, R. (1992). Context effects on responses to attitude questions: attitudes as memory structures. In N. Schwarz & S. Sudman (Eds.) *Context Effects in Social and Psychological Research* (pp. 35-47). New York: Springer-Verlag.

Tourangeau, R., Couper, M. P., & Conrad, F. (2004). Spacing, position, and order: interpretive heuristcs for visual features of survey questions. *Public Opinion Quarterly, 68, 3, 368-393.*

Tourangeau, R., Couper, M. P., & Conrad, F. (2006) Color, labels and interpretive heuristics for response scales. Paper presented at the American Association of Public Opinion Research, Montreal, Quebec, Canada, May 18-21, 2006.

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin, 103, 299-314.*

Tourangeau, R., Rasinski, K., & Bradburn, N. (1991). Measuring happiness in surveys: a test of the subtraction hypothesis. *Public Opinion Quarterly, 55, 255-266.*

Tourangeau, R., Rasinski, K., Bradburn, N., & D'Andrade, R. (1989). Carry-over effects in attitude surveys. *Public Opinion Quarterly, 53, 495-524.*

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

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: the impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60, 275-304.*

Witte, J. C., Pargas, R. P., Mobley, C., & Hawdon, J. (2004). Instrument effects of images in web surveys: a research note. *Social Science Computer Review*, *22, 363-369.*

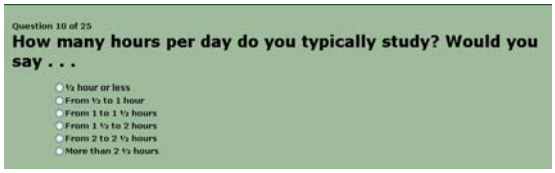
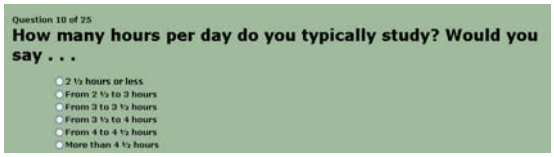
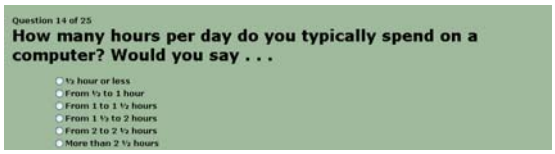
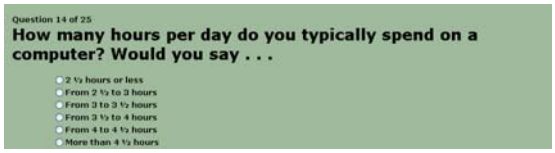
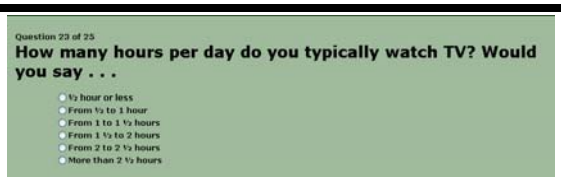
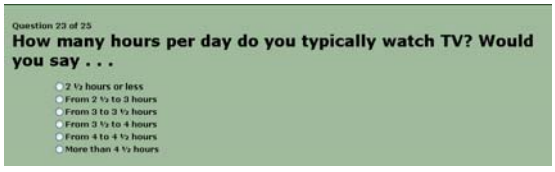
Table 1: Percent of Respondents Reporting Spending up to 2.5 Hours and Over 2.5 Hours/Day Studying, Computing, and Watching TV by Mode

	WEB SAMPLE			TELEPHONE SAMPLE		
	Low Range	High Range	Significance Tests	Low Range	High Range	Significance Tests
Study						
≤ 2 ½ hours	70.2	28.9	X^2= 123.38	65.8	26.5	X^2= 96.60
> 2 ½ hours	29.8	71.1	p = .000	34.2	73.6	p = .000
Computer						
≤ 2 ½ hours	70.8	34.1	X^2= 96.75	74.6	48.9	X^2= 43.56
> 2 ½ hours	29.2	66.0	p = .000	25.4	51.1	p = .000
TV						
≤ 2 ½ hours	91.6	78.8	X^2= 22.99	93.6	82.3	X^2= 18.57
> 2 ½ hours	8.4	21.3	p = .000	6.4	17.7	p = .000
Response Rate	**351/600 (58.5%)**	**379/600 (63.2%)**		**311/536 (58.0%)**	**311/536 (58.0%)**	

Table 2: Logistic Regression Results of the Effects of Response Option Format and Survey Mode

	ODDS RATIO	STANDARD ERROR	Z	P
Q10: Study Time				
High Scale	5.35	.941	9.54	.000
Open-Ended	1.99	.326	4.18	.000
Mode (web)	0.82	.137	-1.21	.227
High * Mode	1.08	.260	0.34	.736
Open * Mode	1.63	.373	2.13	.033
N				2,006
Q14: Computer Time				
High Scale	3.07	.531	6.50	.000
Open-Ended	1.60	.279	2.67	.008
Mode (web)	1.21	.212	1.09	.277
High * Mode	1.52	.361	1.80	.072
Open * Mode	1.62	.384	2.04	.042
N				2,004
Q23: TV Time				
High Scale	3.13	.859	4.15	.000
Open-Ended	2.82	.780	3.75	.000
Mode (web)	1.34	.403	0.96	.338
High * Mode	0.94	.338	-0.17	.865
Open * Mode	1.05	.380	0.13	.898
N				1,996

Figure 1: Experimental Treatments as They Appeared in Each Mode

	WEB SCREENSHOTS	TELEPHONE WORDING
Version 1		How many hours per day do you typically study? Would you say . . . A half hour or less From one half to one hour From one to one and a half hours From one and a half to two hours From two to two and a half hours Or more than two and a half hours
Version 2		How many hours per day do you typically study? Would you say . . . Two and a half hours or less From two and a half to three hours From three to three and a half hours From three and a half to four hours From four to four and a half hours Or more than four and a half hours
Version 3		How many hours per day do you typically study?
Version 1		How many hours per day do you typically spend on a computer? Would you say . . . A half hour or less From one half to one hour From one to one and a half hours From one and a half to two hours From two to two and a half hours Or more than two and a half hours
Version 2		How many hours per day do you typically spend on a computer? Would you say . . . Two and a half hours or less From two and a half to three hours From three to three and a half hours From three and a half to four hours From four to four and a half hours Or more than four and a half hours
Version 3		How many hours per day do you typically spend on a computer?
Version 1		How many hours per day do you typically watch TV? Would you say . . . A half hour or less From one half to one hour From one to one and a half hours From one and a half to two hours From two to two and a half hours Or more than two and a half hours
Version 2		How many hours per day do you typically watch TV? Would you say . . . Two and a half hours or less From two and a half to three hours From three to three and a half hours From three and a half to four hours From four to four and a half hours Or more than four and a half hours
Version 3		How many hours per day do you typically watch TV?