Survey Research Methods (2007)
Vol.1 , No.3 , pp. 121-138
ISSN 1864-3361
http://www.surveymethods.org

Visual Design, Order Effects, and Respondent Characteristics in a Self-Administered Survey

Michael J. Stern
Oklahoma State University

Don A. Dillman
Washington State University

Jolene D. Smyth
University Nebraska-Lincoln

Recent survey design research has shown that small changes in the structure and visual layout of questions can affect respondents' answers. While the findings have provided strong evidence of such effects, they are limited by the homogeneity of their samples, in that many of these studies have used random samples of college students. In this paper, we examine the effects of seven experimental alterations in question format and visual design using data from a general population survey that allows us to examine the effects of demographic differences among respondents. Results from a 2005 random sample mail survey of 1,315 households in a small metropolitan region of the United States suggest that the visual layout of survey questions affects different demographic groups in similar ways.
Keywords: Visual Design Theory, Demographic Differences, Measurement Errors, Self Administered Surveys

Introduction

For many decades, survey researchers have known that minor changes in question or response option wording affect respondents' answers. However, until the 1990s, researchers knew very little about how visual layout changes affect responses. In 1993, Smith provided five examples of how small visual design alterations significantly affected responses in the General Social Survey (see Smith 1993). Based, in part, on Smith's work, researchers have developed a theoretical paradigm that draws on linguistics and Gestalt psychology (Jenkins and Dillman 1997) to articulate a theory of how respondents incorporate visual design elements into the process of understanding and answering survey questions.

Based on these beginnings, a growing body of empirical work now provides a foundation for visual design theory. This research has addressed issues such as how the size and type of answer spaces influences how much and what type of information is entered into them (Christian and Dillman 2004; Couper, Traugott and Lamias 2001), how check-all-that-apply question formats differ from forced choice formats (Smyth, Dillman, Christian and Stern 2006), and how the order of response options affects answers (Tourangeau, Couper and Conrad 2004). However, the research has focused almost entirely on overall effects at the expense of understanding whether (and how) changes in visual design might differentially affect individuals with varied demographic characteristics.

The purpose of this paper is to examine how the effects of visual design may differ based on age, education, and sex of respondents by replicating and extending previous experiments to a general public population. The results of seven visual design experiments are reported for multiple demographic groups. The data for these experiments come from a 2005 random sample mail survey of 1,315 residents of a small metropolitan region in the Western United States. The experiments were embedded in two versions of a questionnaire concerning community satisfaction, technology use, and civic participation. While these seven experiments do not provide us with a large number of replications, our intent here is to address whether there is a basis for further exploring the effects of visual layout across varied demographic groups.

Theoretical Background

Visual Design Theory

In 1984, Tourangeau outlined the four general cognitive steps respondents undergo while answering questions. These steps include comprehending the question, recalling the relevant information, making a judgment, and providing a response. Though this model was originally outlined for interview situations, Redline and Dillman (2002) expanded it to address self-administered surveys by adding one additional stage to the answering process. Specifically, they prefaced the original four steps with a "perception" stage where the respondent takes in all of the visual cues that make up the survey instrument or item.

Research by Schwarz (1996) provides the conceptual basis for why these visual cues are so important in the answering process. He suggests that respondents become *cooperative*

Contact information: Michael J. Stern, Department of Sociology Oklahoma State University CLB 006 Stillwater, OK 74078-4062 (mstern@okstate.edu).

communicators when engaging in the act of completing a survey. That is, conversational norms govern the way that respondents interact with the survey instrument. Schwarz (1996) articulates four underlying maxims involved in the conversational norm, including that the information be clearly expressed and understandable to the intended audience, contributions to the conversation be relevant, contributions to the conversation be informative, and contributions be truthful. The survey instrument serves as the researcher's side of the conversation and should conform to the maxims. Because the researcher is not present during the conversation, the respondent uses all of the visual cues and the layout to guide them in the conversation (Christian and Dillman 2004).

Christian and Dillman (2004) have demonstrated empirically that in addition to verbal language effects, the visual features of questionnaires influence people's answers to survey questions. These additional features include *numeric* language (numbers in the queries and answer categories), *graphical* language (size, spacing, and location of information on the page), and *symbolic* language (e.g. arrows and answer boxes). Survey researchers use these languages both independently and in concert to communicate information to respondents completing self-administered questionnaires.

Tourangeau, Couper, and Conrad (2004) argue that there are five specifically visual interpretive heuristics that respondents follow when evaluating the visual layout of survey questions and the violation of these heuristics can affect the answering process. The five heuristics include: 1) middle means typical; 2) left and top means first; 3) near means related; 4) up means good; and 5) items that are close to each other are similar. As the authors explain, "each heuristic assigns a meaning to a spatial or visual cue" (Tourangeau, Couper and Conrad 2004:370; see also Tourangeau, Couper and Conrad 2007). They also provide evidence to support the underlying principle that when the visual presentation of the question does not conform to the expectation of the respondent, proper interpretation of the meaning of the question can be lost.

A number of Gestalt psychology grouping principles can also help us understand how the presentation of information might affect its interpretation. These principles can be summarized as the Law of Similarity (objects of the same size, brightness, color, or shape are more easily seen together), the Law of Proximity (objects close to each other are grouped together), and the Law of Pragnanz (figures with simplicity, regularity, and symmetry are easier to follow) (Dillman 2007). Jenkins and Dillman (1997) have addressed how these visual and spatial cues affect how respondents interpret information.

One of the lessons we have learned through the research used to develop these perspectives is that respondents often answer questions in ways that were unanticipated by the researcher for reasons having to do with visual layout, rather than their opinion or position on a particular question or issue (Dillman 2007). One of the most well-documented theories concerning these response errors is that of "satisficing" (Krosnick and Alwin 1987). Satisficing occurs when respondents fail to expend the necessary energy needed to answer a question and thus fail to provide an optimal response. While there are reasons why a respondent may satisfice that are outside of the researcher's control, one area that can be controlled by the researcher is the visual design of the question. The more respondents are cognitively taxed by the visual design of questions – including the violation of the interpretive heuristics – the more likely they are to satisfice, thus increasing measurement error (Israel 2006).

Krosnick (1991; 1999) provides evidence for satisficing by showing that respondents tend to choose earlier items in a list because they find the first position that they can reasonably agree with and consider it a satisfactory answer, rather than reading and considering each response option separately. When this occurs in self-administered survey questions, a pattern of primacy is expected.[1] These effects are not only seen in scalar questions, but also, for example, in mark-all-that-apply questions where the respondent is asked to select each option that pertains to them from a list (Smyth et al. 2006).

Respondent Characteristics

Despite growing empirical support for visual design theory (Dillman 2007), the theory itself and the empirical research may be overstated because it has yet to consider demographic variation. In fact, the theory concerning visual layout and design in self-administered surveys is virtually devoid of any reference to respondent characteristics. As a result, the empirical tests to date have not taken into account how various question formats may affect respondents differently based on their demographic characteristics (e.g. education, age, and sex). Furthermore, the research that has examined demographic differences in respondent behavior was developed independently of theories of visual design.

The most common reasoning provided for expecting response effects among individuals of lower education has been their level of cognitive sophistication. Krosnick (1992) argues that respondents with lower levels of education are less likely to do the necessary cognitive work needed for evaluating response categories or fully comprehending the question stem; thus, they are likely to satisfice as evidenced by primacy. However, the research attempting to test this proposition has produced mixed results. For example, Ayidiya and McClendon (1990:244) found that in self-administered mail surveys there were response order effects; however, there was no "reliable evidence" for systematic differences among individuals of lower education. In contrast, Knäuper (1999), Knäuper, Belli, Hill and Herzog (1998), and Krosnick, Narayan and Smith (1996) have all reported that response order effects are more prevalent among people with lower levels of education.

Knäuper (1999) and Knäuper et al. (1998) argue that limits in cognitive ability are present among individuals of older age (60 and older). Using cognitive psychology as the basis of their argument, they suggest that as individuals age, their diminished ability to comprehend questions (which

[1] It should be noted that while patterns of primacy are often found in self-administered surveys, patterns of recency, where the last options in a list are more likely to be selected, are frequently found in interview situations (e.g., telephone surveys).

could be largely affected by visual design) and recall memories (perhaps not as affected by visual design) makes them more susceptible to response order effects and other question problems. This research further suggests that age is a more powerful predictor of response order effects than educational attainment. However, Knäuper (1999) concedes that future theoretical work is needed concerning response effects in older respondents.

Fewer studies have examined response effects by the sex of the respondent. However, studies do suggest that such research is needed. For example, research on the use of a "don't know" category has consistently shown that women are more likely to provide this response than their male counterparts in both adult and adolescent samples (see Rapoport 1982 for a review). Rapoport (1982) suggests that this effect could be the enduring result of differential socialization in that the effect is still persistent after controlling for issues such as question subject knowledge. That is, women are socialized to have lower levels of opinionation than men. This perspective has been supported in research that examines the number of affirmative answers provided in forced choice questions (Glenn 1969).[2] The implications for response effects could be very important. If women and men differ in their level of engagement in survey items, then theories of visual design must address this issue. One way to do this is by testing whether the effects of visual design are equally recognizable across men and women. Notably, the research summarized here precedes the development of the cooperative communicator model of Schwarz (1996) and is in need of further explication.

Research and theory concerning the visual design of self-administered survey questions suggests that respondents actively utilize visual cues (verbal, symbolic, numeric, and graphical) embedded in the design of these questions to guide them in the response process. However, research has not yet addressed in detail how respondents who differ by age, education, and sex may be differentially affected by visual design. Further, the research concerning demographic differences in respondent behavior has not been guided by any unifying theory that seeks to explain the effects reported (Knäuper 1999) and has not yet taken into consideration potential relationships between respondent characteristics and visual design. The experiments described in the following sections seek to provide a basis for unifying these literatures by testing for the effects of visual design in a self-administered mail survey and analyzing whether those effects vary by respondent education, age, and sex. We designed each of the experiments used in this paper to either specifically replicate an earlier study or to test a theoretical postulate relevant to the visual design of self-administered surveys.

Procedures

The experimental comparisons were embedded in two versions of a self-administered mail survey designed to ask residents about their community satisfaction, technology use, and civic participation. Respondents were randomly assigned to one of the two versions. The overall design of the surveys was very similar with only a slight color difference (white and off-white) between the two versions. The 10-page questionnaire with 75 queries for the average respondent was conducted in the winter of 2005. A random sample of 2,000 households was drawn with 1,315 of the households completing and returning the survey, culminating in a response rate of 65.75% (AAPOR [2006] RR1).

The implementation design used three mail contacts (Dillman 2007). The first contact contained a personally signed cover letter explaining the survey's goals and content, a self-addressed stamped envelope, a two-dollar token incentive, and the questionnaire. Two weeks later a follow-up post card was sent to all respondents that thanked those who had responded and encouraged those who had not to please do so. Finally, about two weeks after the post card was sent, a replacement questionnaire was sent to individuals who had not yet responded along with a personally signed letter encouraging them to fill out the questionnaire.

Based on previous research concerning demographic differences in respondent behavior and survey response effects (Knäuper 1999; Krosnick, Narayan and Smith 1996; Rapoport 1982), individuals were partitioned into demographic groups. To examine age, respondents were classified as 60 years old and older or under the age of 60. For education, in keeping with previous studies (Krosnick, Narayan and Smith 1996), individuals were divided into two groups based on presence or absence of a college degree. Sex was partitioned as women or men.

Experiments and Results

Experiment 1: Use of the Number Box versus Polar Point Scale

Design. We begin our analysis by replicating and extending a combined graphical and symbolic language manipulation test. In the original test, Christian and Dillman (2004) examined the effects of two types of non-verbal language (graphical and symbolic) on responses to scalar questions. Respondents provided dramatically different responses based on the version of the question they received. Specifically, respondents provided much higher values (5 = very dissatisfied) in the number box question (where the respondent writes in the number corresponding to their answer) when compared to the polar point (labeled endpoints) versions. It appeared that the removal of the non-verbal languages might have caused respondents to become confused about the direction of the scale in the number box version. That is, the need to carry the information from the question stem to the answer box proved difficult. We sought to examine whether the findings would be 1) replicated and 2) if the results would be reliable across demographic groups.

Similar to Christian and Dillman (2004), we compared a scalar question with a polar point scale and a number box. In addition to the original test, we incorporated a modest extension by including a "don't know" response (Figure 1). The question asked, "On a scale from 1 to 5, how much do

[2] Glenn (1969) does not explicitly discuss this point in his paper, but his data do show this trend.

you think the organizations, clubs, or local groups that exist in the Lewiston/Clarkston area contribute to the quality of life of local residents?" One polar point was represented by "a lot" and the other by "not at all."

We first hypothesize that our findings will replicate those of Christian and Dillman (2004) in that the polar point and number box formats will elicit different responses for the same question and that the direction of this relationship will be the same (e.g. more negative responses for the number box question). In addition, we predict that more respondents will choose the "don't know" option when given the number box version because there is more respondent burden associated with using the number box. Because of this cognitive difficulty and based upon the work of Knäuper (1999) and Krosnick, Narayan and Smith (1996), we expect individuals aged 60 and older or with less than a college degree to be the most affected by the difference in response formats. In addition, we expect these groups to be more likely to choose the "don't know" option.

Findings. Table 1 reports the results comparing the number box to the polar point version of the question, both with and without the "don't know" option. As predicted, there were significant differences in the response distributions between the formats (χ^2= 36.44; p <.001). The direction of the relationship shows that respondents were significantly more likely to provide negative responses (e.g. higher numbers) for the number box as compared to the polar point version of the question (the means were 2.48 and 2.16, respectively). In addition, respondents who received the number box version were significantly more likely to provide a "don't know" response compared to respondents who received the polar point version of the question (23.9% and 11.3%, respectively; χ^2 = 71.18; p < .001).

As shown in Table 2, across all demographic groups there were statistically significant differences in responses based on the version of the question respondents received. In addition, for each demographic group more negative responses were provided for the number box version of the question, meaning that respondents provided larger numbers. The mean differences were significant for all but one of the demographic groups (men), where the mean difference approached significance (p=.06). Furthermore, the likelihood of selecting the "don't know" response was significantly higher for each demographic group when they received the number box version (all differences were statistically significant).

Conclusions. In terms of the demographic hypothesis, there was mixed support. First, the data show that respondents over 60 years of age or possessing less than a college degree appeared to be no more likely than other groups to be influenced by the visual layout of the question. However, when given the number box version, both of these demographic were more likely to provide a "don't know" response than the other demographic groups. Thus, part of the explanation for not seeing a greater overall difference for these groups in the response distributions may have to do with the fact that so many of these individuals chose a "don't know" response. Nonetheless, these findings seem to lend considerable support to the findings of Christian and Dillman (2004), which suggest that responses that are more negative are provided for number box formats, inasmuch as our results show this pattern across demographic groups.

Experiment 2: Manipulating the Size of the Answer Space in Open-Ended Questions

Design. In our second experiment, we examined whether the size of answer spaces in open-ended questions affected all respondents in similar ways. Several previous studies have examined the effects of manipulating the size of the answer space in open-ended questions. For example, Israel (2006) and Christian and Dillman (2004) conducted graphical experiments using various sized answer spaces for open-ended questions. The larger spaces elicited longer responses than the smaller boxes for all items in both experiments. Thus, they concluded that the large box suggests to the respondent that the surveyor wants more information than when small boxes are used.

In this experiment, two different sized boxes were used. The experiment was part of a branching question that asked respondents first whether or not there was any particular change that they believed would make the community a better place to live. If they responded "yes," they were asked to elaborate about what they would change. In one version of the questionnaire the box was 2 inches high and 6.5 inches wide. In the alternate version, the box was 1 inch high and 6.5 inches wide.

Consistent with previous research, we expect the bigger box to yield longer responses for all demographic groups. To clarify, this is a hypothesis about direction of effect; we do not expect all demographic groups to provide answers of the same length. Rather, we are predicting that each demographic group will provide more words when a bigger space is provided than they provide when a smaller space is provided.

Findings. Table 3 shows the mean number of words provided for each size of the box. Overall, the results are consistent with previous research and the hypothesis by showing that the larger box yielded longer responses (17.17 words) than the smaller box (14.95 words). However, the difference failed to reach statistical significance. Turning to the respondent characteristics, as hypothesized, each group provided longer responses in the larger space. However, only among respondents with less than a college degree (p=.008) and men (p=.07) did we find differences that were significant or approached significance.

Conclusions. Regarding the size of the answer space, each demographic group wrote more in the larger space; thus, the graphical language affected all respondents in a similar way. These findings support previous research by showing that the size of the box influences the amount that respondents contribute across demographic groups. However, it should be noted that some groups were more likely to provide longer and more detailed responses. For example, respondents over 60 years of age, with less than a college degree, and women

Table 1: Response Distributions for Polar Point Versus Number Box Questions Without and With the Don't Know Option Included

	Without the Don't Know Option		With the Don't Know Option	
	Polar Point (%)	Number Box (%)	Polar Point (%)	Number Box (%)
1) A lot	34.4	31.4	30.5	23.9
2)	27.3	21.3	24.2	16.2
3)	28.5	24.4	25.3	18.6
4)	7.0	14.1	6.2	10.8
5) Not at all	2.8	8.8	2.5	6.7
Don't know	–	–	11.3	23.9
Mean	2.16	2.48		
n	572	488	645	641
Overall Chi-Square	$\chi^2 = 36.44, p < .001$		$\chi^2 = 71.18, p < .001$	
t-test	t=-4.08, $p < .001$		–	–

Table 2: Response Distributions for Polar Point Versus Number Box Questions Without the Don't Know Response Option by Demographic Group

Response Distributions by Demographic Without the Don't Know Response Option

	60 years +		< 60 years		College +		< College		Women		Men	
	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)
1) A lot	39.7	32.8	32.7	32.2	33.14	35.6	35.9	26.2	38.7	32.1	29.9	33.0
2)	28.4	19.8	27.5	22.7	33.1	18.0	19.1	24.6	24.2	20.6	31.8	21.8
3)	22.7	21.6	30.2	24.8	26.7	24.9	31.8	24.6	26.6	24.8	30.3	22.9
4)	5.7	16.4	7.1	14.0	4.6	14.9	9.1	13.1	7.8	13.3	5.2	16.0
5) Not at all	3.5	9.5	2.5	6.3	2.1	6.6	4.1	11.5	2.7	9.2	2.8	6.4
Mean	2.05	2.50	2.19	2.39	2.09	2.39	2.26	2.59	2.12	2.47	2.19	2.41
n	141	116	324	286	326	289	220	191	256	218	211	188
Overall Chi-Square	$\chi^2 = 13.32$, $p = .010$		$\chi^2 = 15.17$, $p = .004$		$\chi^2 = 38.15$, $p = .000$		$\chi^2 = 13.36$, $p = .004$		$\chi^2 = 14.24$, $p = .007$		$\chi^2 = 19.93$, $p = .001$	
t-test	$t = -2.97$, $p = .003$		$t = -2.19$, $p = .029$		$t = -3.25$, $p = .001$		$t = -2.69$, $p = .007$		$t = -3.17$, $p = .002$		$t = -1.88$, $p = .061$	

Selection of the Don't Know Response Option by Demographic Group

	60 years +		< 60 years		College +		< College		Women		Men	
	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)	Polar Point (%)	# Box (%)
Don't Know	11.9	26.6	8.7	20.3	7.4	17.2	16.3	32.0	10.8	24.6	8.3	20.0
n	160	158	355	359	352	349	263	281	287	289	230	235
Overall Chi-Square	$\chi^2 = 24.29$, $p = .001$		$\chi^2 = 34.42$, $p = .001$		$\chi^2 = 53.67$, $p = .001$		$\chi^2 = 33.38$, $p = .001$		$\chi^2 = 32.88$, $p = .001$		$\chi^2 = 33.02$, $p = .001$	

20. On a scale from 1 to 5, how much do you think the organizations, clubs, or local groups that exist in the Lewiston/Clarkston area contribute to the quality of life of local residents?

- ❑ 1 A lot
- ❑ 2
- ❑ 3
- ❑ 4
- ❑ 5 Not at all

- ❑ Don't know

20. On a 1 to 5 scale where 1 means a lot and 5 not at all, how much do you think the organizations, clubs, or local groups that exist in the Lewiston/Clarkston area contribute to the quality of life of local residents? You may use any number between 1 and 5.

☐

- ❑ Don't know

Figure 1. Polar Point Versus Number Box with a Don't Know Option

Table 3: Mean Differences in the Number of Words Provided for the Entire Sample and by Demographic Group Based on the Size of the Answer Box

	Format	n	Mean	Difference	t-test	*p*
Overall	Big Box	456	17.17	2.22	1.82	.069
	Small Box	466	14.95			
60+ yrs.	Big Box	97	18.40	2.02	0.74	.457
	Small Box	102	16.38			
<60 yrs.	Big Box	284	17.38	2.34	1.46	.144
	Small Box	278	15.04			
<College	Big Box	179	18.74	5.39	2.69	.008
	Small Box	171	13.35			
College +	Big Box	270	16.42	0.29	0.18	.856
	Small Box	276	16.13			
Male	Big Box	167	16.07	3.07	1.78	.077
	Small Box	162	13.00			
Female	Big Box	219	18.59	1.21	0.60	.546
	Small Box	218	17.38			

all provided responses that were one to two words longer than their comparison group, regardless of the size of the box. While beyond the scope of this paper, this is an issue that should be given greater attention in future research.

Experiment 3: Forced Choice versus Check-All-That-Apply Formats

Design. For our third test, we included one experimental comparison of a forced choice versus a check-all-that-apply format. Recent research by Smyth et al. (2006) shows that visual manipulations can help us understand the impact of using one multiple answer question format over another. Specifically, the research shows that respondents provide more affirmative answers to forced choice formats where the respondent is asked to choose either affirmative (e.g. "yes") or a negative response (e.g. "no"), when compared to the more traditional check-all-that-apply format where the respondent marks all the responses they feel apply to them. In addition, their research provided evidence that respondents spend, on average, more time answering the forced choice format. Thus, the inclusion of additional verbal and symbolic elements seems to affect the amount of energy respondents spend on a question and their subsequent responses.

Our experimental test question asked whether respondents had engaged in any of a list of seven activities or none of the above (see Figure 2 for exact wording). Based on previous research (Smyth et al. 2006), we expect the overall response distributions to show that the forced choice format

provides more affirmative responses than the check-all-that-apply formats across demographic groups. That is, all groups should be affected equally.

Findings. Table 4 shows the response distributions for the overall sample. Consistent with previous research, each of the response options was chosen more often in the forced choice than the check-all format (albeit not always at significant levels) with the exception of the non-substantive "none of the above" option. Furthermore, the overall mean number of categories endorsed is larger in the forced choice that the check-all-that-apply format (3.20 vs. 3.01, respectively) at a statistically significant level ($p < .05$). Interestingly, the "none of the above" category is checked by a greater number of respondents in the check-all-that-apply format. This finding seems to further support previous research suggesting that when respondents received the forced choice format they read each option and found at least one response they could mark affirmatively, thus ruling out the "none of the above" category. Conversely, when respondents were provided with the check-all-that-apply format they may not have considered each option separately and instead quickly went through to the "none of the above" response, which appeared last.

Turning to the demographic characteristics, Table 5 shows the mean number of responses checked affirmatively for each group. For every demographic subgroup, the mean number of options endorsed in the forced choice format was higher than in the check-all-that-apply format; in addition, for all but two of the comparisons the differences were significant or approaching significance ($p < .07$). Men and respondents over the age of 60 did not show the same degree of difference as the other groups, although for both groups the mean number of affirmative answers was still higher in the forced choice format than the check-all-that-apply format. With the exception of these two groups, there is reasonable support for the hypothesis that the manipulation affects groups in the same direction.

Conclusions. In comparing the check-all-that-apply and forced choice formats, every demographic group marked more options affirmatively in the forced choice format. Of particular interest, these data lend support to recent empirical studies concerning how respondents answer multiple answer questions. Smyth et al. (2006) have shown that respondents spend more time answering forced choice questions than check-all-that-apply questions because in the forced choice format each response option must be evaluated and answered. Thus, respondents engage in deeper processing with the forced choice format. The fact that 6.3% of respondents chose "none of the above" in the check-all-that-apply question compared to less than 2% in the forced choice format suggests that respondents clearly considered each option in the forced choice format and thus engaged in deeper processing. This finding was true across demographic groups.

Experiments 4 & 5: Response Order Manipulations in Scalar Questions

Design. The order of response options figures prominently into the interpretive heuristics that respondents use in the answering process. For example, Tourangeau, Couper, and Conrad's (2004; 2007) heuristics show that in repeated tests, not only do respondent answers differ by the order of response options, but also when respondents receive versions of the questions where the response options violate the heuristic, they take much longer to complete the answering process.

For our first set of response order experiments, we used two behavior-based scalar questions (Figures 3 & 4). The questions asked "About how often do you travel more than 100 miles outside the area?" and "How often do you use the Internet to access the web or for email?" The question concerning Internet usage was based on a subset of 1,269 individuals who answered affirmatively to a filter question asking whether they used the Internet. For both questions, the response options in one version started with the high end of the behavior scale (e.g. everyday) and the options in the second version started with the low end (e.g. once a week or less). This reversal of visual presentation allowed us to test whether primacy occurred in these questions and then whether any effects were stronger among individuals with certain demographic characteristics. Based on Krosnick's (1999) satisficing theory and the work of Tourangeau, Couper, and Conrad (2004), we hypothesize that primacy will occur in both questions. Furthermore, based on the work of Knäuper (1999) and Krosnick, Narayan, and Smith (1996), we expect individuals aged 60 and older or with less than a college degree to be more affected by the reversal in presentation than those under 60 or respondents with a college degree or more. Men and women were not expected to be differentially affected by the reversing of response options.

Findings. Table 6 shows that the reversal of response categories did not result in differences in response distributions for the question asking about traveling outside the local area. Furthermore, analyses not shown here revealed that there were no significant differences in response distributions among any of the demographic groups. It may be that people were easily able to recall this information and thus not as susceptible to the direction of the response options.

In contrast, the reversal of response options in the Internet use question resulted in significant differences in the response distributions (Table 6 second panel). When "everyday" was located first in the list it was selected at much higher rates than when it appeared last (56.3% and 37.5%, respectively). The reason for this difference appears to be the similarity in the response options "everyday" and "nearly everyday." In version 1, where "nearly everyday" appeared below "everyday," 15.6% of respondents chose it; whereas, in version 2 where it appeared before "everyday," 28.8% of respondents chose the response option. Additionally, when the last response option "once a week or less" was listed first (in the reverse condition), it was selected by 15.3% of the respondents, but when it was listed last, only 12.8% chose it. The differences

10. Have you ever engaged in any of the following activities in order to influence a decision concerning your community?

	Yes	No
Voted in the 2004 general election……..	❑	❑
Attended public hearings…….	❑	❑
Attended a public meeting to discuss public issues/problems…….	❑	❑
Signed a petition …….	❑	❑
Participated in a strike……..	❑	❑
Donated money to community group……	❑	❑
None of the above……	❑	❑

10. Have you ever engaged in any of the following activities in order to influence a decision concerning your community? (Check all that apply)

Voted in the 2004 general election……..	❑
Attended public hearings…….	❑
Attended a public meeting to discuss public issues/problems…….	❑
Signed a petition …….	❑
Participated in a strike…….	❑
Donated money to community group……	❑
None of the above……	❑

Figure 2. Forced Choice versus Check-All-That-Apply Formats

Table 4: Comparison Between Forced Choice and Check-All-That-Apply Formats Shown in Figure 4 for Percentage of Responses Marked Affirmatively

	Forced Choice	Check-all	Significance Tests	
	(%)	(%)	χ^2	p
Voted in the 2004 general election	84.1	83.5	0.09	.764
Attended public hearings	41.5	37.8	1.89	.168
Attended a public meeting	37.0	34.0	1.29	.255
Signed a petition	70.4	65.7	3.30	.069
Participated in a strike	3.7	2.7	0.905	.342
Donated money to community group	81.3	70.9	19.58	.000
None of the above	1.8	6.3	16.35	.000
n	646	656		
Mean number of items endorsed	3.20	3.01	$t = 2.48$	$p = .013$

7. About how often do you travel more than 100 miles outside the area?

- ❑ Once or more a week
- ❑ Once or more a month
- ❑ Once or more a year
- ❑ About once a year
- ❑ Less than once a year

7. About how often do you travel more than 100 miles outside the area?

- ❑ Less than once a year
- ❑ About once a year
- ❑ Once or more a year
- ❑ Once or more a month
- ❑ Once or more a week

Figure 3. Response Order Manipulations in Scalar Questions

Table 5: Comparison Between Forced Choice and Check-All-That-Apply Formats Among Each of the Demographic Groups for Number of Options Marked Affirmatively

	Format	n	Mean	Difference	t-test	*p*
Overall	Forced Choice	646	3.20	0.19	2.48	.013
	Check-all	656	3.01			
60+ yrs.	Forced Choice	159	3.35	0.22	1.45	.148
	Check-all	161	3.13			
< 60 yrs.	Forced Choice	360	3.26	0.20	1.88	.060
	Check-all	358	3.06			
< College	Forced Choice	461	3.08	0.17	1.83	.068
	Check-all	448	2.91			
College +	Forced Choice	171	3.60	0.29	2.08	.039
	Check-all	178	3.31			
Male	Forced Choice	235	3.36	0.13	1.03	.302
	Check-all	230	3.23			
Female	Forced Choice	291	3.21	0.24	2.13	.035
	Check-all	291	2.97			

26. How often do you use an Internet connection to access the web or for email?

- ❑ Every day
- ❑ Nearly every day
- ❑ A few times per week
- ❑ Once a week or less

26. How often do you use an Internet connection to access the web or for email?

- ❑ Once a week or less
- ❑ A few times per week
- ❑ Nearly every day
- ❑ Everyday

Figure 4. Response Order Manipulations in Scalar Questions

in the response distributions were significant for both the chi-square ($\chi^2 = 41.57$; $p < .001$) and the independent samples t-test ($t = -3.93$; $p < .001$).

Table 7 shows the analyses of the response distributions for the Internet usage question by demographic groups. The top two rows of the table show the response distributions of the two response categories (everyday and nearly everyday) for which the overall distributions showed significant differences by version. Across all the demographic groups, the same relationship as seen in the overall distributions is found. In other words, regardless of age, educational attainment, or sex, the everyday category was chosen significantly more when it appeared first than when it appeared last and the nearly everyday response was chosen at significantly higher levels when it appeared before the everyday option (Chi-Square tests show p <.01 for each group). However, the effect seems to be greatest among those under the age of 60, with a college degree or more, and men.

Conclusions. While the question concerning the amount that respondents traveled 100 miles or more outside the local area showed no significant results for the sample as a whole or by any demographic, the question concerning the amount that individuals used the Internet or email produced significant results for the sample as a whole and among each demographic group. What appears to have happened was a clear case of satisficing, where respondents found the first answer they could reasonably justify and selected it. Thus, when the response option "nearly everyday" appeared before "everyday," respondents chose it and moved on to the next question. What is interesting is that we find this effect across demographic groups; therefore, it does not appear that cognitive sophistication, as some have suggested, caused these types of mistakes (Krosnick 1991). Instead, it appears that the effects can be traced back to question design.

These scalar question findings contribute to research on satisficing and primacy. Due to the inconsistency in past findings on primacy, some have suggested that other aspects

Table 6: Response Distributions for Manipulations of Response Options: Percentages of Respondents Choosing Each Response When Given Categories in Order from High to Low Levels of Activity and the Reverse Order

Travel Question			Internet Usage Question		
	High to Low	Reverse Order		High to Low	Reverse Order
(1) Once or more a week	4.1	5.8	(1) Everyday	56.3	37.5
(2) Once or more a month	38.1	37.9	(2) Nearly everyday	15.6	28.8
(3) Once or more a year	45.5	43.0	(3) A few times per week	15.4	18.5
(4) About once a year	6.6	6.3	(4) Once a week or less	12.8	15.3
(5) Less than once a year	5.7	7.0			
Mean	2.71	2.72	Mean	1.85	2.12
n	654	653	n	638	631
Overall Chi-Square	$\chi^2 = 3.33, p = .504$		Overall Chi-Square	$\chi^2 = 41.57, p =< .001$	
t-test	$t = -0.14, p = .885$		t-test	$t = -3.93, p = .000$	

of questions, whether topic, structure or some other feature, might contribute to the occurrence of such effects (Dillman, Sangster, Tarnai and Rockwood 1996). One of the two questions here may provide examples of such effects. In the Internet usage question, either of the two adjacent categories (using the Internet everyday or using it nearly everyday) is probably satisfactory for many respondents to check. Therefore, the one that appears first in the sequence is more likely to get chosen, while the remaining categories are unaffected.

Experiment 6: Response Category Effects in the Presence of a Don't Know Response Option

The next experiment examines response order effects in an opinion-based scalar question that includes a "don't know" response option (Figure 5). Though the response options were reversed in version 2, the "don't know" appeared at the bottom of both lists; that is, only the substantive responses were reversed.

Research has shown that respondents tend to look for more positive answers in scales (Dillman 2007). As such, we expect the "don't know" response to be used less often when the options are ordered from most positive to most negative because in this ordering respondents can quickly find an option that fits them. Moreover, we expect those without an opinion to choose the "neutral" category in this version. In contrast, we expect the "don't know" option to be used more often when the negative options are presented first because respondents have to read further through the list to find the positive answers, making them more likely to see, and then use, the "don't know" option. Furthermore, based on the work of Rapoport (1982) and Knäuper (1999), we expect women to be more likely to provide the "don't know" or "neutral" response than men, and individuals over the age of 60 and with less than a college degree to be most affected by the reversal of response options, leading to higher percentages of non-opinions than among younger individuals.

Findings. Table 8 reports the results for response distributions both with the "don't know" category and with it removed from the response distribution. It is clear that when the response options start with the negative categories, respondents are more likely to choose the "don't know" category, compared to when the response options begin with positive categories (23.4% and 14.5%, respectively [$\chi^2 = 16.75$; $p < .001$]). Additionally, as hypothesized, when the response options appear in the expected order (e.g. positive first), respondents are more likely to choose the "neutral" response than in the alternate version (23.5% and 16.5%, respectively [$\chi^2 = 9.86$; $p < .01$]).

Table 9 reports the percentages for the use of the "don't know" and "neutral" response options by respondents' characteristics. With respect to the "don't know" category, every demographic group except respondents with less than a college degree exhibited the same behavior as that seen in the overall response distributions. While respondents over the age of 60 reported higher levels of non-opinionation than younger respondents, each group appears to be similarly affected by the reversal of response options. This is also true for the difference between women and men. Women reported higher levels of non-opinionation overall, but both sexes were more likely to choose the "don't know" response when given the version that began with "very bad." Turning to the "neutral" category, respondents over 60, those with a college degree or more, and women showed the greatest propensity to choose this option more in the expected order than in the reverse order.

Conclusions. When the response options appeared in an expected order with "very beneficial" listed first, respondents were much less likely to choose "don't know" than when the options began in the reverse order. This effect supports previous research on visual layout and design (Tourangeau, Couper and Conrad 2004) that predicts negative effects of breaking respondents' a priori expectations. When examining whether certain demographic groups were more likely to be affected by the design of this question, it was found that the likelihood of choosing the "don't know" category in the reverse order cut across all but one demographic group (respondents with less than a college degree), meaning that the visual layout seems

Table 7: Response Distributions for Manipulation of Question for Each Demographic Group

	60 years +		< 60 years		College +		< College		Women		Men	
	High to Low	Low to High	High to Low	Low to High	High to Low	Low to High	High to Low	Low to High	High to Low	Low to High	High to Low	Low to High
	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
(1) Everyday	54.9	38.0	59.6	39.3	62.0	42.7	46.6	29.5	57.4	38.9	59.6	37.2
(2) Nearly everyday	14.7	32.0	14.1	28.4	14.9	26.4	17.0	33.2	14.8	29.7	14.0	29.6
(3) A few times per week	12.7	18.0	15.0	17.2	14.6	18.9	15.9	17.9	13.9	17.2	15.5	17.9
(4) Once a week or less	17.6	12.0	11.4	15.1	8.5	12.1	20.5	19.5	13.9	14.2	10.9	15.3
Mean	1.93	2.04	1.78	2.08	1.94	2.22	1.64	1.87	1.84	2.07	1.78	2.11
n	102	100	334	331	316	307	176	190	244	239	193	196
Overall Chi-Square	$\chi^2 = 11.58$, $p = .009$		$\chi^2 = 32.22$, $p = .000$		$\chi^2 = 24.77$, $p = .000$		$\chi^2 = 16.69$, $p = .001$		$\chi^2 = 21.53$, $p = .000$		$\chi^2 = 22.64$, $p = .000$	
t-test	$t = -0.69$, $p = .486$		$t = -3.52$, $p = .000$		$t = -3.91$, $p = .001$		$t = -2.14$, $p = .033$		$t = -2.24$, $p = .025$		$t = -3.08$ $p = .002$	

Percentages of Respondents Choosing Each Response When Given Categories in Order from High to Low Levels and the Reverse Order

23. One of the recent changes that appears to be affecting some people in Lewiston/Clarkston is the Internet. What kind of an effect do you think the Internet is having on most people who live in the Lewiston/Clarkston area?

- ❑ Very beneficial
- ❑ Mostly beneficial
- ❑ Neutral
- ❑ Mostly bad
- ❑ Very bad
- ❑ Don't know

23. One of the recent changes that appears to be affecting some people in Lewiston/Clarkston is the Internet. What kind of an effect do you think the Internet is having on most people who live in the Lewiston/Clarkston area?

- ❑ Very bad
- ❑ Mostly bad
- ❑ Neutral
- ❑ Mostly beneficial
- ❑ Very beneficial
- ❑ Don't know

Figure 5. Category Effects in the Presence of a Don't Know Response Option

9. Which of these do you believe are the largest and smallest problems facing residents of the Lewiston & Clarkston area? Use "1" for the largest problem, "2" for second largest problem and so forth until you have completed all eight.

Lack of community involvement ……..	
Taxes are too high……..	
Lack of affordable health care…….	
Lack of money for local schools……	
Lack of affordable housing……	
Lack of good jobs……..	
Too much crime overall……..	
Too much drug use……	

9. Which of these do you believe are the largest and smallest problems facing residents of the Lewiston & Clarkston area? Use "1" for the largest problem, "2" for second largest problem and so forth until you have completed all eight.

Too much drug use……	
Too much crime overall……..	
Lack of good jobs……..	
Lack of affordable housing……	
Lack of money for local schools……	
Lack of affordable health care…….	
Taxes are too high……..	
Lack of community involvement ……..	

Figure 6. Manipulations of Response Options in a Ranking Question

Table 8: Response Distributions for Scalar Question When the Response Options Appear in Expected and Reverse Order and With the Don't Know Option Included and Removed from the Distribution

	With the Don't Know Option		Without the Don't Know Option	
	Expected[a] Order (%)	Reverse Order (%)	Expected[a] Order (%)	Reverse Order (%)
Very beneficial	19.1	15.4	22.4	20.1
Mostly beneficial	40.0	39.4	46.8	51.5
Neutral	23.5	16.5	27.4	21.5
Mostly bad	2.3	4.5	2.7	5.8
Very bad	0.6	0.8	0.7	1.0
Don't know[b]	14.5	23.4	–	–
n	638	631	554	497
Overall Chi-Square	$\chi^2 = 28.65, p = .000$		$\chi^2 = 11.91, p = .018$	

[a] Expected order refers to response options that start with the most positive option (e.g., very beneficial) and end with the first option's opposite (e.g, very bad).

[b] The "don't know" response option appeared as the last option in both versions.

Table 9: Percentage of Respondents that Used the Don't Know and Neutral Categories for Each Demographic Group by Whether the Response Options Appeared in Expected or Reverse Order[a]

	Use of Don't Know Category[b]						Use of Neutral Category					
	Response Option Order	n	%	Difference	χ^2	*p*	Response Option Order	n	%	Difference	χ^2	*p*
Overall	Expected	648	14.5	-8.9	16.75	≤ .001	Expected	649	23.5	+7.0	9.86	≤ .01
	Reverse	649	23.4				Reverse	648	16.5			
60+ yrs.	Expected	160	21.5	-12.9	6.16	≤ .01	Expected	158	23.4	+10.3	5.65	≤ .05
	Reverse	158	34.4				Reverse	160	13.1			
< 60*yrs.*	Expected	357	7.3	-7.3	9.95	≤ .01	Expected	357	24.1	+5.6	3.34	ns
	Reverse	362	14.6				Reverse	362	18.5			
< *College*	Expected	178	11.8	-3.8	1.08	ns	Expected	266	22.6	+3.4	0.97	ns
	Reverse	173	15.6				Reverse	287	19.2			
College +	Expected	442	14.7	-11.6	18.56	≤ .001	Expected	354	24.0	+10.3	12.80	≤ .001
	Reverse	464	26.3				Reverse	350	13.7			
Male	Expected	227	9.3	-9.0	7.92	≤ .01	Expected	227	22.9	+6.3	2.91	ns
	Reverse	235	18.3				Reverse	265	16.6			
Female	Expected	283	13.8	-8.3	6.75	≤ .01	Expected	289	24.6	+7.6	5.07	≤ .05
	Reverse	294	22.1				Reverse	294	17.0			

[a] Expected order refers to response options that start with the most positive option (e.g., very beneficial) and end with the first option's opposite (e.g., very bad).

[b] The don't know response option appeared as the last option in both versions.

to have affected most respondents in the same way, regardless of demographic characteristics.

With the "don't know" category, we see that this option is much more likely to get checked when the most used categories (very beneficial and mostly beneficial) appear at the end of the response options. As suggested in our hypotheses, in this situation the "don't know" response option is visually more accessible to respondents. While in the other version, the respondents may check the beneficial categories at the top of the list and never see the "don't know" category, thus rendering the "don't know" category visually inaccessible. Previous research has also shown that when an undecided category is placed in the middle of scalar responses, respondents that are neutral as well as undecided use it (Willits and Janota 1996).

Experiment 7: Response Order Manipulations in a Ranking Question

Design. Although response order effects are widely documented in scalar questions (Dillman 2007; Krosnick 1999; Schuman and Presser 1981) as well as in mark-all-that-apply questions (Smyth et al. 2006), very little research has addressed whether they are present in ranking questions. However, survey researchers often ask respondents to rank a series of items. In these ranking questions, respondents are generally given a list of response options and asked to rank them based on a given criterion that is generally specified in the query (e.g. What is the biggest problem?). These questions may be very difficult for respondents to answer for two reasons. First, respondents must carry the information from the query to the list of response options. For example, the query might ask the respondent to rank the options from best to worst or vice versa. However, the respondent must recall when answering the question what the highest and lowest numbers are meant to represent. Second, respondents must evaluate each response option relative to the others in order to provide a ranking; thus, the more options there are, the more difficult the task.

In the only test of reversing response options in a ranking question that we could locate, Ali (2003) found that the reversal affected how respondents ranked the options on a scale from 1 to 10 where 1 represented the biggest problem in the community and 10 represented the smallest problem. Specifically, the data showed that options at the top and bottom of the list showed significant differences, with options listed first receiving a higher ranking (i.e. were listed as "bigger problems"). Nonetheless, some researchers have suggested that ranking questions actually produce better data than the traditionally used rating questions because respondents tend to answer rating questions very quickly (see Krosnick 1999 for review). However, Krosnick's (1999) work on satisficing would suggest that one should expect response order effects in ranking questions due to the cognitive difficulty of answering them. As a result of this work, we expect to see an indication of respondents satisficing, as evidenced through primacy by ranking options as larger problems (i.e., giving them lower numbers where 1 means the largest problem) when they appear early in the list than when they appear later in the list, because of the cognitive difficulty of ranking questions. Thus, the major difference is expected to appear in the first and last few response options; the middle categories are not expected to exhibit the same magnitude of differences. Based on previous research, we also expect older respondents and those with less than a college degree to be most affected by the reversal of the response options. Furthermore, one might expect the need to carry information from the query and evaluate each option in relation to others to be more cognitively taxing for older individuals as well as those with lower levels of education (Knäuper 1999), resulting in more response order effects among these groups.

To examine these expectations the final experiment examines two versions of a question that asked respondents to rank eight problems facing the area from the largest to the smallest. Exact question wording as well as the original and reverse order presentations of the response options can be seen in Figure 6. In the analyses below, 303 cases were removed (reducing the sample size from 1,315 to 1,012) because they either provided the same ranking to different items or did not provide a ranking for all eight items.

Findings. Table 10 reports the results for each of the response options. As hypothesized, the response options that appeared in the first and last two positions, depending on questionnaire version, ("lack of community involvement," "taxes too high," "too much crime," and "too much drug use"), showed the largest effects (t-scores are negative in the first columns and positive in the last columns). The middle categories seemed virtually unaffected by the reversal as evidenced by the t-scores. For example, when "too much drug use" appeared first in the list of options, 22.8% of respondents labeled it as the biggest problem, whereas when it appeared last, only 16% of respondents reported that it was the largest community problem, thus increasing the mean for the version where the option was listed eighth. These findings support those of Ali (2003) and show that primacy occurs in ranking questions in the same way as rating questions.

Having seen the effect of reversing the response options among the first and last two categories in the overall distributions, we will now examine whether there are differences by demographic group in the mean values provided for these four response options. The first row in Table 11 shows the overall means and difference of mean t-tests for the four response options. The first option, "lack of community involvement," approaches significance (p=.065), while the other three options reach statistical significance. Again, regardless of significance level, it is important to notice that the t-scores are negative in the first two columns and positive in the second two columns, indicating the mean scores in each case were lower (i.e., respondents ranked the option as a larger problem) when the response option appeared earlier in the list. Respondents over the age of 60 do not appear more likely than those under the age of 60 to be affected by the reversal, as evidenced by the fact that both groups showed mean differences in three out of the four response options, albeit not the same options. Respondents with a college degree or more were the least affected by the reversal of the response options. Of the four

Table 10: Percentage of Respondents Ranking Problems in the Community by Position the Option Was Listed in the Response Categories.

		Lack of community involvement		Taxes too high		Lack of affordable health care		Lack of money for local schools		Lack of affordable housing		Lack of good jobs		Too much crime		Too much drug use	
Position		1^{st} (%)	8^{th} (%)	2^{nd} (%)	7^{th} (%)	3^{rd} (%)	6^{th} (%)	4^{th} (%)	5^{th} (%)	5^{th} (%)	4^{th} (%)	6^{th} (%)	3^{rd} (%)	7^{th} (%)	2^{nd} (%)	8^{th} (%)	1^{st} (%)
Largest	1	3.0	3.5	15.0	11.5	5.8	6.6	4.8	6.6	3.4	1.2	49.9	46.0	2.2	4.1	16.0	22.8
Problems	2	5.6	5.5	17.4	13.8	14.0	14.6	12.8	13.1	9.6	12.6	13.8	16.6	8.6	11.7	22.4	17.5
	3	8.4	7.4	10.4	11.9	16.0	16.4	15.6	12.9	8.2	12.5	16.6	13.8	12.8	10.1	13.6	15.2
	4	13.0	9.4	12.2	10.5	16.8	15.0	15.0	14.0	14.6	12.3	7.0	10.9	13.4	15.6	8.8	11.7
	5	11.4	9.2	10.0	11.3	17.2	13.6	14.4	14.8	18.2	15.0	5.6	6.4	11.8	16.4	10.4	12.5
	6	12.2	11.9	10.2	9.9	11.8	12.3	16.2	13.6	17.8	18.7	2.8	3.5	17.4	18.3	8.4	9.6
Smallest	7	13.4	13.3	13.6	15.8	11.2	14.0	9.4	14.4	14.4	17.0	2.4	1.8	19.6	13.6	13.2	8.4
Problems	8	32.9	40.0	11.0	15.2	7.0	7.4	11.6	10.5	13.6	11.3	0.9	0.5	14.0	10.1	7.0	2.3
Mean		5.79	6.04	4.25	4.65	4.41	4.44	4.27	4.34	5.14	5.08	2.31	2.38	5.25	4.89	3.88	3.47
n		499	513	499	513	499	513	499	513	499	513	499	513	499	513	499	513
Overall Chi- Square		$\chi^2 = 8.36$, $p = .302$		$\chi^2 = 10.19$, $p = .178$		$\chi^2 = 4.60$, $p = .700$		$\chi^2 = 9.34$, $p = .204$		$\chi^2 = 16.27$, $p = .023$		$\chi^2 = 10.08$, $p = .184$		$\chi^2 = 20.96$, $p = .005$		$\chi^2 = 30.19$, $p = .000$	
t-test		$t = -1.85$, $p = .065$		$t = -2.69$, $p = .007$		$t = -.266$, $p = .790$		$t = -.205$, $p = .837$		$t = .460$, $p = .646$		$t = -.553$, $p = .581$		$t = 2.92$, $p = .004$		$t = 2.99$, $p = .003$	

response options, individuals with a college degree or more only showed a primacy effect for "lack of community involvement" (listed first in one version and last in the other). For respondents with less than a college degree, a primacy pattern was present in three of the four options. In terms of sex, both men and women were affected by the reversal in similar ways.

Conclusions. Overall, the position of the response option clearly affects how it was ranked. However, as seen above, there are some mixed results concerning how powerfully the response option's position affected respondents across the demographic groups. For example, there is little support for the hypothesis that age would affect the propensity for satisficing in this experiment. However, there is evidence that education affected the likelihood of satisficing. As a result, in keeping with previous research (Krosnick, Narayan and Smith 1996), it appears that respondents with less than a college degree were more affected by the reversal than those with a college degree or more. Nonetheless, most of the demographic groups showed a propensity to provide low rankings when the item appeared high in the list, suggesting that the question design outweighed individual demographic differences.

Discussion & Overall Conclusions

This paper has attempted to contribute to our understanding of how to reduce measurement error in self-administered surveys by unifying two literatures. On the one hand, the emerging literature concerning visual layout and design theory has not addressed how respondents with varied demographic characteristics may be differentially affected by the questionnaire design. Indeed, a major shortcoming of this research has been the homogeneity of the samples used to test these theories, in particular, the use of college student samples. On the other hand, the research concerning demographic differences in respondent behavior has not addressed the recent contributions of visual design theory.

The results from our seven experiments embedded in a general population survey provide substantial evidence that the visual design of questions (graphical and verbal manipulations) in self-administered surveys affects respondents' behavior regardless of age, educational attainment, and sex. Thus, the results of this paper lend considerable support to previous work that has served to explicate and/or test visual design theory. For example, the replications of Christian and Dillman's (2004) number box versus polar point scalar questions and use of different sized open-ended answer spaces resulted in findings that suggest that it is the visual design that was influential across demographic sub-groups. In addition, we were able to replicate the findings from Smyth et al.'s (2006) check-all-that-apply versus forced choice formats for all but two of demographic groups (men and respondents over 60).

A very interesting result is that the use of the don't know category was so dramatically affected by the reversal of response options. These findings seem to support the suggestion of Tourangeau, Couper and Conrad (2004) that when the response options do not meet the respondent's a priori expectations, the way they answer the question is affected. In our test, the neutral category was used more often when the beneficial categories were placed first. This use seems consistent with respondents reading down a list that starts with positive responses, but when noting that they were moving into negative responses (which few wanted to pick) they opted for what seemed a reasonable category (i.e. neutral). It is impossible to know whether this finding is the product of higher levels of non-opinionation, as suggested by Rapoport (1982), or the effects of the "up means good" and "middle means typical" heuristics. In all likelihood, it is a combination. However, it is clear that the reversal of response options did not equally affect all demographic groups in terms of their propensity to choose the conceptual middle category ("neutral"). These findings suggest that it is not only category order that may influence people's answers, but characteristics of those categories and their layout, as well as the visibility of response options. This explanation may also address why there were not consistent results across demographic groups in the ranking question. This is an issue that clearly needs further research, which may help explain the highly inconsistent results observed in previous primacy experiments.

In short, while the majority of these seven experiments have shown that the effects of visual design affect people of different ages, educational attainment, and sex in similar ways, what remains unresolved is why different demographic groups show effects to be of the same kind and yet different in degree and/or magnitude. For example, while all demographic groups provided more answers in the larger space in the open-ended experiment, there was quite a bit of variation between the groups, particularly, in terms of the number of words used regardless of box size. Several of these issues may benefit from the use of cognitive interviews where more in-depth qualitative data can be collected and analyzed. Thus, while this research bolsters previous studies on the importance of visual design theory, more studies are needed to test the varying effects on demographic groups.

Acknowledgements

Support for conducting this research was provided by the Department of Community and Rural Sociology and the Social and Economic Sciences Research Center at Washington State University, Pullman, WA. Additional support was provided by the USDA-National Agricultural Statistics Service and the National Science Foundation Division of Science Resource Statistics under Cooperative Agreement #43-3AEU-1-80055 with the Washington State University Social and Economic Sciences Research Center, Don A. Dillman, Principal Investigator.

We wish to acknowledge with thanks thoughtful comments on earlier versions of this paper provided by Kendra W. Stern, Thomas Rotolo, and Scott Myers.

References

Ali, Y. (2003). *Local and global environmental issues and attitudes: The case of the snake and clearwater rivers.* (Unpublished Doctoral Dissertation. Washington State University)

Table 11: Percentage for Each Demographic Group Ranking Problems by When the Option Was Listed in the First Two or Last Positions in the List

	Lack of Community Involvement						Lack of Affordable Housing					
	Position	n	Mean	Difference	t-test	*p*	Position	n	Mean	Difference	t-test	*p*
Overall	1^{st}	499	5.78	-0.25	-1.84	.065	2^{nd}	499	4.25	-0.40	-2.69	.007
	8^{th}	513	6.03				7^{th}	513	4.65			
60+ yrs.	1^{st}	108	5.48	-0.52	-1.82	.069	2^{nd}	108	3.90	-0.62	-1.93	.054
	8^{th}	119	6.00				7^{th}	119	4.52			
< 60 yrs.	1^{st}	305	5.91	-0.09	-.516	.606	2^{nd}	305	4.42	-0.38	-2.01	.044
	8^{th}	311	6.00				7^{th}	311	4.80			
< College	1^{st}	335	6.04	-0.07	-.477	.633	2^{nd}	335	4.07	-0.50	-2.72	.007
	8^{th}	338	6.11				7^{th}	338	4.57			
College +	1^{st}	156	5.21	-0.69	-2.84	.005	2^{nd}	156	4.62	-.020	-.768	.443
	8^{th}	161	5.90				7^{th}	161	4.82			
Male	1^{st}	185	5.52	-0.31	-1.35	.178	2^{nd}	185	4.11	-0.42	-1.68	.094
	8^{th}	196	5.83				7^{th}	196	4.53			
Female	1^{st}	232	5.97	-0.17	-.893	.372	2^{nd}	232	4.38	-0.48	-2.23	.026
	8^{th}	236	6.14				7^{th}	236	4.86			
	Too Much Crime						Too Much Drug Use					
	Position	n	Mean	Difference	t-test	*p*	Position	n	Mean	Difference	t-test	*p*
Overall	8^{th}	499	5.25	0.37	2.91	.004	7^{th}	499	3.88	0.41	2.99	.003
	1^{st}	513	4.88				2^{nd}	513	3.47			
60+ yrs.	8^{th}	108	4.96	0.60	2.20	.029	7^{th}	108	3.70	0.39	1.34	.180
	1^{st}	119	4.36				2^{nd}	119	3.31			
< 60 yrs.	8^{th}	305	5.46	0.33	2.12	.034	7^{th}	305	4.03	0.52	3.02	.003
	1^{st}	311	5.13				2^{nd}	311	3.51			
< College	8^{th}	335	5.14	0.39	2.51	.012	7^{th}	335	3.83	0.44	2.60	.010
	1^{st}	338	4.75				2^{nd}	338	3.39			
College +	8^{th}	156	5.49	0.25	1.20	.231	7^{th}	156	3.95	0.24	.999	.318
	1^{st}	161	5.24				2^{nd}	161	3.71			
Male	8^{th}	185	5.24	0.44	2.20	.028	7^{th}	185	4.04	0.62	2.82	.005
	1^{st}	196	4.80				2^{nd}	196	3.42			
Female	8^{th}	232	5.39	0.38	2.05	.041	7^{th}	232	3.89	0.42	2.05	.040
	1^{st}	236	5.01				2^{nd}	236	3.47			

Ayidiya, S., & McClendon, M. (1990). Response effects in mail surveys. *Public Opinion Quarterly*, *54*(2), 229-247.

Christian, L., & Dillman, D. (2004). The influence of graphical and symbolic language manipulations on responses to self-administered questions. *Public Opinion Quarterly*, *68*, 58-81.

Couper, M., Traugott, M., & Lamias, M. (2001). Web survey design and administration. *Public Opinion Quarterly*, *65*, 230-253.

Dillman, D. (2007). *Mail and internet surveys: The tailored design method*. New York: Wiley.

Dillman, D., Sangster, R., Tarnai, J., & Rockwood, T. (1996). Understanding differences in people's answers to telephone and mail surveys. *New Directions for Evaluation*, *70*, 45-61.

Glenn, N. (1969). Aging, disengagement and opinionation. *Public Opinion Quarterly*, *33*(1), 17-33.

Israel, G. (2006, February). *Visual cues and response format effects in mail surveys.* (Paper presented at the annual meeting of the Southern Rural Sociological Association. Orlando Florida)

Jenkins, C., & Dillman, D. (1997). Towards a theory of self-administered questionnaire design. In L. E. Lyberg et al. (Eds.), *Survey measurement and process quality.* New York: John Wiley & Sons.

Knäuper, B. (1999). The impact of age and education on response order effects in attitude measurement. *Public Opinion Quarterly*, *63*(3), 347-370.

Knäuper, B., Belli, R., Hill, D., & Herzog, A. (1998). Question difficulty and respondents' characteristics: Effects on data quality. *Journal of Official Statistics*, *13*, 181-199.

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

Krosnick, J. (1992). The impact of cognitive sophistication and attitude importance on response-order and question-order effects. In N. Schwarz & S. Sudman (Eds.), *Context effects in social and psychological research.* New York: Springer.

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

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

Krosnick, J., Narayan, J., & Smith, W. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, *70*, 29-44.

Rapoport, R. (1982). Sex differences in attitude expression: A generational explanation. *Public Opinion Quarterly*, *46*(1), 86-96.

Redline, C., & Dillman, D. (2002). The influence of alternative visual designs on respondents' performance with branching instructions in self-administered questionnaires. In R. M. Groves, D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse.* New York: Wiley.

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, New Jersey: Lawrence Erlbaum.

Smith, T. (1993). *Little things matter: A sampler of how differences in questionnaire format can affect survey response.* National Opinion Research Center, Methodological Report #78. (University of Chicago General Society Survey)

Smyth, J., Dillman, D., Christian, L., & Stern, M. (2006). Comparing check-all and forced-choice question formats in web surveys. *Public Opinion Quarterly*, *70*(1), 66-77.

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

Tourangeau, R., Couper, M., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly*, *71*, 91-112.