IS SATISFICING A SLIPPERY SLOPE? AN INVESTIGATION INTO THE EFFECTS OF SATISFICING WHEN ANSWERING QUESTIONS IN A MAIL SURVEY

Kumar Rao [*,1] **Olena Kaminska** [**]

[*] *The Gallup Organization*
[**] *Program in Survey Research and Methodology, University of Nebraska-Lincoln*

Abstract: In this article, we investigate response-order effects across various candidate questions (i.e., questions potentially prone to satisficing) from a 2007 Social Issues mail survey conducted by The Gallup Panel. We also investigate extreme responding and analyze patterns of extreme responses. Then, we investigate if the odds of providing a extreme straight non-differentiated response to a target question is affected by providing response in the same way to questions prior to the target question. The results suggest the possibility that response-order effects are especially strong for response choices placed in the lower half of the response list. We discuss the findings from the study and conclude with recommendations for future research.

Keywords: Satisficing, Response Behavior

1. INTRODUCTION

The order in which questions, tasks, sections or choices appear in a questionnaire have been shown to effect the responses elicited (Schuman and Presser, 1996). Most commonly, ordering effects refer to the effect that the content of early questions (tasks or sections) may have on the answers given in later questions (tasks or sections). Along the same lines, response-order effects refer to the order in which choices or lists are presented within a question. This study reports on a direct testing of this effect across various candidate questions (i.e., questions potentially prone to satisficing) using data gathered from two questionnaire versions of a mail survey. The two versions were the same in every aspect other than the ordering of the response lists, which were reversed for the candidate questions in one of the versions. An important and unique distinction of this study pertains to the type of questions analyzed. In this study, all the candidate questions are of likert-type, in contrast to most studies that have examined response-order effects in mail surveys have asked respondents to pick the top few items in the response list (see, for instance, Krosnick and Alwin (1987)). Lastly, this study provides preliminary evidence about whether or not respondents who satisfice the first set of candidate questions are likely to satisfice all of the remaining candidate questions.

There have been a number of studies that have examined response-order effects (see, e.g., Krosnick and Alwin (1987); Krosnick *et al.* (1996); McClendon (1986); Mingay and Greenwell (1989); Bishop and Smith (2001); Sudman *et al.* (1996)), and a number of theories put forward as to why

[1] Kumar Rao is a graduate of Survey Research and Methodology program in University of Nebraska-Lincoln and currently works for the Gallup Panel, Gallup Organization. Email: kumar_rao@gallup.com

they might occur. Early investigators have theorized response-order effects as a consequence of asking long, difficult, or complex questions (Payne (1951); Rugg and Cantril (1944)). But as Schuman and Presser (1981) have demonstrated, significant and substantial response-order effects are evident even in some short and simple questions. Some theories focus on the impact of memory and the notion that response-order effects arise because of the inability of respondents to remember all of the response alternatives, particularly true, it is suggested, if questions are read alound without a visual aid (Blankenship (1943); Knauper (1995)). Few other theories have focussed on the cognitive processes that respondents go through when choosing items (Krosnick and Alwin, 1987), including the theory of Satisficing (Simon (1957); Krosnick (1991)), which argues that people will choose adequate answers rather than optimal. The tendency to satisfice in this manner, according to this theory, depends on three factors: 1) the difficulty of the question; 2) the respondent's ability to optimize (i.e. to retrieve, process, and integrate information from memory); and 3) the respondent's motivation to optimize. In the case of response-order, the theory predicts a form of "weak satisficing" that leads respondents to choose the first acceptable alternative that is presented to them in a closed-ended question, depending on the mode in which the data are collected. If the response alternatives are read aloud to the respondent without a visual aid, the satisficing theory predicts recency effects (i.e., respondents are more likely to choose the last items in the list). But if the response alternatives are presented in a visual format, satisficing theory predicts primacy effects (i.e., respondents are more likely to choose those at the top of the list). For a more detailed account of the satisficing model and its recent evolution, refer to Krosnick (1991) and Krosnick *et al.* (1996).

Even though the various theories about response-order effects have certainly enriched the science of survey research as they have evolved, none appear as yet to provide a plausible general explanation for response-order effects. In an important exception to this generalization, however, Krosnick and his associates (see, Krosnick and Alwin (1987); Krosnick *et al.* (1996)) conducted various experiments and meta-analyses and confirmed the satisficing theory's predictions on response-order effects. At this stage, it would clearly be useful to re-examine the cognitive basis of satisficing theory to test the validity of its statistical support. A caveat merits a brief mention at this juncture: This study is not intended to test the satisficing theory's explanation of the cognitive processes underlying the effects of response-order. Instead, we limit our research in this study to test the response-order effects, predicted by the theory as a form of "weak satisficing". According to Krosnick (1991), motivation to optimize is probably greatest at the beginning of a questionnaire and decreases as more and more questions are asked and answered. He further argues that the longer an interview has been under way, the lower is the respondent's motivation to optimize, and the more likely satisficing will flourish. Along the same lines as an interview, for an individual question with a long list of items/response choices, one can expect that the satisficing behavior will be more prominent for later items in that question, perhaps even confounding with the position of that question in the questionnaire. However, it is worthy to note response-order effects like ones investigated in this study can occur due to both satisficing and context effects. While context effects determine each single item, the overall expected result of satisficing is an increasing trend of satisficing responses for later items in a question, independent of the item's topic. Then, the question is, how would a trend like that look?

Let's consider, for example, one of the candidate questions used in this study. Figure 8 in the Appendix section (see page 8) shows the first candidate question analyzed in this study. When we observe the pattern of mean difference for each item in this question before and after the item's position in the list has been reversed, we can expect higher mean values when items are positioned later in the list, as opposed to when they are placed earlier in the list. The reason being that while answering a question with a long response list like the one mentioned above, respondents satisfice by exhibiting behavior that involves choosing response ratings that are closer to the other extreme end (i.e., right end) of the scale as they move down the long list. Though not optimal, this is certainly a reasonable strategy, because it provides them a "quick and easy" way to answer (for the question in the illustration) that they worry a great deal to all problems, thereby minimizing the psychological costs required to respond. According to Krosnick (1991), when answering many questions using the same rating scale, satisficing respondents could provide identical or nearly identical ratings across questions by simply selecting a point on the response scale that appears to be reasonable for the first object, and then rating all of the remaining objects at that point. The end result would then be higher mean values obtained for later items as opposed to earlier items in the list when respondents choose ratings close to the right end (i.e., "A fair amount" or "A great deal" in the illustration) of the scale as the reasonable point as they move down the list.

In addition to analyzing mean values, we also investigate extreme responding [2], which is the tendency to select either end of the scale for all response choices in the list. Following this, we analyze the patterns of extreme responses across both regular and reversed versions of the survey questionnaire. Then, we investigate if the odds of providing a extreme straight non-differentiated response to a target question is affected by providing response in the same way to questions prior to the target question. Finally, we discuss the findings from the study and conclude with recommendations for future research.

2. METHODS

2.1 Sample

The data for this study were collected in a mail survey conducted by The Gallup Panel [3] in the spring of 2007. The survey [4] was entitled "Social Issues Survey," and was sent to a random sample of 28,747 active adult (ages 18 and above) Gallup Panel members assigned to receive mail surveys. The purpose of the survey was to obtain panel members' attitudes, opinions, and beliefs about several pressing social issues. The survey received a total of 20,691 responses at a net response rate of 72%.

As previously noted, two versions of the questionnaire were used such that the ordering of the response lists was reversed for the candidate questions in one of the versions. For classification purposes we will call these two versions as Regular version (FORM1 survey) and Reversed version (FORM2 survey). Table 1 shows the distribution of the total number of responses received across both versions.

Table 1. Total number of responses received across mail survey versions

Mail Survey Versions	*n*	%
FORM1 survey (regular version)	10,356	50.1
FORM2 survey (reversed version)	10,335	49.9
Total	20,691	100

2.2 Measures

Out of twenty six questions that were part of the Social Issues Survey, only three questions [5] were selected [6] as the candidate questions and then subjected to analysis. The first candidate question asked respondents to indicate how much they personally worry about various social problems starting from hunger and homelessness to the cost of higher education. Respondents rated their extent of worry as "Not at all," "Only a little," "A fair amount," or "A great deal". The second candidate question asked respondents to rate how much of a priority various public health issues should be for the federal government, with response options ranging from AIDS to Poverty. Respondents rated their priority as "Low," "Medium," "High," or "Top". Lastly, the third candidate question asked respondents the effectiveness of various measures aimed at reducing obesity in the United States. Respondents rated a particular measure's effectiveness as "Not at all", "Not too", "Fairly," or "Extremely,". Values of one through four were assigned to all these descriptors in order to compute means.

A multivariate analysis of variance (MANOVA) test was conducted to determine if vectors of means for the two forms are significantly different when responses to all items are considered simultaneously. In addition to this multivariate test, independent t tests were conducted to compare mean values of items across both forms for all three candidate questions. Another reason for conducting these multiple t tests later on was to

[2] For the first candidate question which was rated as "Not at all", "Only a little", "A fair amount", or "A great deal", a extreme response would be one that is rated either "Not at all" or "A great deal" for all items in the response list. Refer to Figure 7 in page 8 for an example.

[3] The Gallup Panel is one of the nation's few research panels that are representative of the entire U.S. population. The Gallup Panel selects households using random digit dialing (RDD) methodology. Panel members are randomly recruited by telephone and, depending on their level of usage of the Internet, are assigned to receive surveys either through the mail or online (in addition to telephone surveys).

[4] The survey was also sent to 27,130 active Gallup Panel members assigned to receive Web surveys. The Web survey received a total of 17,488 responses at a net response rate of 64%. The data obtained from the Web survey is not part of this current investigation.

[5] Please refer to the Appendix section for illustrations of all candidate questions.

[6] While one could argue about the subjective nature of the decision to select a question as a potentially satisfiable question before-hand, the decision was largely guided by the insights derived from reviewing literature on satisficing theory. For instance, studies have found non-differentiation to be more common toward the end of a questionnaire than toward the beginning (Herzog and Bachman (1981); Kraut *et al.* (1975)), and that a longer response list might be more susceptible to primacy effects (Krosnick and Alwin, 1987). In light of such theoretical perspectives, three candidate questions with long response lists were selected. The first two candidate questions had a response list containing 14 and 13 response choices, and the third question had 19 response choices. The positions of candidate questions in the questionnaire ensured that there exists a nice spread.

observe the pattern of mean difference for each item in the candidate question before and after the item's position in the list was reversed. Then, patterns of extreme responses (i.e., responses to ratings placed either at the extreme left or the extreme right end of the scale) were analyzed to assess the effect of response-order on this style of reporting, starting from the top to the bottom of the response list.

3. RESULTS

3.1 Response-order effects

We first tested the hypothesis that responses to candidate questions vary depending on the order in which the response options are presented. Using one-way or single factor multivariate analysis of variance (MANOVA), we found clear evidence of response-order effects for all three candidate questions. For the first candidate question, a significant difference was found resulting in a Wilks' $f(\lambda) = 0.93$, F(14, 19333)=97.59, p=0.001, η^2=0.07. When the analysis when repeated for the second candidate question containing thirteen items, the following results were obtained: Wilks' $f(\lambda) = 0.91$, F(13, 19092)=141.77, p=0.001, and η^2=0.09. Lastly, for the third candidate question, the analysis resulted in a Wilks' $f(\lambda) = 0.89$, F(19, 18919)=121.70, p=0.001, and η^2=0.11. Overall, we can conclude that the analysis of responses to candidate questions from a multivariate approach (i.e., considering the responses to all items in a candidate question simultaneously) indicated that there is a significant difference in how subjects responded depending on the position of the item in the response list.

Subsequent univariate t tests indicated that, in most cases, items had a higher mean value when presented later in the list than when presented earlier in the list. The results are shown in Tables 2, 3, and 4. The tables present the means, standard deviations, and univariate t test values for all items in the response list across both forms. For the first candidate question, all univariate t values are significant, except for response option 4L. Interestingly, the effects of order appear particularly high for the items that are close to the top and bottom of the list, since these were the items that had widely different placement on the two forms. As expected, the items located in the middle of the list were far more stable [7] because their placement was not widely different between the regular and reversed ordered forms. For the second candidate question, univariate t values of items 14C and 14G were not significant, whereas the remaining were all significant. For the third candidate question, only response item 18I was found to be not significant, while all other items' univariate t values were significant.

We followed the above analysis by computing the mean difference for each item in the candidate question before and after the item's position in the list was reversed. Specifically, we calculated the difference in mean values as shown in Tables 2, 3, and 4 for each item before and after the item's position in the list has been reversed. The results are plotted in Figures 1, 2 and 3.

Fig. 1. Pattern of mean differences for items in candidate question #4

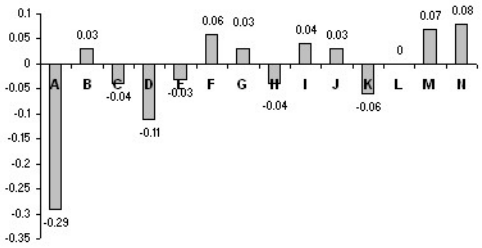


Fig. 2. Pattern of mean differences for items in candidate question #14

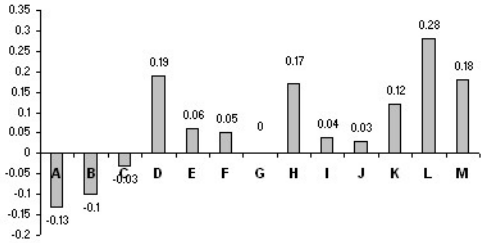


Fig. 3. Pattern of mean differences for items in candidate question #18

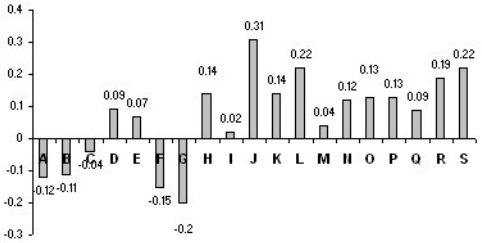


Figure 1 shows a mean difference of -0.29 for response choice "A," which means that this response choice received a higher rating when it was shown at the end (i.e. in FORM2) than when it was shown at the beginning of the list (i.e. in FORM1). The same is true at the other extreme end of the list i.e. response choice "N". For this item, the mean difference is 0.08, meaning this item

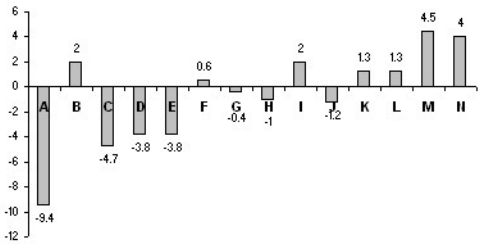
[7] This assertion is based on the finding for the first candidate question. Note that the absolute difference in mean values between the two forms (FORM1-FORM2) ranged from 0.03 to 0.04 for middle response items (i.e. 4G and 4H), whereas for the top and bottom items (i.e.4A and 4N), the absolute difference in mean values ranged from 0.08 to 0.29

was rated higher when it was shown at the end (i.e., in FORM1) than when it was shown at the beginning of the list (i.e., in FORM2). The figures magnify previously noted finding that the effects of order appear particularly high for the items that are closer to the top and bottom of the list, in contrast to those in the middle. However, there are some exceptions. For instance, response choice "J," which is shown in the exact middle of the list in both forms (see, Figure 3), has the highest mean difference (0.31). With response-order playing no role in this case, it simply means respondents who answered FORM1 rated this item higher than their FORM2 counterparts. All in all, the figures show a topsy-turvy pattern of difference in mean values, with negative values (i.e. FORM2 is higher) from the beginning to the middle of the list are followed by positive values (i.e. FORM1 is higher) from the middle to the end of the list. In simple terms, this analysis indicates items in most cases received a higher rating when they were placed close to the lower half of the response list.

3.2 Patterns of extreme responses

The positive and negative swiveling of mean values because of response-order variation led us to examine the possibility of extreme responding. Specifically, we computed the difference in proportion of those responding either at the extreme left or right end of the scale for each item before and after the item's position in the list has been reversed. The idea was to determine if the increase or decrease in mean values because of response-order variation for each item in the question is a function of reporting at the extreme ends of the scale. If that is the case, then patterns of mean differences and extreme responding should look similar. The results of the analysis are shown in Figures 4, 5, and 6.

Fig. 4. Pattern of differences in extreme responding for candidate question #4



As we can see from Figure 4, the difference between FORM1 and FORM2 for an extreme response is -9.4% for the first response choice (i.e., choice "A"). In other words, this item received 9.4% more extreme responses when it was placed at the end of the list (i.e., in FORM2) than when it

Fig. 5. Pattern of differences in extreme responding for candidate question #14

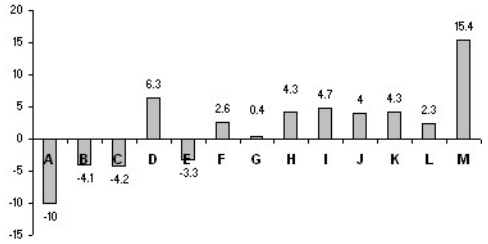
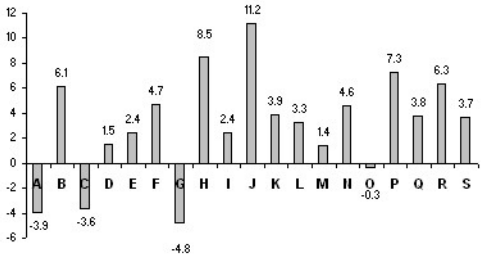


Fig. 6. Pattern of differences in extreme responding for candidate question #18



was placed at the top of the list (i.e., in FORM1). In a similar fashion, response choice "N" received 4.0% more extreme responses when it was placed at the end of the list than when it was placed at the top of the list. Notice how the difference tapers down at the middle section, becoming as low as -0.4% for response choice "G". A similar pattern can be seen for candidate question #14 (see, Figure 5), but with much higher values at the extremes (-10% for choice "A" and 15.4% for choice "M").

The outcome of the analysis for the third candidate question, shown in Figure 6, provides a somewhat different perspective about extreme responding and response-order effect from the one obtained by first two questions. While the pattern for the top and bottom items for this question is topsy-turvy (-3.9% for choice "A" and 3.7% for choice "S"), much like in two other questions, the effect of extreme responding is very different in the middle section. For instance, response choice "J" received 11.20% higher extreme responses in FORM1 than FORM2, for a position in the response list that is same in both forms (i.e., choice "J" is the 10th item in FORM1 and FORM2). But the key question is: are patterns of mean differences and extreme straight responses similar? By looking at the positive and negative fluctuations in differences across both sets of graphs, we can say that, with few exceptions, both patterns do look distinctively similar. For instance, items "P" to "S" for question #18 are associated with positive difference in means and proportions of extreme responses (see Figures 3 and 6). This tells us that when these items were shown at the lower half of the list, respondents were more likely

Table 2. Mean values and unpaired t test across both forms for the first candidate question (Question #4)

Response Option	FORM1				FORM2				Analysis	
	Position	n	Mean	SD	Position	n	Mean	SD	t	p
4A	1	10,293	2.71	0.91	14	10,291	3.00	0.90	-23.69	***
4B	2	10,303	3.26	0.79	13	10,295	3.23	0.80	2.73	**
4C	3	10,265	2.98	0.85	12	10,268	3.02	0.88	-3.70	***
4D	4	10,312	3.32	0.85	11	10,300	3.43	0.78	-9.93	***
4E	5	10,293	2.92	0.99	10	10,275	2.95	1.01	-2.02	*
4F	6	10,277	2.53	0.96	9	10,266	2.47	0.96	5.03	***
4G	7	10,297	3.06	0.98	8	10,285	3.03	1.00	2.51	*
4H	8	10,310	3.14	0.85	7	10,282	3.18	0.82	-4.11	***
4I	9	10,310	3.12	0.85	6	10,274	3.08	0.85	3.82	***
4J	10	10,291	2.68	0.98	5	10,267	2.65	1.00	2.55	**
4K	11	10,297	3.16	0.91	4	10,291	3.22	0.85	-4.89	***
4L	12	10,292	2.64	1.07	3	10,268	2.64	1.06	0.21	
4M	13	10,322	2.93	0.92	2	10,266	2.86	0.90	5.56	***
4N	14	10,313	2.83	1.01	1	10,260	2.75	0.99	6.29	***

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; (two-tailed)

Table 3. Mean values and unpaired t test across both forms for the second candidate question (Question #14)

Response Option	FORM1				FORM2				Analysis	
	Position	n	Mean	SD	Position	n	Mean	SD	t	p
14A	1	10,234	2.74	0.91	13	10,247	2.87	0.98	-9.94	***
14B	2	10,219	2.36	0.90	12	10,236	2.46	0.95	-8.47	***
14C	3	10,209	2.81	0.88	1	10,239	2.84	0.92	-1.91	
14D	4	10,212	3.27	0.80	10	10,242	3.08	0.88	16.62	***
14E	5	10,239	2.41	1.02	9	10,237	2.35	1.04	5.01	***
14F	6	10,233	3.38	0.74	8	10,234	3.33	0.76	4.95	***
14G	7	10,225	2.29	0.91	7	10,236	2.29	0.90	0.40	
14H	8	10,233	2.90	0.86	6	10,218	2.73	0.87	14.45	***
14I	9	10,228	3.03	0.93	5	10,239	2.99	0.89	2.84	**
14J	10	10,238	2.56	0.95	4	10,221	2.53	0.91	1.99	*
14K	11	10,208	2.64	0.90	3	10,173	2.52	0.86	9.66	***
14L	12	10,247	2.57	0.98	2	10,238	2.29	0.93	21.02	***
14M	13	10,241	3.09	0.93	1	10,221	2.91	0.85	14.86	***

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; (two-tailed)

to rate "Extremely" than when they were shown at upper half of the list. In a nut-shell, all of these findings seem to suggest the possibility that respondents would rate items placed at the lower half of the list higher (i.e., at extreme end of the scale) than if they were to rate the same items placed at the upper half of the list.

3.3 Extreme straight non-differentiation

In this section of our analysis, we use binary logistic regression to test our hypothesis[8] that

[8] According to Schaeffer *et al.* (2005), the number of questions a person has already answered prior to a target question can affect response quality: respondent fatigue is

Table 4. Mean values and unpaired t test across both forms for the third candidate question (Question #18)

Response Option	FORM1				FORM2				Analysis	
	Position	n	Mean	SD	Position	n	Mean	SD	t	p
18A	1	10,150	2.28	0.95	19	10,229	2.40	1.00	-9.03	***
18B	2	10,156	1.51	0.75	18	10,200	1.62	0.80	-10.39	***
18C	3	10,167	2.77	1.04	17	10,191	2.81	1.06	-2.41	*
18D	4	10,258	3.22	0.94	16	10,259	3.13	0.98	6.71	***
18E	5	10,258	2.82	1.01	15	10,248	2.75	1.00	5.41	***
18F	6	10,222	1.93	1.02	14	10,231	2.08	1.05	-10.78	***
18G	7	10,245	2.78	1.05	13	10,247	2.98	1.02	-14.91	***
18H	8	10,273	3.57	0.70	12	10,256	3.43	0.79	14.77	***
18I	9	10,239	2.88	1.01	11	10,240	2.86	0.99	1.51	
18J	10	10,245	3.09	0.89	10	10,239	2.78	0.92	25.51	***
18K	11	10,183	2.36	1.09	9	10,205	2.22	1.02	9.75	***
18L	12	10,216	2.71	1.03	8	10,242	2.49	1.02	15.95	***
18M	13	10,250	2.67	1.02	7	10,254	2.63	1.02	3.28	**
18N	14	10,247	2.91	0.92	6	10,256	2.79	0.92	9.74	***
18O	15	10,209	2.46	1.04	5	10,234	2.33	1.03	9.15	***
18P	16	10,250	3.28	0.84	4	10,251	3.15	0.87	11.25	***
18Q	17	10,223	2.83	0.94	3	10,224	2.74	0.93	7.33	***
18R	18	10,223	2.60	1.02	2	10,215	2.41	0.96	14.01	***
18S	19	10,226	2.27	1.14	1	10,207	2.05	1.04	14.51	***

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; (two-tailed)

whether providing extreme straight nondifferentiated [9] response to the third candidate question is related to reporting in the same way to prior two questions (i.e., second and first candidate questions). In this process, we determine the odds of extreme straight non-differentiation in question #18 as a function of extreme straight non-differentiation in the first and second candidate questions.

In the regression analysis, the dependent variable is whether the respondent provided an extreme straight non-differentiated response to third candidate question (i.e., question #18). The independent variables are whether the respondent provided a extreme straight non-differentiated response to the first and second candidate questions (i.e., question #4 and question #14 respectively). Both the dependent and independent variables were coded to range from 0 to 1. The dependent variable was coded 1 for an extreme straight response and 0 for not an extreme straight response. A reversed coding scheme (i.e., 0 for an extreme straight response and 1 for not an extreme straight response) was followed for the two independent variables. Table 5 shows the results of the analysis.

Table 5. Logistic regression analysis of predictors of extreme straight non-differentiation in question #18

Independent Variable	β	*SE*	*Odds ratio*
Question #4	-1.59	0.26	0.20***
Question #14	-1.28	0.28	0.27***

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$; (two-tailed)

often greater after a person has answered a lot of prior questions, and this may reduce the quality of responses. On a somewhat related note, Krosnick and Alwin (1989) show that non-differentiation is more common toward the end of a questionnaire than towards the beginning. keeping this in mind, the reason we theorized in this way has to do with the fact that in addition to being asked in the later half of the survey, this question had the most number (19) of response items in the list, which raises the possibility of respondent fatigue (before and while answering this question) and the likelihood to satisfice.

[9] For example, for the first candidate question which was rated as "Not at all", "Only a little", "A fair amount", or "A great deal", a extreme straight non-differentiated response would be one that is rated "Not at all" for all items or "A great deal" for all items in the response list.

Both independent variables significantly predict extreme straight non-differentiation in question #18 ($\chi^2(2) = 65.89$, N=20,691, $p < .001$). As can be seen from Table 5, the odds of providing a extreme straight non-differentiated response in question #18 is 0.20 if such a response was not provided in question #4 and 0.27 if such a response was not provided in question #14. Stated differently, the odds of providing extreme straight non-differentiated response in question #18 are 5 times (1/0.20) greater if such a response is provided in question #4 and 3.7 times (1/0.27) greater if such a response is provided in question #14.

4. LIMITATIONS, SUMMARY AND CONCLUSION

The interpretation of these results is, of course, limited by the fact that this is a exploratory study, and that other unmeasured variables may account for the differences in responses. There is also the possibility of context effects and problem of confounds. The response-order effect for an item may have been influenced by context of prior items in the list. For example, in question #4, items "I" and "J" (Economy and Unemployment) are related and could certainly influence each other depending on which one comes first in the list when the order is reversed. Response-order effect can be confounded with question characteristics such as the position of the question, difficulty of answering the question, and topic of the question. We did not account for any of these factors in our analysis. Having done so, this could have explained, for example, why the pattern of extreme responses is different for question #18 (see Figure 6). However, within these limitations, numerous interesting insights nonetheless appeared.

This study confirms response-order effects in mail surveys are real and certainly cannot be ignored. Our research suggests the possibility that response-order effects are especially strong for items or response choices placed in the lower half of the list. Consequently, respondents would rate response items placed at the lower half of the list higher (i.e., by selecting a point at an extreme end of the scale) than if they were to rate the same items placed at the upper half of the list. This finding is consistent with previous research showing response-order effects are stronger for longer questions, questions with longer response options, and those with more difficult language (Payne (1951); Bishop and Smith (2001). We also found the odds of providing a extreme straight non-differentiated response to a target question are affected by providing responses in the same way to questions prior to the target question.

There are lessons to learn about how respondents read response list in surveys. A significant proportion of respondents may not read lists as we would expect, and the unpredictable effects of this behavior could be reduced by simple instructions to respondents to read through the list thoroughly before making their selection. Motivational comments such as *"Your opinions are valuable, please continue..."* before a long response list may help in reducing some of the effects of satisficing, such as non-differentiation. While we can have such comments dynamically shown in Web surveys [10], the dynamics of mail surveys introduce a completely different terrain of challenges and possibilities. Having said that, the future course of this research will focus on some of these limitations while at the same time investigating new ways to obtain a more accurate reading of opinions in mail studies.

5. APPENDIX

Fig. 7. Example of extreme and extreme straight response style

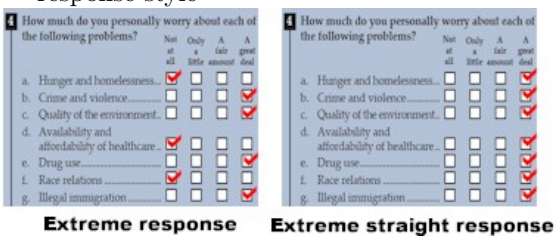
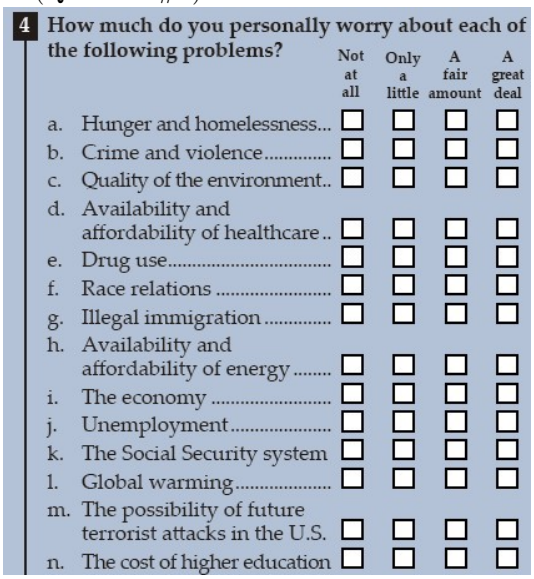
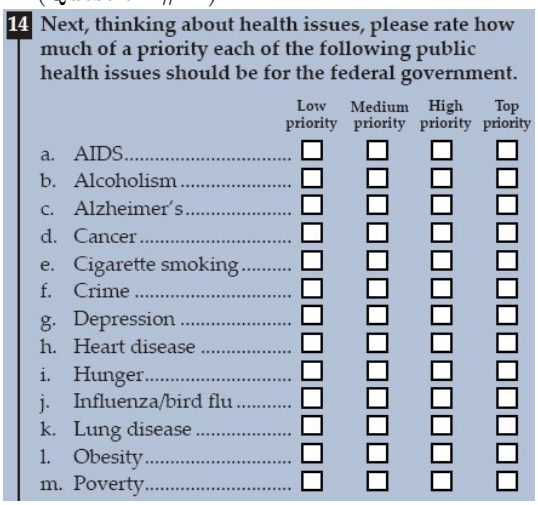


Fig. 8. Illustration of first candidate question (Question #4)



[10] For instance, we can show a pop-up message automatically after a long response list and even vary the linguistic "power" of the message such that those who give a non-differentiated response are shown a message that has a higher degree of coercing for people to be truthful about their selections.

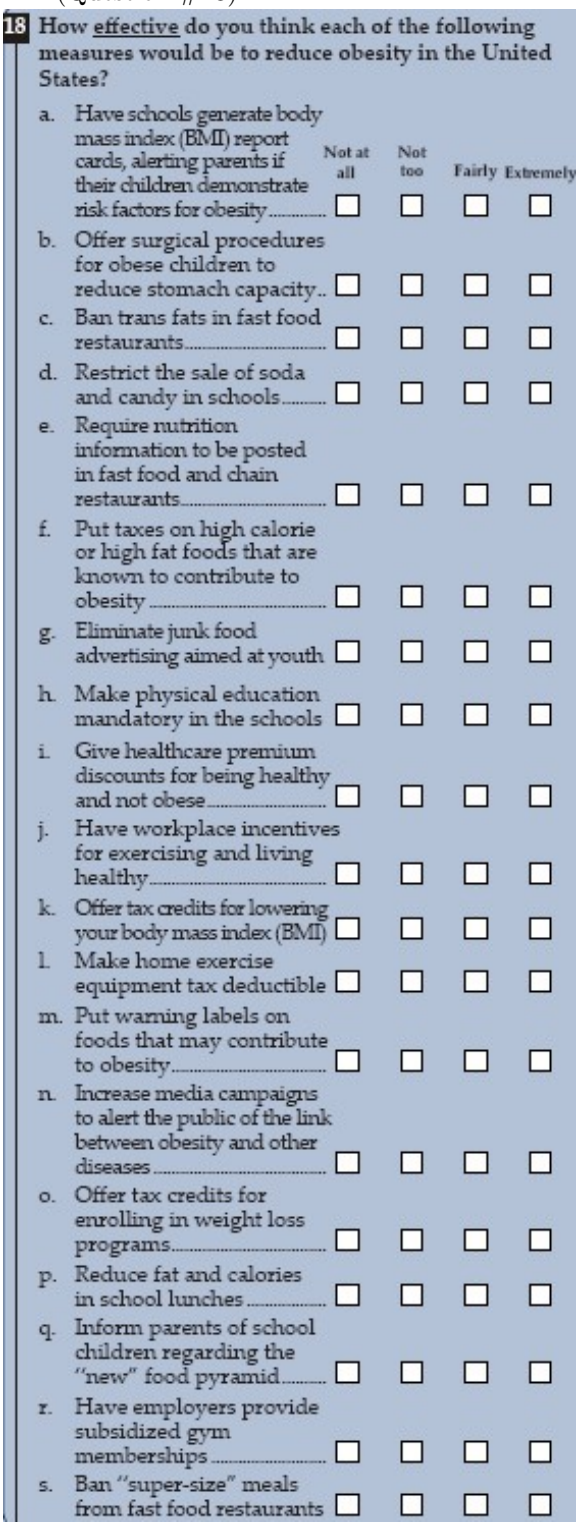
Fig. 9. Illustration of second candidate question (Question #14)



14 Next, thinking about health issues, please rate how much of a priority each of the following public health issues should be for the federal government.

	Low priority	Medium priority	High priority	Top priority
a. AIDS	☐	☐	☐	☐
b. Alcoholism	☐	☐	☐	☐
c. Alzheimer's	☐	☐	☐	☐
d. Cancer	☐	☐	☐	☐
e. Cigarette smoking	☐	☐	☐	☐
f. Crime	☐	☐	☐	☐
g. Depression	☐	☐	☐	☐
h. Heart disease	☐	☐	☐	☐
i. Hunger	☐	☐	☐	☐
j. Influenza/bird flu	☐	☐	☐	☐
k. Lung disease	☐	☐	☐	☐
l. Obesity	☐	☐	☐	☐
m. Poverty	☐	☐	☐	☐

Fig. 10. Illustration of third candidate question (Question #18)



18 How effective do you think each of the following measures would be to reduce obesity in the United States?

	Not at all	Not too	Fairly	Extremely
a. Have schools generate body mass index (BMI) report cards, alerting parents if their children demonstrate risk factors for obesity	☐	☐	☐	☐
b. Offer surgical procedures for obese children to reduce stomach capacity	☐	☐	☐	☐
c. Ban trans fats in fast food restaurants	☐	☐	☐	☐
d. Restrict the sale of soda and candy in schools	☐	☐	☐	☐
e. Require nutrition information to be posted in fast food and chain restaurants	☐	☐	☐	☐
f. Put taxes on high calorie or high fat foods that are known to contribute to obesity	☐	☐	☐	☐
g. Eliminate junk food advertising aimed at youth	☐	☐	☐	☐
h. Make physical education mandatory in the schools	☐	☐	☐	☐
i. Give healthcare premium discounts for being healthy and not obese	☐	☐	☐	☐
j. Have workplace incentives for exercising and living healthy	☐	☐	☐	☐
k. Offer tax credits for lowering your body mass index (BMI)	☐	☐	☐	☐
l. Make home exercise equipment tax deductible	☐	☐	☐	☐
m. Put warning labels on foods that may contribute to obesity	☐	☐	☐	☐
n. Increase media campaigns to alert the public of the link between obesity and other diseases	☐	☐	☐	☐
o. Offer tax credits for enrolling in weight loss programs	☐	☐	☐	☐
p. Reduce fat and calories in school lunches	☐	☐	☐	☐
q. Inform parents of school children regarding the "new" food pyramid	☐	☐	☐	☐
r. Have employers provide subsidized gym memberships	☐	☐	☐	☐
s. Ban "super-size" meals from fast food restaurants	☐	☐	☐	☐

REFERENCES

Bishop, G. and A. Smith (2001). Response-order effects and the early gallup split-ballots. *Public Opinion Quarterly*.

Blankenship, A.B. (1943). *Consumer and Opinion Research*. Harper Collins. New York.

Herzog, A.R. and J.G. Bachman (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly* **45**, 549–59.

Knauper, B. (1995). Response effects in older age. Midwest Association for Public Opinion Research. Chicago.

Kraut, A.L., A.D. Wolfson and A. Rothenberg (1975). Some effects of position on opinion survey items. *Journal of Applied Psychology* **60**, 774–776.

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

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

Krosnick, J. and D. Alwin (1989). A test of the form-resistant correlation hypothesis: Ratings, rankings, and the measurement of values. *Public Opinion Quarterly*.

Krosnick, J., S. Narayan and W.R. Smith (1996). *Satisficing in Surveys: Initial Evidence*. advances in survey research ed.. Jossey-Bass. San Francisco.

McClendon, M. (1986). Response-order effects for dichotomous questions. *Social Science Quarterly* **67**, 205–11.

Mingay, D. J. and M. T. Greenwell (1989). Memory bias and response-order effects. *Journal of Official Statistics* **5**, 253–263.

Payne, S. L. (1951). *The Art of Asking Questions*. Princeton University Press. Princeton, NJ.

Rugg, D. and H. Cantril (1944). *"The Wording of Questions" In Gauging Public Opinion*. Princeton University Press. Princeton, NJ.

Schaeffer, M. E., J. Krosnick, E. G. Langer and M. D. Merkle (2005). Comparing the quality of data obtained by minimally balanced and fully balanced attitude questions. *Public Opinion Quarterly*.

Schuman, H. and S. Presser (1981). *Questions and Answers in Attitude Surveys*. Academic Press. New York.

Schuman, H. and S. Presser (1996). *Questions and answers in attitude surveys : experiments on question form, wording, and context*. Sage. London.

Simon, H. (1957). *Models of Man*. Wiley. New York.

Sudman, S., N.M. Bradburn and N. Schwarz (1996). *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass. San Francisco.