SATISFICING IN WEB SURVEYS: IMPLICATIONS FOR DATA QUALITY AND STRATEGIES FOR REDUCTION

by

Chan Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Survey Methodology)
in the University of Michigan
2013

Doctoral Committee:

Professor Frederick G. Conrad, Chair
Research Scientist Steven G. Heeringa
Associate Professor Frauke Kreuter
Vice President Roger Tourangeau, Westat

Acknowledgements

I am very grateful to have this dissertation committee – Fred Conrad, Steve Heeringa, Frauke Kreuter and Roger Tourangeau. Although they were technically the judges of this dissertation, I never felt I was being judged. They challenged me with the goal of making the research more rigorous. Their comments were always more about how to improve the work, rather than just criticizing it.

Without data, this dissertation would still be ideas in my head. The first two studies used data from the LISS panel (Longitudinal Internet Studies for the Social Sciences) in Netherlands. The LISS panel is a great platform to conduct survey research and other social science research. I hope that the LISS panel continues to thrive. I am also very grateful to Roger Tourangeau, Mick Couper, and Fred Conrad for giving me space in their survey to conduct my third study.

I greatly appreciate the financial support provided by the U.S. Census Bureau for their dissertation fellowship and the Rackham School of Graduate Studies at the University of Michigan for the one term dissertation fellowship. These fellowships helped me to focus on my dissertation in the final year of my PhD study.

Writing is an important part of dissertation. Over the past couple of years my writing skills have greatly improved, thanks to Chris Feak for her truly amazing writing classes. Without her help, writing the dissertation would be much more painful.

I also want to thank Mario Callegaro and Charles DiSogra for offering my first internship, and Stephanie Teasley for my first research assistant position. I was very lucky to have their trust and belief in my early career in survey research.

I would also like to express my gratitude towards the administrative staff at the Program in Survey Methodology. Their excellent work makes everything easier for students. I also give thanks for my fellow students, colleagues and friends in the Program for their advice and encouragement, and cheerful happy hours. I also thank my parents for always believing in me, although I don't think I have done a good job explaining to them what Survey Methodology is about in Chinese. I am grateful to those who have become my "family" in Ann Arbor, especially the Drongowskis.

Finally, I would like to thank the most important person during my PhD study – my advisor and dissertation chair, Fred Conrad. I am extremely grateful for his tremendous input in mentoring me during the past several years. This included weekly meetings (some are Skype chats between continents), reviewing, commenting, and editing almost all the work that I produced, and numerous other things that would be too many to list. I cannot thank enough for Fred's support and guide for my research and career.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

With the increasing use of the Web in mixed mode surveys, especially those conducted by the Census and other federal statistical agencies, it has become more urgent than ever to develop methods to enhance online measurement quality. This dissertation research (including three studies) focuses on respondent satisficing as a source of online measurement errors, and interactive intervention to reduce satisficing behaviors. The first study evaluates speeding (or very fast responding) as an indicator by investigating how it is associated with another well-known satisficing behavior – non-differentiation in grid questions. The second and third studies examine intervention design in Web surveys to curtail respondent satisficing. Specifically, the second study examines whether intervention for different satisficing behaviors could produce different effects on overall response quality. The third study explores whether intervention in Web surveys can induce the feeling of interacting with a human agent. Study 1 shows that respondents who speed more often tend to straightline on more grid questions, suggesting that the tendency to speed is indeed related to satisficing. The results of Study 2 demonstrate that intervention in a survey can have a broad impact of improving respondents' reporting effort, which is not restricted to the satisficing behavior it targets nor the type of survey questions where it occurs. The different intervention designs in Study 3 did not yield consistent differences in respondent behaviors. However, the intervention conditions,

regardless of the design, produced more reports of socially desirable answers compared to the no-intervention condition. This pair of observations – that intervention can help increase respondent effort (Study 2) but also make respondents less willing to disclose undesirable information (Study 3) – seem to converge on one explanation on how intervention works. That is, the interactive feedback about respondents' behaviors may increase their sense of social presence as they complete the online questionnaire. As a result, this may motivate respondents to present themselves in a more positive light as a respondent (by working harder on the survey) as well as a person (by not reporting undesirable information about themselves).

Introduction

Krosnick (1991) borrowed the term "satisficing" from Economists and used it to describe a survey phenomenon in which respondents simply provide a satisfactory as opposed to an optimal answer, when doing so would require substantial cognitive effort. In the paper, Krosnick specified three factors affecting satisficing (task difficulty, respondent ability, and respondent motivation) and discussed various response strategies (often referred to as satisficing behaviors or indicators) that may be employed by a satisficing respondent.

Since its introduction to survey research, satisficing has become a prominent research theme. Satisficing behaviors have been widely examined as a means of evaluating response quality, which would be otherwise very difficult given that accuracy of survey answers are often very hard to measure. For example, satisficing behaviors have been frequently assessed in mode comparison studies to draw conclusions on which survey mode yields better response quality (e.g., Fricker, Galesic, Tourangeau and Yan 2005; Holbrook, Green, and Krosnick 2003).

Over the years, satisficing seems to have become a synonym for low data quality and a problem that survey researchers and practitioners struggle with. An important goal of survey work seems to be that of encouraging optimization (or maximization) and suppressing satisficing among respondents. Optimization, according to Krosnick (1991) occurs when respondents carefully and comprehensively perform all of the cognitive

steps required to answer a survey question. These steps include understanding a question, retrieving relevant information, making judgments and estimating, and reporting an answer (a response process model proposed by Tourangeau and his colleagues; see Tourangeau, Rips and Rasinski 2000, p.7-16). While optimization is in theory possible in the survey response process, in reality it may an unattainable goal. Simon (1955, 1956) proposed that instead of maximizing, satisficing is a rational choice given the limitations of the "environment", which in Simon's model represent various constraints such as alternative choices, physiological and psychological limitations, and goals. In a survey setting, similar constraints may lead people to satisfice as opposed to optimize when answering survey questions. In the case where optimizing requires a substantial amount of mental effort with no apparent gain or pay-off of doing so, satisficing, as opposed to optimizing, seems to be the rational choice for respondents.

Even though it may be reasonable for respondents to satisfice, satisficing can compromise the quality of their responses. The impact on data quality depends on the degree of satisficing, which range from almost no effort at all (e.g., randomly choosing a response option) to substantial but still not maximal effort (e.g., providing informative and detailed, but not exhaustive, responses to an open-ended question). Krosnick (1991) distinguished two forms of satisficing: weak satisficing (all cognitive stages are performed by respondents when answering a question, but less thoroughly) and strong satisficing (retrieval and judgment are omitted). He proposed that strong satisficing and optimizing constitute two ends of a continuum of thoroughness in answering a question (see Figure i).

Here, I propose a three-level distinction with regard to respondent effort: optimization (or maximum), attainable maximum and actual achievement (see Figure II). Optimization, same as in Krosnick's framework, represents most thoroughness in the responding process. The next level is attainable maximum. The gap between optimization and attainable maximum takes into account the limitations imposed by respondent ability and task difficulty. In the case of an easy task and high respondent ability, the attainable maximum can reach optimization. Another level down is actual achievement of thoroughness. The difference between the attainable maximum and actual achievement is mainly determined by respondent motivation to expend effort in answering a question. Krosnick (1991) identified a variety of possible determinants of respondent motivation, including need for cognition, topic interest, perceived survey importance, interviewer behavior, accountability, and the length of interview preceding the question. In addition to these factors, it seems quite plausible to assume that motivation can also be affected by task difficulty and ability. Specifically, the motivation could be thwarted by a combination of a difficult task and low ability respondents, and boosted if the conditions are reversed.

Respondent ability, task difficulty and respondent motivation are the three factors affecting satisficing originally proposed in Krosnick's framework. Rather than discussing these three factors as jointly determining satisficing, the current proposed framework uses them to define the attainable maximum and actual achievement. This framework implies that the actual achieved thoroughness can be very close to the attainable maximum, but can never exceed it; and the same relationship exists between the attainable maximum and optimization. Thus, if a question is difficult to answer (e.g., "In the past 6 months,

how much did you spend on purchasing clothes?"), no matter how motivated respondents are, they can only use certain strategies (e.g., rough estimation) to achieve a satisfactory answer. This framework also suggests that to reduce satisficing, it is important to first consider ways to reduce mental burden on respondents. For example, if an open numeric answer requires a daunting recall process, it may help to shorten the reference period or offer closed-ended response options instead.

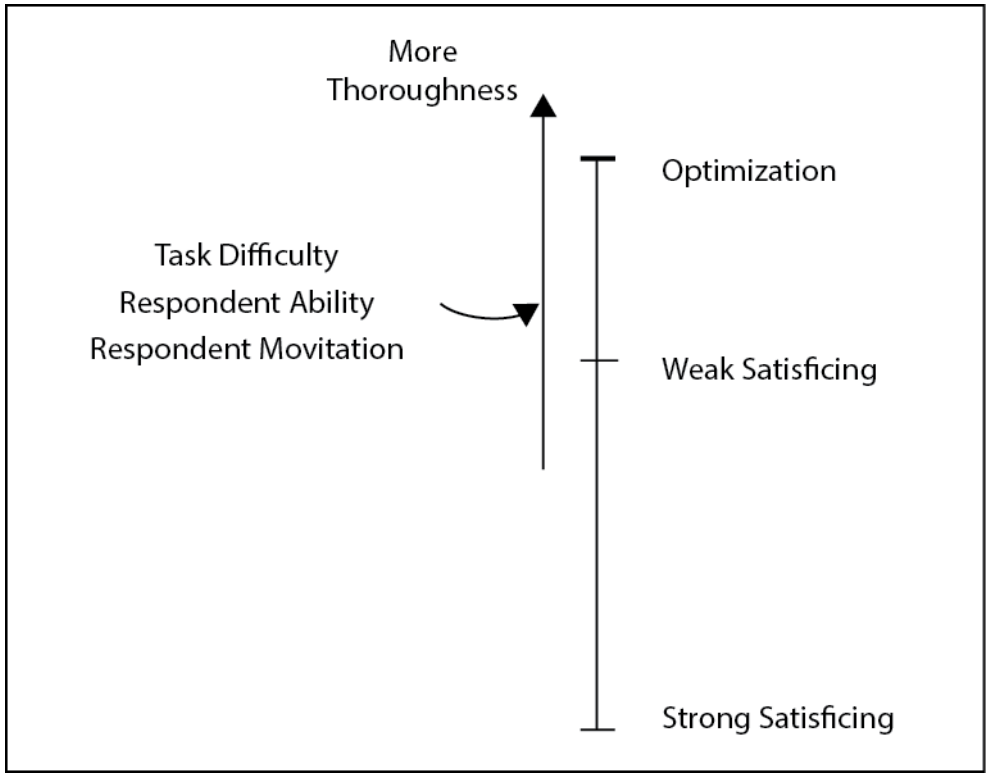


Figure i. Satisficing framework proposed by Krosnick (1991)

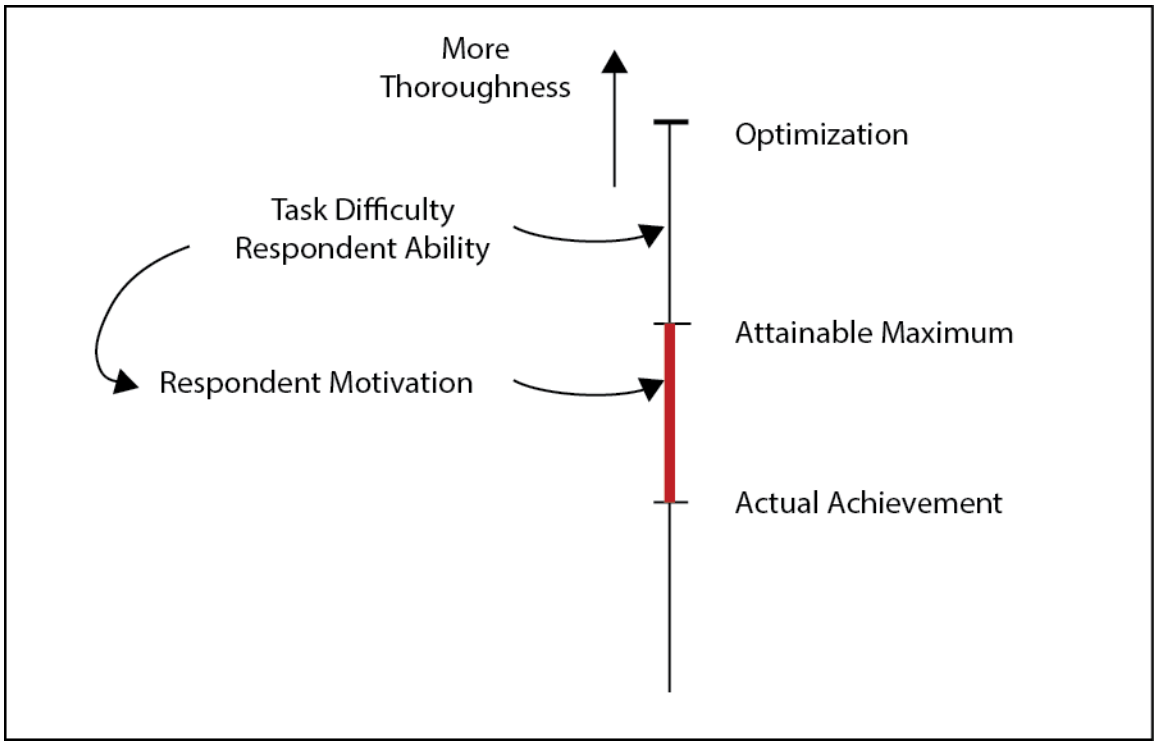


Figure ii. Revised satisficing framework

Focus of This Dissertation

The current dissertation research investigates satisficing due to lack of motivation to expend mental effort on answering a question – i.e., when a respondent's actual performance fails to reach the attainable maximum of thoroughness (the gap shown in red in Figure ii. In other words, this research focuses on lack of thoroughness in response process as a result of unmotivated respondents, in contrast to the situation that more thoroughness is difficult to achieve because the task is too difficult given respondents' cognitive ability.

Many of the satisficing strategies discussed in the literature are likely to arise from inadequate motivation. Specifically, this dissertation research focuses on two behaviors: very fast responding (or speeding) and non-differentiation in grid questions. These

behaviors occur not likely because respondents cannot afford to expend more effort; rather they arise most likely because they are not motivated to invest more effort in answering the question. (For example, speeding is most likely when respondents are unwilling to spend more time thinking about answers, not when they cannot.)

This research focuses on Web surveys. This is because respondents taking self-administered surveys may be particularly unmotivated compared to interviewer-administered surveys where interviewers can help engage respondents and motivate them to expend effort. Among various self-administered modes, I chose to examine Web surveys because of the rich paradata that Web surveys provide on how respondents fill out a questionnaire. These data can be used to identify unmotivated respondents. In addition, the interactivity of the Web allows for the design of interactive prompts in Web surveys that may help curtail satisficing behaviors and encourage respondents to spend extra time and effort. This could be the result of clearer communication with respondents with regard to the expected level of effort and precision in their answers, which may motivate respondents to increase effort to meet expectation of those collecting the data. Moreover, prompts may also induce a sense of accountability if respondents feel that their effort-saving strategies are noticed by the survey organization.

This dissertation consists of three main chapters corresponding to three studies. Study 1 evaluates whether speeding can be useful to capture unmotivated respondents by investigating the relationship between speeding and straightlining. Given that unmotivated respondents often engage in more than one satisficing behavior during a survey (a finding in Study 1), Study 2 addresses the question as to which satisficing behaviors warrant intervention prompts, and whether prompts in response to different

satisficing behaviors have different effects on data quality. Specifically, an experiment is conducted to compare two types of interactive prompts in grid questions: one targeting speeding and the other targeting non-differentiation. Study 3 explores another aspect of the design of these interactive prompts. Specifically, Study 3 compares human-like and computer-like prompts with regard to their impact on data quality.

References

Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly, 69*(3), 370.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires - comparisons of respondent satisficing and social desirability response bias. *Public Opinion Quarterly, 67*(1), 79.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129-138.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics, 69*(1), 99.

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

CHAPTER 1: Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining (Study1)

1.1 Background

A major concern about self-report data is satisficing – the tendency of respondents to provide satisfactory but not optimal answers (Krosnick 1991). Satisficing is generally assumed to reflect the expenditure of inadequate effort by respondents. While satisficing can be difficult to detect in paper and pencil surveys, Web surveys can capture various data on respondents' behaviors (i.e., process data or *paradata*) that can be used to assess satisficing and its impact on data quality. Examples of these paradata include keystroke[1] and mouse-movement data.

Perhaps the most commonly analyzed Web survey paradata is response time, usually defined as the time from when a page is loaded (question is displayed) until the answer is submitted. Fast responding (or "speeding") is often considered as evidence of satisficing and low quality data. However, fast responses are not always problematic. Respondents can sometimes answer a question both quickly and accurately depending on factors such as task difficulty and accessibility of an attitude. Therefore, it is not quite clear what speeding really means for response quality. A related question, on which there is a lack of relevant research, is whether speeding is related to other satisficing behaviors

[1] Heerwegh (2003) discusses the use of JavaScript to capture various respondent actions involving clicks and keystrokes.

that have been identified in the literature, such as primacy/recency effects, acquiescence, and non-differentiation in using rating scales. The idea is that if speeding is related to these satisficing behaviors, we probably can claim more confidently that speeding is a useful indicator for satisficing and low response quality. However, the only published findings seem to be that by Malhotra (2008), who found that shorter completion times were associated with stronger primacy effects among low education respondents.

The aim of this study is to further the understanding of speeding in Web surveys and its implications for data quality. I believe speeding can be a useful tool to identify satisficing respondents. Specifically, response time thresholds (below which a respondent can be said to be speeding) can be set low enough that accurate answers are unlikely. In addition, repeated speeding over an entire questionnaire, as opposed to speeding on a particular question, is likely to reflect a motivation to rush through a survey and therefore can be a good indicator for satisficing. This study also explores the association between speeding and another well-known satisficing behavior in Web surveys – straightlining (or non-differentiation) in grid questions. Speeding is expected to be positively related to straightlining, since both are likely to be the response strategies of a satisficing respondent. This study examines two issues: (1) the characteristics of respondents who speed more frequently than others throughout a survey (referred to as "persistent speeders"), and (2) whether persistent speeders are more likely to engage in straightlining.

1.2 Methods

1.2.1 Dataset and Calculation of Response Time

The data analyzed in this study are from the wave 5 Politics and Values Survey conducted by the MESS project (http://www.centerdata.nl/en/MESS) and administered to its LISS panel (Longitudinal Internet Studies for the Social Sciences). LISS panel is a probability-based Web panel of households in the Netherlands drawn from the population register by Statistics Netherlands. The Politics and Values Survey is one of the core LISS panel surveys that are conducted annually. The survey on which the current article is based was fielded in December 2011 and again in January 2012 for the December non-respondents. The participation rate is 78.9% (5,814 completes out of 7,372 invited panel members).

Response times are calculated as the elapsed time between submission of an answer to the previous and current question. Among the 5,814 respondents, 291 respondents have missing timestamps on one or more questions. Because this study compares respondents on their speeding status over the entire questionnaire, these 291 respondents are not included in the analyses. I also exclude follow-up questions, because they are not administered to all respondents. If respondents answer a question more than once, it is only the first response time that is considered in the determination of speeding.

The final dataset for the analyses include 5,523 panelists, with response times for 54 questions. (Appendix 1.1 shows the demographic distributions of the respondents in the final dataset.)

1.2.2. Threshold for Speeding

In principle, any response time that is shorter than the optimal response time (i.e., the amount of time required to produce the optimal response) can be considered an instance of speeding. Although this is conceptually straightforward, in practice it is very

difficult to determine the optimal response time, since it depends on a variety of factors. Although Yan and Tourangeau (2007) analyzed the influences of question-level and respondent-level characteristics on response times, it should be noted that the response times examined in their study were the actual time respondents spent, which can be very different from the optimal time required to answer the questions accurately.

This study employs a simple measure of speeding. Specifically, I set the speeding threshold as 300 milliseconds (msec) per word, a rough estimate of reading speed, multiplied by the number of words in the question[2]. The idea behind this approach is that when response times are faster than likely reading times, respondents are unlikely to have given the question adequate thought. Note that this study does not attempt to accurately determine speeding on individual questions for individual respondents. Rather, our goal is to use this generic threshold to identify the respondents whose response times tend to fall on the lower end more often than others.

1.3. Results

1.3.1 Prevalence of Speeding

On average respondents sped on about 15 out of the 54 questions (see Table 1.1). Respondents varied considerably in how often they sped, with those in the top quartile speeding on 21 or more out of 54 questions (see Table 1.1). The question, then, is whether the respondents who sped more often did so consistently throughout the questionnaire. To examine this, I divided the questionnaire into two parts with completion times on average about the same in each part. For each part of the

[2] This reading speed, 300 msec per word, is slower than the typical reading rate among college students for comprehension, which is about 200 msec per word as found in a number of early studies (e.g., Carver 1992).

questionnaire, I grouped respondents into four quartiles based on the number of questions on which they sped, and compared their speeding status between the first and second half the questionnaire.

Table 1. 1. Mean and Quartiles of the Numbers of Questions on Which Respondents Sped (out of a total of 54 questions)

Mean	25th Percentile	Median	75th Percentile
15.4	8	14	21

As shown in Table 1.2 respondents' speeding tendency was relatively consistent across the questionnaire: among respondents who sped least frequently (1st quartile) and most frequently (4th quartile) in the first part of the questionnaire, the majority (66.9% and 65.0%, respectively) remained in the same quartile for the second part of the questionnaire in terms of their speeding frequency relative to other respondents. These findings suggest that speeding is not a random behavior, but a characteristic of respondents.

Table 1. 2. Comparisons of Speeding Frequency in the First and Second Half of the Questionnaire*

	Speeding Frequency in the **SECOND Part** of the Questionnaire				
Speeding Frequency in the **FIRST Part** of the Questionnaire	1st Quartile (least)	2nd Quartile	3rd Quartile	4th Quartile (most)	Total
1st Quartile (least)	66.9%	28.7%	4.3%	0.2%	100%
2nd Quartile	23.0%	50.3%	23.1%	3.6%	100%
3rd Quartile	4.4%	26.8%	45.2%	23.7%	100%
4th Quartile (most)	1.9%	9.5%	23.7%	**65.0%**	100%

*Quartiles are calculated based on the number of questions on which respondents sped in each part of the questionnaire.

In this study, I am particularly interested in the group of respondents ranked in the highest quartile of speeding frequency in both parts of the questionnaire (bolded in Table

1.2). These respondents are referred to as "persistent speeders" in this study. We next examine the characteristics of these persistent speeders.

1.3.2 Characteristics of Persistent Speeders

I used logistic regression to model the likelihood of being a persistent speeder (i.e., being in the highest quartile of speeding frequency in both parts of the questionnaire). The explanatory variables included age (18-34, 35-44, 45-54, 55-64, >=65), gender, level of education (primary school, junior high, senior high, junior college, college, university)[3], respondent origin (Dutch vs. first/second-generation immigrants), tenure on the panel (whether the household joined the panel in 2007 or after), early vs. late respondents (whether the respondent completed the survey in December or in the non-respondent follow-up in January), and whether the household received any device (computer[4], Internet connection, or both) from the panel to complete surveys. Table 1.3 below presents the regression results from the final model. Since education, gender and origin do not have significant impacts given the other covariates in the model, they were not included in the final model.

The regression reveals a strong monotonic decrease in persistent speeding as respondents get older. This can also be seen in Figure 1.1, which shows the percentage of persistent speeders dropping from over 40% among those age 18-34 to less than 5% among those 65 and older. Certainly this pattern reflects the established finding that older respondents tend to be slower than younger respondents because of cognitive aging (cf. Schwarz, Park, Knauper, and Sudman 1999).

[3] These education categories are used by CBS (Statistics Netherlands).
[4] Either a laptop or a simPC is provided. SimPC is a small and simple computer (more information at http://www.lissdata.nl/lissdata/About_the_Panel/Equipment).

Table 1. 3. Parameter Estimate for Final Logistic Model of Persistent Speeding

Parameter	Estimate (β)	Standard Error	Odds Ratio (e^{β})	*p*-value of χ^2
Intercept	0.041	0.08	1.042	0.5960
(ref: 18-34)				
35-44	-0.845	0.10	0.430	<.0001
45-54	-1.333	0.11	0.264	<.0001
55-64	-2.199	0.13	0.111	<.0001
>=65	-3.474	0.21	0.031	<.0001
Received any device *(ref: no device received)*	-0.564	0.19	0.569	0.0034
Joined the panel after 2007 *(ref: Joined before 2007)*	-1.077	0.12	0.341	<.0001
Responded in Jan *(ref: responded in Dec)*	-0.334	0.11	0.716	0.0020
Responded in Jan × Joined the panel after 2007	0.896	0.27	2.450	0.0008

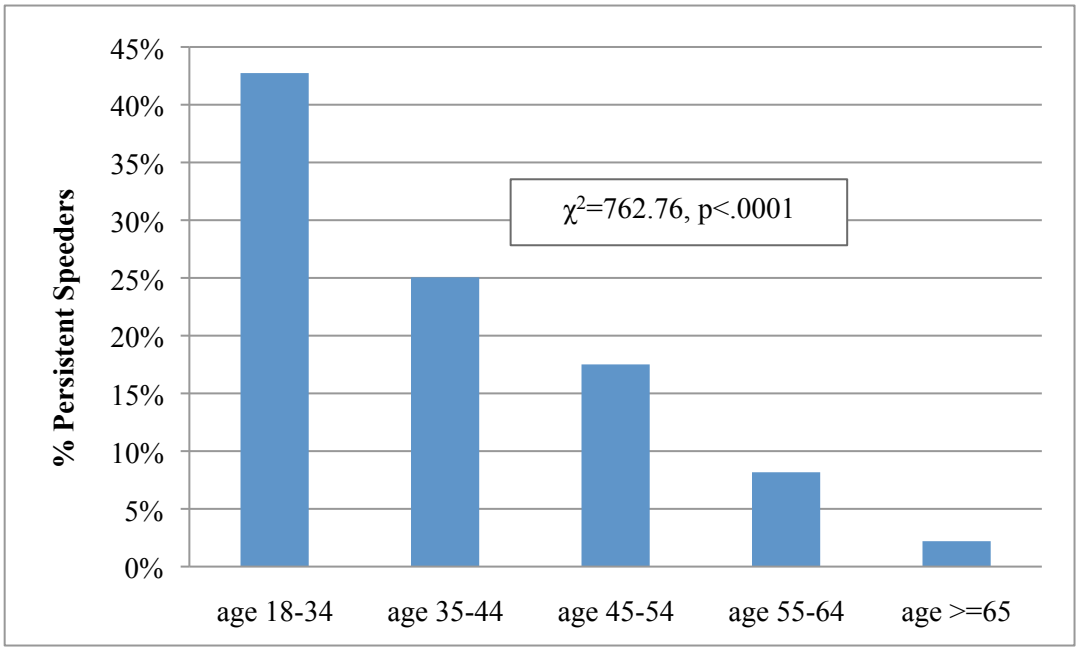


Figure 1. 1. Percentage of Persistent Speeders by Age Groups

The regression analyses also show lower prevalence of persistent speeders among respondents who received any device from the panel compared to others. One possible explanation is that respondents with a device provided by the panel may have less experience using computers and the Internet and, therefore, may take more time to navigate through the questionnaire than other respondents. Another possibility is that these respondents may feel more obliged to expend effort on the surveys because the survey organization has provided them the device.

Regarding the effect of tenure, I expected that speeding might be more prevalent among respondents with longer tenure on the panel. Respondents who had been on the panel longer might be more familiar with the questions (as the survey is conducted annually with some questions recurring). Thus, these respondents might need less time to complete the questionnaire, as a result of less reading time of the familiar questions or less time answering them. In addition, these veteran respondents might be more subject to survey fatigue compared to the newer panel members and therefore, they might be more likely to rush through the survey. As predicted, Table 1.3 shows less persistent speeding among respondents who joined the panel after 2007 compared to those joined the panel earlier, although this tenure effect seems to be smaller among January respondents (December non-respondents) than that among the December respondents (odds ratio is 0.341 for December respondents and $0.341 \times 2.450 = 0.834$ for January respondents).

Regarding the differences between early and late respondents, researchers generally believe that response propensity can be associated with response quality if there are some common factors (e.g., interest in the topic) that correlate with both people's decision to participate and the level of effort they are willing to spend in the survey. However,

evidence is mixed regarding the relationship between response propensity and response quality. (For example, this relationship is found in Fricker and Tourangeau (2010), but is only spurious in Kaminska et al. (2010).) So in this study I speculated that, if there is any difference, January respondents (December non-respondents) will be more likely to speed persistently than the December respondents. As shown in Table 1.3, this is only observed among those who joined the panel after 2007 (odds ratio=0.716 × 2.450=1.754). Among those who joined the panel earlier, the difference is reversed (odds ratio=0.716). One explanation is that respondents who have been in the panel longer are generally more committed to the study than newer respondents; nonresponse among relatively committed panel members may be less related to response quality than is nonresponse among those who are less committed.

While there are clear demographic predictors of persistent speeding, the value in identifying speeders depends on whether speeding is associated with reduced data quality. If fast responding results from respondents' unwillingness to expend effort answering the questions, then persistent speeders should be more likely to exhibit satisficing, in particular straightlining. This is examined in the next section.

1.3.3 Relationship between Speeding and Straightlining

Toward the end of the survey, respondents answered a series of grid questions. These are a type of survey question in which multiple items with the same rating scales are displayed in a matrix, with the items being the rows and the rating scales being the columns. These grid questions in the questionnaire asked about opinions on a variety of topics, mostly on 5-point scales (see Appendix 1.2 for the wording). The analysis focused

on straightlining – i.e., choosing the same response option for all the items in a grid so that the selected answers are in a vertical line. (One grid question with only two statements was excluded.)

I first investigated at the question-level how speeding is associated with straightlining. The findings are shown on the left side of Table 1.4. Across all the eight grid questions examined, respondents who sped on the question were substantially more likely to straightline. In addition, if the tendency to speed relates to satisficing, we can expect the previous speeding behaviors to be also indicative of response quality on the grid questions. This is shown on the right side of Table 1.4 which compares the amount of straightlining between the respondents who sped a lot before the grid questions (in the top quartile in terms of the speeding instances) and those who sped less often. The straightlining in the grid questions was positively correlated with the earlier speeding frequency, although the correlation was not as strong as when speeding was evaluated with regard to the grid questions.

I then examined the association between the overall speeding tendency throughout the questionnaire and straightlining. The simple bivariate analysis showed that across the eight grid questions, persistent speeders on average straightlined on approximately two questions, while others on average straightlined on approximately one question (1.9 vs. 1.0, t= -20.28, p<.0001).

I also conducted regression analyses to further understand the relationship between speeding and straightlining, controlling for the demographic variables. Specifically, I used negative binomial regressions to model the number of grid questions on which

respondents straightlined.[5] The explanatory variables included respondents' speeding tendency (persistent speeder vs. not) as well as the demographic variables that were included in the speeding model (i.e., age, gender, education, origin, tenure, early vs. late respondents, and whether received any device from the panel). In addition to the main effects, I also tested the interaction effects between speeding tendency and the demographic variables. The findings are shown in Table 1.5 (only the effects significant at the 0.05 level are included).

[5] The dependent variable here (i.e., the number of grid questions where respondents straightlined) is essentially count data and we have tried both Poisson regressions and negative binomial regressions. Because of the overdispersion issue with the Poisson model, we prefer the negative binomial models and report the results here.

Table 1. 4. Comparison of Straightlining Based on Speeding Status <u>on</u> and <u>before</u> the Grid Question

	% Straightlining			% Straightlining		
Grid Questions	Speeding	Not Speeding	*diff*	Speeding frequently before the grids (top quartile)	Others	*diff*
Working mothers	28.3%	9.1%	***	16.4%	10.5%	***
Role of father/mother in households	21.7%	8.3%	***	19.5%	10.4%	***
Foreigners/immigrants	32.4%	0.9%	***	9.4%	1.7%	***
Marriage	30.6%	0.7%	***	9.6%	1.8%	***
Taking care of parents	41.8%	13.8%	***	28.7%	16.8%	***
Women w/ kids & working	49.6%	26.9%	***	37.0%	30.4%	***
Attitudes towards women	27.6%	12.9%	***	28.5%	17.7%	***
Questionnaire evaluations	21.6%	8.4%	***	13.9%	7.9%	***

*** *p*-value of CHISQ test <.0001

Table 1. 5. Parameter Estimate for Final Negative Binomial Model to Predict Number of Grid Questions on Which Respondents Straightlined (across 8 grid questions)

Parameter	Estimate (β)	Standard Error	Rate Ratio (e^{β})	*p*-value of CHISQ
Intercept	0.165	0.07	1.180	0.0209
Persistent speeder *(ref: not)*	1.125	0.11	3.081	<.0001
(ref: 18-34)				
35-44	0.035	0.05	1.036	0.4555
45-54	-0.011	0.05	0.989	0.8169
55-64	-0.139	0.05	0.871	0.0038
>=65	-0.300	0.05	0.741	<.0001
Female *(ref: Male)*	0.137	0.03	1.146	<.0001
Persistent speeder x female	-0.154	0.07	0.857	0.0204
(ref: primary school)				
Junior high	-0.125	0.07	0.883	0.0579
Senior high	-0.150	0.08	0.860	0.0600
Junior college	-0.201	0.07	0.818	0.0035
College	-0.016	0.07	0.984	0.8102
University	-0.026	0.08	0.975	0.7586
Persistent speeder × junior high	-0.367	0.13	0.693	0.0041
Persistent speeder × senior high	-0.630	0.14	0.533	<.0001
Persistent speeder × junior college	-0.395	0.13	0.674	0.0017
Persistent speeder × college	-0.810	0.13	0.445	<.0001
Persistent speeder × university	-0.835	0.15	0.434	<.0001
Joined the panel after 2007 *(ref: in 2007)*	-0.195	0.04	0.823	<.0001
*Dispersion coefficient**	*0.211*	*0.02*		

***** This is constrained to zero for Poisson models. A positive value suggests overdispersion (i.e., observed variability in the dependent variable exceeds that predicted by the Poisson model).

Among all the explanatory variables, respondent speeding tendency (i.e., whether respondents engaged in persistent speeding or not) seems to have the strongest impact on how often they straightlined. As shown in Table 1.5, persistent speeders are expected to straightline about 3 times as many grid questions as did others ($e^{1.125}=3.081$).

With regard to education, I found both the main effects, as well as strong interactions with the speeding tendency. This can also be seen in Figure 1.2 which shows the mean number of straightlined answers (across eight grid questions) by respondents' overall speeding status and further by their education levels. While persistent speeders appear to straightline more than others across all educational groups, the effects are particularly large among the less educated respondents. If the focus is on the impact of education on straightlining, Figure 1.2 tells us that when respondents engaged in persistent speeding, the amount of straightlining increased considerably among the less educated groups but when respondents do not speed persistently, the level of straightlining is quite similar across education groups.

Another finding is that overall straightlining was somewhat less common among the older respondents. However, unlike education, there was no significant interaction between speeding status and age. This suggests that the positive association between speeding and straightlining seems to be similar across all age groups. This finding does not support the argument that speeding is less problematic for the young respondents because they are cognitively faster (i.e., they can read and retrieve information quickly). Rather, the evidence suggests that although younger respondents may be cognitively capable to answer more quickly, too quickly (i.e., speeding) is still problematic and likely to reflect satisficing.

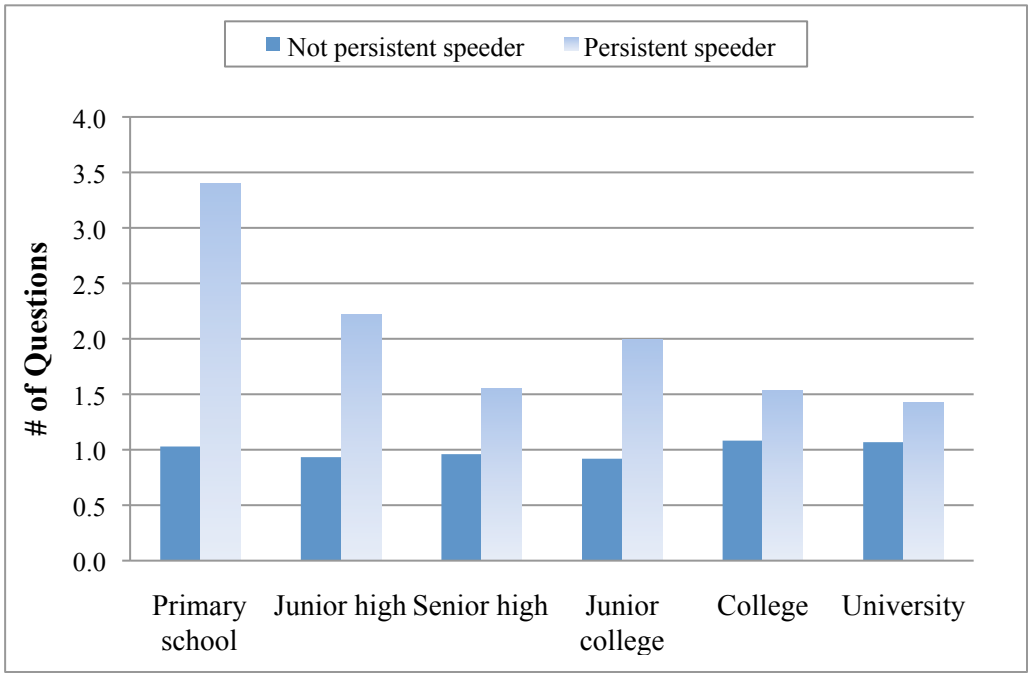


Figure 1. 2. Number of Grid Questions with Straightlining Answers by Speeding Tendency and Education

I also found an interaction between gender and speeding tendency where the effect of speeding on straightlining seems to be slightly reduced among female respondents. (This can be seen from the negative but small coefficient of the interaction between speeding and gender in Table 1.5.) In addition, the regression also shows that straightlining was somewhat less among the respondents who joined the panel after 2007 compared to those who joined the panel earlier.

1.4. Discussion

This study shows that the tendency to speed is strongly related to age (substantially more among younger respondents), with no further impact of respondent education. The analyses of speeding and straightlining reveal a strong positive association between these two behaviors even when various demographics are controlled. In other words, it seems

to be a universal phenomenon that people straightline more as they speed, suggesting that speeding can be a very useful indicator for satisficing and response quality.

The relationship between speeding and straightlining appears to be strongly moderated by respondent educational level. Specifically, speeding seems to only modestly increase straightlining among highly educated respondents, but has a much more dramatic impact among those with relatively low levels of education. This suggests that the very educated respondents may be able to read and comprehend the questions quickly and at the same time give quality responses (i.e., without straightlining), while this could be more challenging for respondents with a relatively low level of education. This finding suggests the value of strategies to slow down respondents, especially those with low education levels.

Moreover, the extent to which speeding relates to straightlining is surprisingly similar between younger and older respondents. This suggests that greater cognitive ability (less cognitive aging) cannot be the only reason why younger respondents are faster, and that at least some of the speeding among the young results from satisficing.

Web surveys vary in many aspects, from recruiting methods to target populations (Couper 2000). These differences can affect the level of effort people are willing to expend on a survey, and therefore the overall speeding tendency. However, it is worth noting that this study focuses on relative speeding frequency and most of the findings are based on the comparisons of persistent speeders (i.e., respondents who speed more than others across the entire questionnaire) and other respondents. Therefore, the major conclusions from this study regarding who tends to speed and the impact on response quality is likely to have broad implications for other types of Web surveys.

Appendix 1. 1. Respondent Demographics in the Final Dataset

	Final dataset
n	5523
Gender	
Male	46.7%
Female	53.3%
Age	
18-34	20.1%
35-44	16.7%
45-54	19.3%
55-64	21.7%
>=65	22.2%
Education	
Primary school	8.0%
Junior high	25.9%
Senior high	10.9%
Junior college	23.6%
college	23.1%
university	8.5%
Origin	
Dutch	11.5%
Foreigner	88.5%
Tenure	
Joined the panel in 2007	73.0%
Joined the panel after 2007	27.0%
Early vs. late respondents	
Responded in December 2011	83.5%
Responded in January 2012	16.5%
Received any device	8.6%

Appendix 1. 2. Wording of the Eight Grid Questions

Grid_Q1:
For each statement, please indicate to what extent you agree or disagree.

A working mother's relationship with her children can be just as close and warm as that of a non-working mother.

A child that is not yet attending school is likely to suffer the consequences if his or her mother has a job.

Overall, family life suffers the consequences if the mother has a full-time job.

1 fully disagree
2 disagree
3 neither agree nor disagree
4 agree
5 fully agree

Grid_Q2:
And to what extent do you agree or disagree with the following statements?

Both father and mother should contribute to the family income.

The father should earn money, while the mother takes care of the household and the family.

Fathers ought to do more in terms of household work than they do at present.

Fathers ought to do more in terms of childcare than they do at present.

1 fully disagree
2 disagree
3 neither agree nor disagree
4 agree
5 fully agree

Grid_Q3:
What is your opinion on the following statements?

It is good if society consists of people from different cultures.

It is difficult for a foreigner to be accepted in the Netherlands while retaining his/her own culture.

It should be made easier to obtain asylum in the Netherlands.

Legally residing foreigners should be entitled to the same social security as Dutch citizens.

There are too many people of foreign origin or descent in the Netherlands.

People of foreign origin or descent are not accepted in the Netherlands.

Some sectors of the economy can only continue to function because people of foreign origin or descent work there.

It does not help a neighborhood if many people of foreign origin or descent move in.

1 fully disagree
2 disagree
3 neither agree nor disagree
4 agree
5 fully agree

Grid_Q4:
What is your opinion on the following statements?

Married people are generally happier than unmarried people.
People that want to have children should get married.
A single parent can raise a child just as well as two parents together.
It is perfectly fine for a couple to live together without marriage intentions.
For a couple that wants to get married, it is good to first start living together.
A divorce is generally the best solution if a married couple cannot solve their marital problems.
It is all right for a married couple with children to get divorced.

1 fully disagree
2 disagree
3 neither agree nor disagree
4 agree
5 fully agree

Grid_Q5:
What is your opinion on the following statements?

Children ought to care for their sick parents.
When parents reach old age, they should be able to live with their children.
Children that live close by ought to visit their parents at least once a week.
Children ought to take unpaid leave in order to care for their sick parents.

1 fully disagree
2 disagree
3 neither agree nor disagree
4 agree
5 fully agree

Grid_Q6:
Do you think that women, under the circumstances described below, should be able to have a full-time job, a part-time job, or no job at all?

If she has a baby (a child younger than 1 year).
If she has a child that does not yet attend school.

After the youngest child starts primary school.
After the youngest child starts secondary school.

1 full-time
2 part-time
3 no job at all

Grid_Q7:
The following statements are on marriage, the duties of husbands and wives, and about rearing boys and girls. Please read each statement and indicate to what extent you agree or disagree.

A woman is more suited to rearing young children than a man.
It is actually less important for a girl than for a boy to get a good education.
Generally speaking, boys can be reared more liberally than girls.
It is unnatural for women in firms to have control over men.

1 fully disagree
2 disagree
3 neither agree nor disagree
4 agree
5 fully agree

Grid_Q8:
Finally; what did you think of this questionnaire?

Was it difficult to answer the questions?
Were the questions sufficiently clear?
Did the questionnaire get you thinking about things?
Was it an interesting subject?
Did you enjoy answering the questions?

1 certainly not
2
3
4
5 certainly yes

References

Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading, 36*(2), 84-95.

Couper, M. P. (2000). Web surveys - A review of issues and approaches. *Public Opinion Quarterly, 64*(4), 464.

Fricker, S., & Tourangeau, R. (2010). Examining the relationship between nonresponse propensity and data quality in two national household surveys. *Public Opinion Quarterly, 74*(5), 934-955.

Heerwegh, D. (2003). Explaining response latencies and changing answers using client-side paradata from a web survey. *Social Science Computer Review, 21*(3), 360-373.

Kaminska, O., McCutcheon, A., & Billiet, J. (2010). Satisficing among reluctant respondents in a cross-national context. *Public Opinion Quarterly, 74*(5), 956-984.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly, 72*(5), 914-934.

Schwarz, N., Park, D., Knauper, B., & Sudman, S. (1999). *Cognition, aging, and self-reports*. Philadelphia, PA: Psychology Press.

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology, 22*(1), 51-68.

CHAPTER 2: Intervention for Satisficing Behaviors in Web Surveys: Different targeted behaviors and similar impact on response quality (Study 2)

2.1 Background

The proliferation of Web surveys is one of the most noticeable phenomena in the survey industry during the past decade (Couper 2000; Couper and Miller 2008). For a long time the growth of Web surveys has mostly occurred in market research or surveys of high internet-penetration populations (e.g., the young and the educated). Recently, along with the rise of mixed-mode designs, Web has been increasingly offered as a mode option in general household surveys, including those conducted by federal statistical agencies. (For example, the American Community Survey has just added the option of responding online.) Lately, a growing interest in collecting survey data from mobile devices has increased the importance of Web surveys even further.

Web surveys have many attractive features. In particular, its cost advantage and convenience (especially with the development of mobile Web surveys) may hold keys to meeting some of the critical challenges survey researchers face today. However, one major concern with Web surveys, as with other self-administered modes, is that respondents have to be self-motivated to expend sufficient effort. When respondents fail to invest sufficient effort to respond thoughtfully – a phenomenon often described as "satisficing" (cf. Krosnick 1991), measurement error increases and data quality is compromised.

In telephone and face-to-face surveys, it is the interviewers' task to engage and motivate respondents to provide accurate responses. In the case of self-administered surveys, however, it is much more challenging to influence respondent motivation. A small number of studies have attempted to address the issue. The early work by Charles Cannell and his colleagues showed that a commitment procedure – asking respondents to agree to "be conscientious and hard-working" – was effective in improving respondent performance in interviewer-administered surveys, although the effect was smaller in telephone than face-to-face surveys (Miller and Cannell 1982). In a very recent study, Conrad et al. (2011) applied this commitment technique to a Web survey of opt-in panelists and found that the commitment condition (where almost all the respondents chose to commit) reduced speeding compared to the control condition. These findings suggest the possibility of improving the response quality of a Web survey by making respondents commit to expending effort.

While commitment can be considered a preventive approach to curtailing satisficing, a few recent studies have started to explore a corrective approach – that is, to intervene with respondents when behaviors likely to result from satisficing are detected. These behaviors are often referred to as satisficing behaviors and examples include speeding (responding too quickly to have given adequate thought for accurate responding), item nonresponse (skipping a question without providing an answer), and non-differentiation (giving the same or very similar ratings to all the items in a grid or matrix of questions).

So far, this intervention technique in Web surveys has been tested in a few studies to curtail speeding (Conrad et al. 2009; Conrad et al. 2011), reduce item nonresponse

(DeRouvray and Couper 2002), and increase answers to open-ended questions (Holland and Christian 2009). In all these studies, the intervention involved a message to respondents indicating why they were being prompted and encouraging them to take certain actions. For example, in the studies of speeding interventions conducted by Conrad and his colleagues, respondents who answered faster than a response time threshold were prompted with a message saying, "You seem to have responded very quickly. Please be sure you have given the question sufficient thought to provide an accurate answer. Do you want to go back and reconsider your answer?" Overall, these studies on interventions have found very promising results: all these studies have shown that the interventions are reasonably successful in affecting the targeted behavior (e.g., less speeding and reduced item nonresponse); in addition, contrary to the concern that respondents might get annoyed by the interventions and drop off the survey, the studies found no substantial increase in break-off as a result of the interventions (Conrad et al. 2009; Conrad et al. 2011; DeRouvray and Couper 2002).

2.2 This Study

Despite these promising findings, there are still many questions yet to be explored regarding intervention design and how it might affect the success of interventions in improving data quality. Specifically, this current study examines the choice of behavior that is targeted for an intervention. Satisficing respondents are likely to engage in more than one satisficing behavior. This can be seen in the correlation of satisficing behaviors across questions. For example, Conrad et al. (2011) found that respondents who sped more often in a series of behavioral frequency questions were also more likely to straightline in later grid questions. In addition, it is also possible to observe multiple

satisficing behaviors within a question. For example, Study 1 of this dissertation showed that within a grid question speeding is associated with straightlining. Given that satisficing behaviors tend to be related both across and within questions, this raises the question of which behavior prompts should be designed to discourage. If we intervene when respondents engage in one behavior will that reduce the frequency of other satisficing behaviors?

Specifically, this study focuses on intervention in grid questions – a question format ubiquitous in Web surveys. Although the grid design is an economical way of presenting multiple items with the same response option, one concern is that the design of the grid may encourage non-differentiated answers (i.e., the selection of similar answers for all items or statements without carefully evaluating the individual items or even reading the items). Therefore, one direct approach is to prompt respondents who give non-differentiated answers in a grid question. Another option is to prompt respondents for speeding, given the evidence that speeding is related to straightlining in grid questions.

This study investigates whether these two prompts (speeding and non-differentiation) have different impacts on respondent behavior. Specifically, this study addresses two research questions:

Research Question (RQ)

RQ1: Do speed and non-differentiation interventions produce different impact on respondent behaviors on grid questions (i.e., instances of speeding and non-differentiation)?

The previous studies have shown that prompts can at least reduce the targeted behavior. Thus, with regard to the impact on grid questions (RQ1), I expected the speeding prompts to reduce speeding, and non-differentiation prompts to reduce the occurrence of non-differentiated responses. Given the association between speeding and non-differentiation, I also expected that intervention with one behavior would curtail both behaviors.

RQ2: For each type of prompt can the effect, if any, be carried over to other types of questions?

Regarding the carry-over effect of prompts on other types of questions (RQ2), I expected speeding prompts to have broader impact on respondent performance than the non-differentiation prompts. The reason for this intuition concerns the generality of the targeted behavior. Respondents can speed on any type of question, while non-differentiation is specific to grid questions. Therefore, speeding prompts in one question could lead respondents to slow down in other questions regardless of the question types. By contrast, non-differentiation prompts are designed to address a behavior specific to grid questions and so do not explicitly provide respondents guidance on how they should perform in other types of questions. Thus, the influence of non-differentiation prompts might be limited to grid questions.

To summarize, this study has two hypotheses:

Hypotheses:

H1: Prompting either speeding or non-differentiation reduces both behaviors on grid questions (RQ1).

H2: Speeding prompts in grid questions can also improve respondent performance in other types of questions, while the influence of the non-differentiation prompts will be evident only for grid questions. (RQ2)

2.3 Experimental Design

The experiment included two phases: an ***intervention*** phase and an ***evaluation*** phase. (To make the question topics cohesive, all the questions in this experiment are related to health.) The intervention phase contained four grid questions, one grid per screen. Each of the grid questions consisted of 4 or 5 statements with a 5-point rating scale. Respondents were randomly assigned to one of the following three conditions for the intervention phase:

(1) **Speeding prompt condition**: respondents prompted for speeding (response time < 300 msec per word)

(2) **Non-differentiation prompt condition**: respondents prompted for non-differentiation, which includes both straightlining (same responses for all the statements in a grid) and "near-straightlining" (same responses for all the statements in a grid, except for one);

(3) **Control condition (no prompt)**

When respondents' behavior triggered a prompt (either because the response was very fast, or the ratings in a grid were very similar), a window popped up with a message about the behavior. The pop-up window also presented respondents an option to go back

and re-consider their answers. If they chose to go back, respondents were returned to the grid question that triggered the intervention. There was no intervention for this second pass through a question. If they chose to go on, respondents would move on the next question. This optional review was designed to be less irritating than an obligatory approach that would have forced respondents to go back. (See Table 2.1 for details of the experimental conditions.)

Following the intervention phase was the evaluation phase, which was designed to assess the impact of the prompts. This section contained three types of questions: one grid question (as a benchmark), two health-related knowledge questions, and one open-ended question. The order among these three types of questions was randomized. There was no intervention for the evaluation section.

Figure 2.1 summarizes the design of this experiment. The wording of all the questions in the experiment is presented in Appendix 2.1.

Table 2. 1. Three Experimental Conditions for the Intervention Phase (Four Grid Questions)

Experimental Conditions	Behavior Triggering Intervention	Intervention Message Displayed in a Pop-up Window
No prompt (control): (1/5 sample)		
Non-differentiation (ND) prompt: targeting both straightlining and near-straightlining (2/5 sample)	When respondents straightline (same responses for all the items in a grid) or near-straightline (same responses except for one item)	"You seem to have given very similar ratings for the different items in this question. Please think about each item on its own and be sure to give it enough thought so that your answer is informative and accurate. Do you want to go back and reconsider your answers?" (Yes/No)
Speeding (SP) prompt: targeting speeding (2/5 sample)	When the total response time to the grid question is less than 300 msec per word (i.e., 0.3 sec multiplied by the number of words in the question)	"You seem to have answered very quickly. Please be sure you have given all the items in the question sufficient thought so that your answer is informative and accurate. Do you want to go back and reconsider your answers?" (Yes/No)

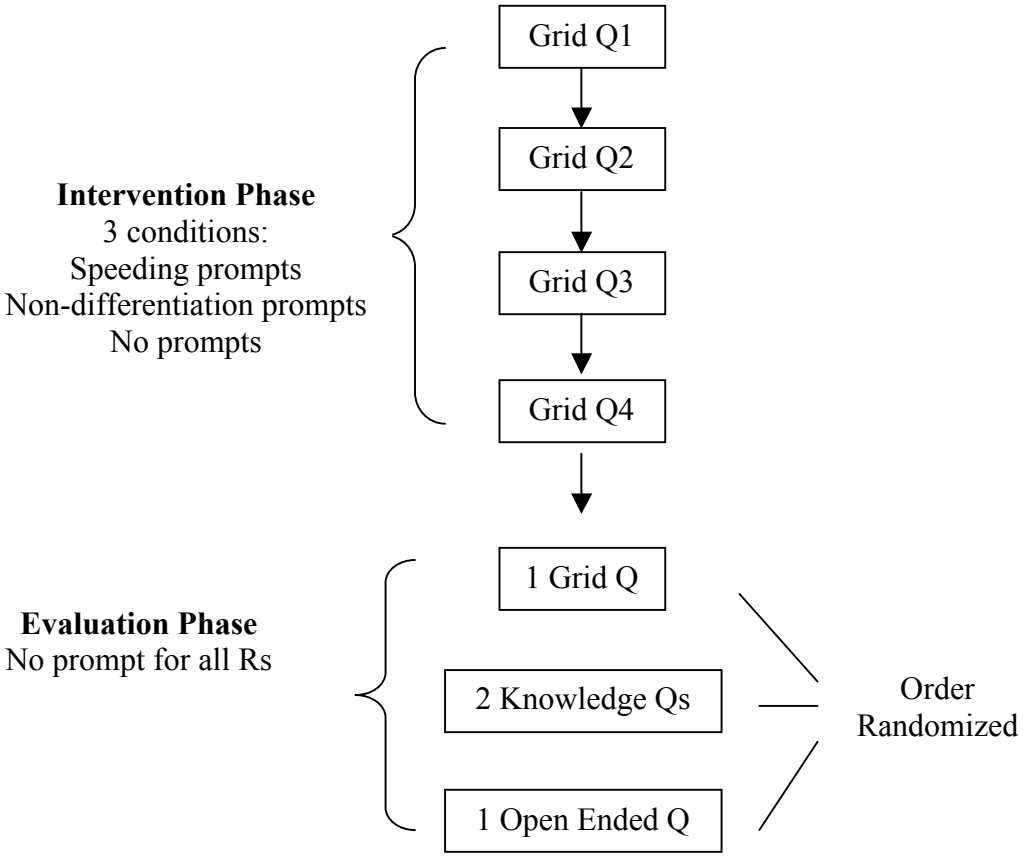


Figure 2. 1. Experimental Design

Given the question included in the evaluation phase, the hypothesis regarding the carry-over effect of the prompts (H2) allows me to derive the following three predictions: (1) both speeding and non-differentiation prompts on the earlier intervention phase can reduce speeding and non-differentiation on the grid question in the evaluation phase; (2) in addition, speeding prompts on the earlier grid questions can improve respondent behaviors on both the knowledge (less speeding and more correct answers) and open-ended questions (less speeding and longer answers); and (3) non-differentiation prompts

on the earlier grid questions have little impact on respondent performance in the knowledge and open-ended questions.

The experiment was attached to the end of a monthly survey on the LISS panel (Longitudinal Internet Studies for the Social Sciences). The survey is about work and schooling and it is one of the core surveys conducted annually on the panel. The data were collected in April 2012. The completion rate was 78.3% (5,848 completes out of 7,472 invited panel members).

2.4 Results

Manipulation Check & Break-offs

As can be seen in Table 2.2 below, break-offs are very rare in all three conditions. Thus, consistent with previous studies, this study does not find an increase in break-offs as a result of the intervention. Table 2.2 also shows that majority of the respondents (over 80%) either near-straightlined or straightlined one or more times, and thus were eligible for at least one non-differentiation prompt, while only about one third of respondents were eligible for any speeding intervention.

It is worth noting that the primary manipulation in this experiment only affects respondents who warrant an intervention. Respondents who never sped during the intervention phase received the same treatment in both the control (no prompts) and speeding prompt condition. This suggests that the percentage of respondents who never sped – or equivalently the respondents who sped at least once (i.e., were eligible for any speeding prompts) – should not differ between the control and speeding prompt condition. Similarly, there should be no significant difference in the percentage of respondents eligible for any non-differentiation prompt between the control and non-

differentiation prompt condition[6]. Both of these predictions are confirmed (see the *t*-tests in Table 2.2).

Table 2. 2. Intervention Eligibility and Break-offs

	Control	Speeding prompts	Non-diff prompts	Diff (*t*-tests)
Respondents starting the experiment	1138	2329	2389	
Break-off	0.09% (n=1)	0.17% (n=4)	0.17% (n=4)	
Eligibility for any SP prompt	35.0%	37.3%	--	*t*= -1.31, *p*=0.191
Eligibility for any ND prompt	83.6%	--	83.0%	*t*= 0.40 , *p*=0.692

Intervention Phase (Effect of Intervention on Grid Questions)

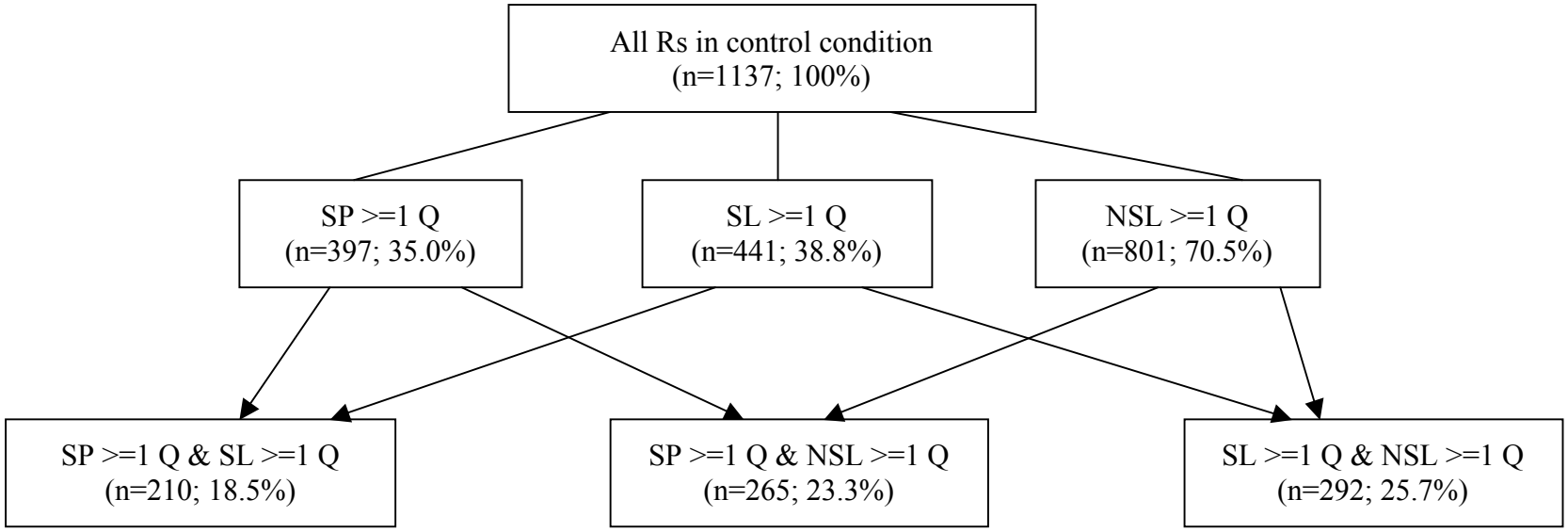
Before I describe the results, I want to point out that the four grid questions in the intervention section are purposely designed so that valid responses are likely to involve some differentiation across the items in the grid. This is either because the wording of items implies opposite valence or because the items have somewhat distinctive traits. For example, one grid question contained five diet-related behaviors. Four of them are typically considered healthy choices (e.g., avoiding fast food). The other one is not as healthy (i.e., emphasizing the taste of food rather than its nutritional value) and respondents are expected to rate this item differently from others. Another example is the question about obtaining health-related information; respondents are expected to rate the information from doctors as more important compared to that from other sources (see Appendix 2.1 for all the question wording). Thus, straightlining in this study is

[6] However, it is worth noting that among the respondents eligible for any prompts, the number of instances of the behavior (speeding or non-differentiation) should be lower in the prompt than in the control condition (as shown later in Table 3 and Table 4).

considered clear evidence of satisficing. Near-straightlining in this study is considered a milder satisficing behavior than straightlining.

I started by examining how respondents answered the four grid questions when there was no prompt (control condition). An overview of respondent performance in this condition is presented in Figure 2.2 with focus on three behaviors: speeding, straightlining, and near-straightlining.

There are a few things worth noting in Figure 2.2. One is the high prevalence of near-straightlining: approximately 70% of the respondents near-straightlined on at least one question. In contrast, straightlining and speeding were less endemic: about 40% of all the respondents straightlined at least once and 35% sped at least once. In addition, these two groups are composed of overlapping, but not identical, respondents. Specifically, about half of the respondents in each of the two groups (n=210) also belonged to the other group.



Notes:
SP >=1 Q: sped on at least one question
NSL >=1 Q: near-straightlined on at least one question
SL >=1 Q: straightlined on at least one question
All the percentages are with regard to the total number of respondents in the control condition (i.e., denominator is 1137).

Figure 2. 2. Overview of respondent behaviors on the four grid questions (control condition only)

To evaluate the impact of prompts, I first examined the overall occurrence of speeding, straightlining and near-straightlining in the three conditions. This analysis focused on respondent behaviors when they answered the questions for the first time. In the two prompting conditions, prompted respondents could choose to go back and answer the question again. Because the prompts could only occur after respondents answered the first question, the comparisons excluded the first question and examined respondent behaviors on the second, third and fourth question.

The findings are presented in Table 2.3. The left side of the table presents the means and standard deviations of the incidences of each satisficing behavior. The right side of the table shows pairwise comparisons of these behaviors across the three experimental conditions. Compared to the control condition, all three behaviors were significantly less frequent when non-differentiation prompts were given, with a particularly clear reduction in straightlining (0.25 vs. 0.45, t=8.89, p<.001). By contrast, the overall impact of the speeding prompts seemed much smaller and not significant.

It is worth pointing out that the comparisons in Table 2.3 include all respondents in each condition. This approach of including all participants in the comparison across randomized treatment groups (regardless whether the subjects satisfy the entry criteria or they comply with the treatment) is referred to as intention to treat (ITT) analysis in clinical trial studies. One important advantage of ITT analysis is that it produces unbiased pragmatic estimates of the treatment effect. However, ITT analysis tends to underestimate the treatment effect if some participants in the treatment group do not receive the treatment because they are not eligible. In this study, both prompt conditions included this type of ineligible and not treated participants (i.e., respondents who never

sped in the speeding prompt condition and those who never straightlined nor near-straightlined in the non-differentiation prompt condition). In particular, the effect of speeding prompts could be particularly understated by the ITT analysis, since only about one third respondents in the speeding prompt condition were eligible and given a prompt (see Table 2.2)

Table 2. 3. Effect on Grid Questions: Comparison across three experimental conditions with all respondents included (ITT analysis)

	Experimental Conditions			*p*-value of pairwise *t*-tests (with Bonferroni correction)		
	(1) Control	(2) Speeding prompts	(3) Non-diff prompts	(1) vs. (2)	(1) vs. (3)	(2) vs. (3)
sample size	1137	2325	2385			
mean # of speeding (std)	0.48 (0.76)	0.47 (0.71)	0.39 (0.68)	1.000	0.002	<.001
mean # of straightlining (std)	0.45 (0.75)	0.42 (0.72)	0.25 (0.57)	0.880	<.001	<.001
mean # of near-straightlining (std)	0.67 (0.73)	0.65 (0.72)	0.55 (0.66)	1.000	<.001	<.001

Notes: The analyses compared respondent behaviors on the second, third, and fourth grid questions.

To better capture the effect of the prompts, I compared each of the prompt conditions to the control condition and restricted the analyses to respondents who were eligible for prompts[7]. The comparisons were separate because prompt eligibility differed between the speeding and the non-differentiation prompts. Specifically, to compare the effect of speeding prompts, I removed respondents who never sped in the control and the speeding prompt condition; similarly, I excluded respondents who never straightlined nor near-straightlined in the control and the non-differentiation prompt condition to evaluate the effect of non-differentiation prompt. The findings are shown in Table 2.4. For the non-differentiation prompts, this "eligible-only" approach yields similar conclusions to those in Table 2.3. For speeding prompts, the "eligible-only" comparisons reveal that speeding prompts reduced speeding and straightlining, which were "disguised" in the overall analyses.[8]

In summary, the analyses of respondent behaviors on the grid questions (intervention phase) showed that prompting either speeding or non-differentiation can reduce both behaviors. In addition, non-differentiation prompts seem to have a much more noticeable overall impact on respondent performance because more respondents engage in non-differentiation than speeding.

[7] This method of excluding ineligible respondents should yield equivalent samples. This is confirmed in Table 2.2.

[8] The reduction in straightlining is only significant by the measure of number of straightlining incidence. The percentage of any straightlining appears to be also reduced, but the difference is not significant.

Table 2. 4. Effect on Grid Questions: Comparison of each prompt condition to the control condition including only respondents eligible for at least a prompt

Satisficing Behaviors	**Control condition** (Rs eligible for one or more speeding prompts)	**Speeding prompt condition**	*t*-tests	**Control condition** (Rs eligible for one or more non-diff prompts)	**Non-diff prompt condition**	*t*-tests
Mean # of Speeding	1.36	1.26	t=2.71, p=0.007	0.50	0.40	t=3.56, p<.001
% Any Speeding	98.0%	96.8%	t=1.21, p=0.225	35.5%	30.2%	t=2.87, p=0.004
Mean # of Near-straightlining	0.61	0.65	t=-0.81, p=0.419	0.80	0.66	t=5.27, p<.001
% Any Near-straightlining	47.1%	49.8%	t=-0.88, p=0.380	62.6%	55.3%	t=3.77, p<.001
Mean # of Straightlining	0.77	0.66	t=2.12, p=0.034	0.54	0.30	t=9.03, p<.001
% Any Straightlining	48.1%	43.9%	t=1.41, p=0.160	38.0%	23.4%	t=8.30, p<.001

Notes: The analyses compared respondent behaviors on the second, third, and fourth grid questions.

Implications of behavior changes as a result of the prompts

Despite the success of the prompts, one concern is that the change of behaviors may not reflect more conscientious responding – respondents might change their behaviors superficially simply to avoid triggering prompts. However, the finding that prompting either speeding or non-differentiation can reduce both behaviors on grid questions seemed to suggest that it was not the case. If respondent slow-down reflected an adaptive strategy (e.g., they still respond without much thought, but wait a few more seconds before submitting their answers), we would not observe the reduction in non-differentiated answers by the speeding prompts. Similarly, if respondents randomly clicked on the response options to make their answers appear to be more differentiated, non-differentiation prompts would have little impact on speeding.

To further assess whether respondents took the prompts seriously, I examined the content of responses in the grid questions. Specifically, I looked at the grid question, which asked respondents to rate the importance of various sources for health-related information. As can be seen from Table 2.5, respondents in the two prompt conditions rate the information from "health service professionals" as more important than those in the control group. Although it is not possible to verify whether a higher rating of "health service professionals" means more accurate responses, it certainly suggests that the different reports as a result of the intervention are at least thoughtful.

Table 2. 5. Comparisons of Ratings to One Grid Question in the Intervention Section

Mean (1=not important at all; 5= very important)	**Control condition** (Rs eligible for one or more SP prompts)	**Speeding prompt condition**	*t*-tests	**Control condition** (Rs eligible for one or more ND prompts)	**Non-diff prompt condition**	*t*-tests
Television	3.07	2.96	*t*=1.62, n.s.	3.09	3.08	*t*=0.33, n.s.
Books, newspaper, and magazines	3.27	3.22	*t*=0.92, n.s.	3.34	3.34	*t*= -0.17, n.s.
Internet	3.43	3.36	*t*=1.11, n.s.	3.28	3.31	*t*= -0.63, n.s.
Family and friends	3.50	3.39	*t*=1.89, *p*=0.059	3.44	3.45	*t*= -0.10, n.s.
Health service professionals	**3.57**	**3.68**	***t*= -1.78, *p*=0.076**	**3.83**	**3.93**	***t*= -2.67, *p*=0.008**

Evaluation Phase (Effect of intervention on other types of questions)

The results above have shown that both speeding and non-differentiation prompts improved respondent performance on the grid questions. Next, I report the findings on whether this positive effect is transferred to other types of questions. Specifically, I examined respondent behaviors during the evaluation phase of the experiment, which contained three types of questions administered right after the intervention phase.

I first conducted ITT analysis of all the respondents including those not eligible for a prompt. (Note that because the break-offs in this experiment were very rare, the sample in the evaluation phrase should still be equivalent across experimental conditions.) The findings are presented in Table 2.6. For the grid question in the evaluation phase, the overall speeding and straightlining seemed to be curtailed by the earlier non-differentiation prompts; the earlier speeding prompts seemed to only affect overall speeding, with little impact on non-differentiation. There was no evidence that intervention affected how respondents answered the knowledge questions: both the response times and the correctness of their answers were similar across the experimental conditions. For the open-ended question, despite the limited impact on overall response time, both prompts seemed to lead to longer answers. (As shown in Table 2.6, the increase was significant for the non-differentiation prompt and marginal significant for the speeding prompt.)

To highlight the effect of prompts, Table 2.7 shows separate comparisons for each type of prompt including only the respondents eligible for a prompt in the intervention phase. For the grid question, as expected, interventions – either for speeding or non-differentiation – reduced both speeding and straightlining on this question, much the

same as the impact on the grid questions in the intervention phase. For the open-ended question, both types of prompts led to significantly and noticeably longer response times and longer responses (more characters). This contradicts the original hypothesis that only the effect of speeding prompts can transfer to other types of questions. Despite the significant effect on the open-ended question, both prompts failed to increase correct answers to the knowledge questions. My explanation is that the response process to these knowledge questions may be different than the process for closed-ended survey questions. For the knowledge questions, respondents are likely to be in a situation of either knowing the answer or not, which may have less to do with effort than knowledge, and therefore were not much influenced by the prompts.

Table 2. 6. Respondents' Behaviors in the Evaluation Phase: Comparison across three experimental conditions with all respondents included (ITT analysis)

	Experimental Conditions			*p*-value of pairwise *t*-tests (with Bonferroni correction)		
	(1) Control	(2) Speeding prompts	(3) Non-diff prompts	(1) vs. (2)	(1) vs. (3)	(2) vs. (3)
Grid Q						
% Speeding	22.7%	17.5%	18.6%	<.001	0.013	0.939
% Straightlining	7.7%	6.1%	3.8%	0.142	<.001	0.002
Knowledge Qs						
Total response time	40.12	40.18	40.27	1.000	1.000	1.000
Total correct Qs	3.95	3.99	3.94	0.484	1.000	0.099
Open-ended Q						
Response time	40.34	41.39	42.84	1.000	0.155	0.492
Answer length (characters)	26.8	29.41	29.99	0.092	0.024	1.000

Table 2. 7. Respondents' Behaviors in the Evaluation Phase: Comparison of each prompt condition to the control condition including only respondents eligible for at least a prompt

	Control condition (Rs eligible for SP prompts in the intervention section)	**Speeding prompt condition**	*t*-tests	**Control condition** (Rs eligible for ND prompts in the intervention section)	**Non-dif prompt condition**	*t*-tests
Grid Q						
% Speeding	49.4%	36.5%	t=4.37, p<.001	24.0%	18.7%	t=3.31, p=0.001
% Straightlining	14.1%	10.2%	t=2.03, p=0.042	9.0%	4.0%	t=5.42, p<.001
Knowledge Qs						
Total response time	27.71	30.72	t=-2.18, p=0.029	39.48	39.94	t=-0.37, p=0.709
Total correct Qs	3.94	3.94	t=-0.11, p=0.913	3.94	3.94	t=0.07, p=0.941
Open-ended Q						
Response time	25.93	29.20	t=-2.12, p=0.034	38.72	42.57	t=-2.73, p=0.007
AAwer length (characters)	22.55	26.14	t=-2.13, p=0.034	25.36	29.46	t=-3.05, p=0.002

2.5 Discussion and Conclusions

This study compared two types of interventions on grid questions – one for speeding and the other for non-differentiation – with the goal of exploring which intervention is more effective in enhancing response quality. The results show that prompting after either speeding or non-differentiation can reduce both behaviors on grid questions, not just the specific behavior targeted in the prompt. This study also demonstrates that both prompts, although implemented only on grid questions, have effects beyond grid questions. Specifically, this study shows that the earlier prompts on grid questions substantially improved the quality of responses to a later open-ended question and this is true for both speeding and non-differentiation prompts. Overall, the findings suggest that the prompts, although only occurring on grid questions and targeting a particular behavior, can improve overall response quality. There are a few possible explanations. One is that the prompts simply serve as reminders to improve answer quality, and that, once reminded, at least some respondents will work harder. It is also possible that prompts inform respondents that the survey organization is monitoring their effort. As a result, respondents may feel more accountable for their behaviors, motivating them to expend more effort.

This study shows that speeding prompts and non-differentiation prompts work in quite similar ways –both seem to draw respondents' attention to response quality. One difference is that the two prompts target an overlapping but not identical set of respondents. In this study, non-differentiation prompts are administered to many more respondents than speeding prompts. Therefore, although both prompts have a positive

impact on the quality of answers to grid questions, non-differentiation prompts seem to have more noticeable effects.

Intervention for satisficing has to rely on the behavioral indicators such as speeding, item non-response, and non-differentiation. Although these behaviors are likely to be the response strategies employed by satisficing respondents, they do not always result from satisficing (as noted in Krosnick 1991). Therefore, the intervention technique may miss some satisficing respondents (i.e., false negative) and mistakenly intervene with some conscientious respondents (i.e., false positive). In this study, the speeding intervention used a general threshold of 300 msec per words. Although this measure of speeding can be very useful to capture the overall tendency to speed across various questions, it is less informative when used to pinpoint satisficing respondents on individual questions. The non-differentiation intervention in this study prompted respondents for both straightlining and near-straightlining. In this study straightlining constituted strong evidence of satisficing, because the grid questions were designed so that valid responses in a grid were likely to vary. The near-straightlining in this study seemed to be less problematic, especially for the grid with only four items[9]. Some of the near-straightlining answers could be legitimate. This probably explains why the intervention in this study (both for speeding and non-differentiation) had larger impact on straightlining compared to that on near-straightlining. To maximize the potential of intervention, research is needed to explore how to better identify satisficing respondents.

[9] This can be seen from response times. For the grids of five items, there is a steady increase of response time from straightlining respondents, near-straightlining respondents, and other respondents, For a grid question of only four items, response times are still the shortest among the straightlining respondents, but the response times of near-straightlining respondents seems to be similar to those who gave more differentiated answers, i.e., these near-straightlining respondents might expend as much effort as the respondents with more differentiated responses.

To conclude, this study demonstrates the promise of intervention to positively influence behaviors of Web respondents and enhance answer quality. Although this study focuses on satisficing, the findings seem to suggest that in general, better communication and more interaction with Web respondents (e.g., tailored feedback about particular behaviors and instructions of what is expected of respondents) may produce better respondent performance. Compared to traditional Web surveys that are essentially online versions of paper questionnaires, an interactive Web survey interface designed to actively guide the interviewing process (similar to the role of a human interviewer) may help keep respondents informed and engaged, and even indicate the seriousness of the survey, thus yielding more conscientious respondent behavior.

Appendix 2. 1. Question Wording

I. Intervention Phase

INT_Q1 [GRID DESIGN]: Scientists and doctors now recognize that lifestyle can have a major impact on people's health. Please indicate how important you believe the following lifestyle choices are for your overall health.

1 not important at all
2
3
4
5 very important

A balanced diet
Regular exercise
Adequate sleep
Effective stress management
Prayer, faith, and meditation

INT_Q2 [GRID DESIGN]: Please indicate how much you favor or oppose each of the following behaviors.

1 Strongly oppose
2 Somewhat oppose
3 Neither oppose nor favor
4 Somewhat favor
5 Strongly favor

Avoiding fast food
Paying close attention to the nutritional information on food packaging
Monitoring cholesterol levels closely
Limiting the amount of red meat in diet
Emphasizing the taste of food rather than its nutritional value

INT_Q3 [GRID DESIGN]: How important is each of the following sources for you to obtain health-related information?

1 not important at all
2
3
4
5 very important

Television
Books, newspapers, and magazines
Internet
Family and friends
Health service professionals

INT_Q4 [GRID DESIGN]**:** Modern technology, such as computers, televisions and the recent invention of smartphones, has played an important role in people's lives and their health. Please indicate how much you agree or disagree with the following statements.

1 Strongly disagree
2 Somewhat disagree
3 Neither disagree nor agree
4 Somewhat agree
5 Strongly agree

Modern technology makes our lives healthier, easier, and more comfortable.
Modern technology causes us to become more and more inactive in our daily lives.
People would do better by living a simpler life without so much technology.
The next generation will live better because of advances in technology.

II. Evaluation Phase

EVA_Q1 [GRID DESIGN]: We are interested in what you think of the following statements related to health.

1 Strongly disagree
2 Somewhat disagree
3 Neither disagree nor agree
4 Somewhat agree
5 Strongly agree

Eating healthy means sacrificing taste.
I wish I could make better choices with the things I eat and drink.
It is possible for a diet to be both healthy and satisfying.
I am satisfied with the way I manage my diet.

EVA_Q2: Eating foods that contain saturated fats raises the level of cholesterol in your blood. Based on your knowledge, are saturated fats usually found in

Animal products like meat and dairy products
Vegetables and vegetable oils
Not sure

EVA_Q3: Based on your knowledge, which has more saturated fat in the following series of paired foods?

Liver, or
T-bone steak

Butter, or
Margarine

Egg white, or
Egg yolk

Skim milk or
Whole milk

EVA_Q4: Based on your knowledge, what things that people eat might increase the chance of getting heart disease?

[TEXT BOX]

References

Conrad, F. G., Tourangeau, R., Couper, M. P., & Kennedy, C. (2009). Interactive Interventions in Web Surveys Can Make Respondents More Conscientious. Paper presented at the annual meeting of the American Association for Public Opinion Research, Hollywood, FL, May.

Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2011). Interactive Interventions in Web Surveys Can Increase Response Accuracy. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, AZ, May.

Couper, M. P. (2000). Web surveys - A review of issues and approaches. *Public Opinion Quarterly, 64*(4), 464.

Couper, M. P., & Miller, P. V. (2008). Web survey methods: Introduction. *Public Opinion Quarterly, 72*(5), 831-835.

DeRouvray, C., & Couper, M. P. (2002). Designing a strategy for reducing "no opinion" responses in web-based surveys. *Social Science Computer Review, 20*(1), 3-9.

Holland, J. L., & Christian, L. M. (2009). The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Social Science Computer Review, 27*(2), 196-212.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213-236.

Miller, P. V., & Cannell, C. F. (1982). A study of experimental techniques for telephone interviewing. *Public Opinion Quarterly, 46*(2), 250-269.

CHAPTER 3: Design of Interactive Interventions in Web Surveys: Humanness, Social Presence, and Data Quality (Study 3)

3.1 Background

The findings from the existing literature and the Chapter 2 of this dissertation suggest that intervention prompts in Web surveys can be a useful tool for resolving respondent satisficing and enhancing answer quality. This is manifested not only in the reduction of the behaviors targeted by interventions, but also in the evidence that interventions seem to have the potential to improve respondent performance overall.

Despite promising results, the studies on interventions have also revealed the challenge in this approach – how to get respondents to cooperate. For example, Holland and Christian (2009) found that probes in open-ended questions elicited more responses from some of the respondents (25% respondents responded to the probes for the first open-ended question, and only 9% did so for the second open-ended question). Conrad and his colleagues also found that while there was clear evidence of slow-down by the speeding interventions, some respondents (hard-core speeders) seem to be unaffected by prompts and very few respondents went back and re-considered their answers after the prompts (Conrad et al. 2009 & 2011). These findings suggest that while respondents are generally responsive to the interventions from human interviewers, they can be much less

cooperative with computer interventions. To exploit the potential of interventions as a means of enhancing online response quality, research is needed to improve understanding factors that affect intervention effectiveness.

Social Presence in Human-Computer Interaction (HCI)

Intervention prompts in Web surveys are, essentially, the automated equivalent of a task traditionally carried out by human interviewers: monitoring respondent behaviors and intervening if necessary to ensure quality of answers. In this sense, intervention prompts serve the role of a computer agent to interact with respondents.

One important research question in computer agent design is whether making a computer agent look, sound, or behave like a human influences users' interactions with the computer agent. A critical concept for this line of research is social presence – the sense of being with another human. When social presence is triggered, users interact with a computer agent as if the agent were a human.

So far, some research has demonstrated social presence in Human-Computer Interaction (HCI). Specifically, studies have shown that visual and aural cues of humanness can trigger social presence. Sproull et al. (1996) found that users felt more aroused and tended to present themselves in a more positive light when the computer interface had an autonomous talking face than when it had a text-only interface. Evidence of social presence is also found in computer-assisted self-administered surveys containing pre-recorded videos of interviewers. Fuchs (2009) manipulated the gender of the interviewer in the videos in a survey on sex-related topics and found gender-of-interviewer effects similar to those in a face-to-face interview. Krysan and Couper (2003) examined race-of-interviewer effects in a video survey of racial attitudes but only found

evidence for social presence among African American respondents.[10] Evidence of social presence is also found by Conrad, Schober and Nielsen (2011) which manipulated the race of the animated virtual interviewers. However, not all attempts to induce social presence are successful. For example, Tourangeau, Couper and Steiger (2003) found that presenting an image of the researcher in the Web survey had little impact on answers to both the sensitive questions and the scales measuring social desirability and impression management.

In addition to visual and aural cues of humanness, interactivity is another important factor to trigger social presence. One well-known finding is from Nass et al. (1999). The study found that people, after a series of interactions with a text-based computer for a tutoring program, rated the tutoring performance higher when the evaluation questions were asked on the same computer compared to when the evaluation was administered either on a different computer or in a paper-and-pencil.

Some studies reported evidence of social presence based on self-reports to a series of items that asked people how much their interaction with the computer felt like they were with an actual person (e.g., Lee and Nass 2005; Skalski and Tamborini 2007). Such self-reported measures of social presence contradict the view that social presence is automatic and unconscious (Nass et al. 1999). The differences in these self-reported measures do not necessarily imply behavioral differences. In fact, Tourangeau, Coupter and Steiger (2003) found that the manipulation of the humanized cues in the survey

[10] For the black respondents, the race-of-interviewer effect was similar to that in the condition of live interviewers, indicating the videos of the interviewers induced social presence among the black respondents. However, for the white respondents, the study found that the video and live interviewer condition had the opposite direction with regard to race-of-interviewer effect, implying that the image of the black interviewer activated stereotypic responses to the racial issues, rather than inducing the sense that a black interviewer was present.

interface affected respondents' evaluation of their survey experiences, but had little impact on how respondents answered the survey.

Social Presence in Computer-Mediated Communication (CMC)

Social presence discussed above is essentially the illusion of a computer being a human. This is different from the social presence created when computers are used to interact with other people (i.e., computer-mediated communication, such as video chat). The degree of social presence for CMC depends on how much the medium is transparent and invisible. (See Lombard 1997 for detailed discussion on definitions for social presence.)

Other Evidence of Social Responses to Computers

Two other types of evidence are sometimes referred to as social responses to computers. One type of evidence demonstrates that people can treat computers as social actors. For example, Nass et al. (1995) and Moon and Nass (1996) found that users attributed different personalities (dominant vs. submissive) to text-based computers depending on how the computer gave feedback. These studies also showed that people responded to computer personalities the same way as they would react to human personalities. Nass, Fogg, and Moon (1996) showed that subjects readily formed a team with a computer simply because they were told that the evaluation of their performance depended on another computer. Although these studies showed similarity between how people interact with computers and other humans, the findings do not necessarily suggest that people mistake computers for people.

Another type of evidence for social responses to computers comes from the studies showing that humanized features of a computer agent can evoke stereotypes from the

users. For example, Nass, Moon, and Green (1997) found that the gender of the voice of the computer triggered various gender-related responses (e.g., the tutoring computer was perceived to be more informative when the gender of voice was stereotypically matched to the topic, such as men know more about computers than women). Pratt et al. (2007) examined how the ethnicity of computer agents affects people's behaviors in item-ranking tasks. They found that the subjects (mostly white) tended to change their attitudes more based on a white agent's advice than an African-American agent's advice and rated the white agent higher on intellectual-related personality attributes. These findings can be interpreted as evidence that people use human features of a computer agent to infer the source of the messages presented by the computer, but do not necessarily indicate that the users feel that a human agent is present.

3.2 This Study

This study explores whether interactive interventions in Web surveys can induce social presence – that is, the illusion of interacting with a human agent. This illusion might yield more respondent compliance with Web intervention and be more effective in motivating respondents and improving response quality than when respondents treat interventions simply as automated computer feedback. In the latter case, they may more readily ignore the interventions and thus limit the positive effects of interventions.

However, the feeling of interacting with a human agent, if triggered, might also lead to less candid reports to sensitive questions. If so, this could begin to help us understand what it is about removing interviewers from question-asking that produces the advantages of self-administration. In particular, it could be that it's the presence of discrete cues of humanness more than a physically present human that promote socially

desirable responding. So to the extent these cues are displayed in what are traditionally considered self-administered modes, there is socially desirable responding.

While the studies of HCI have been mostly conducted in laboratories, this study examines interventions in real survey settings. This is because participants in labs tend to be more focused on and cooperative with the tasks compared to when they are in natural environments. Therefore, people may be less likely to engage satisficing behavior like speeding, making it harder to measure the impact of interactive prompts on such behaviors.

The empirical findings about social presence (reviewed in the previous section) suggest that "talking faces" (Sproull et al. 1996; Fuchs 2009; Krysan and Couper 2003; Conrad, Schober, and Nielsen 2011) and the interactivity of the interface could be effective cues to trigger social presence. Despite strong human cues in "talking faces", there are some practical challenges of using "talking faces" in real survey settings. First, the sudden appearance of a "talking face" in the middle of a survey could feel disruptive or strange to respondents. Second, respondents can control the volume of the audio output or even whether it is on or off. As a result, we cannot guarantee that respondents will hear the speech output. Therefore, this study employs only visual cues – images and texts – to manipulate respondents' perception of whether they are interacting with a computer or a human agent. In addition, to manipulate the perceived interactivity of the prompts, this study also varies how the intervention messages are presented. The details of the experimental design are described in the next section.

Experimental Design (3 phases)

Phase 1: Intervention (5 conditions)

The experiment consisted of three phases. The first was the intervention phase. This phase included five behavioral frequency questions (e.g., "During the past 2 years, how many overnight trips have you taken?"). Respondents were randomly assigned to one of five conditions: four intervention conditions (where fast responses are prompted) and a control condition with no intervention.

The four intervention conditions formed a 2 × 2 factorial design. The first factor controlled the picture and the text displayed in the interventions and varied whether it was human-like or computer-like. In the human-like condition, a picture of a female face was presented with the intervention message (i.e., "You seem to have answered very quickly. Please be sure you have given all the items in the question sufficient thought so that your answer is informative and accurate. Do you want to go back and reconsider your answers? Yes/No"). In the computer-like condition, a yellow triangle error sign was displayed instead. The text was the same as that in the human-like condition, except that a statement, "The page at https://[insert the current http address] says:" is added to the beginning of the text.

The second factor varied the perceived interactivity of the intervention. In one condition, the intervention (image and text) appeared on the next screen and looked like the next question in the questionnaire. In the other condition, the intervention was displayed in a pop-up window on the top of the current survey question for which the intervention was triggered; this was intended to increase the salience of the interactive prompt.

In summary, the 2 x 2 factorial design produced four intervention conditions: (1) human-like intervention in a pop-up window (HM-POP), (2) human-like intervention in the next survey screen (HM-SCN), (3) computer-like intervention in a pop-up window (CMP-POP), and (4) computer-like intervention in the next survey screen (CMP-SCN). Appendix 3.1 presents the screen captures of the interventions in each of the four conditions. The experiment also included a control condition with no intervention. In total, there were five conditions.

In the human-like condition, this study used images of three different human faces (displayed in Appendix 3.2). Respondents in this condition were randomly assigned to one of the images, which were kept the same across all the interventions throughout the survey. The images were obtained from the Microsoft gallery, an online database providing royalty-free images.

The faces in the three photos are all white, female, smiling and wearing a telephone headset. Those images were intended to approximate interacting with a human interviewer in a telephone survey. This experiment explores whether images of faces can influence respondents' behavior; it is not intended to explore whether the impact of faces depends on their characteristics (e.g., female vs. male, blond versus brunette). Nonetheless, we will test for differences between the three faces.

Phase 2: Intervention Evaluation

Following the intervention phase was a set of questions evaluating the impact of interventions. This included two grid questions and three sensitive questions. (No speeding intervention is implemented for these questions.) Respondent behaviors on the

two grid questions were evaluated as additional evidence of the effect of the prompts on answer quality. The sensitive questions were designed to assess whether the earlier interventions affect the report of socially undesirable behaviors in these questions.

Phase 3: Design Assessment

At the end of the survey, respondents who received any intervention during phase 1 were asked to evaluate their experience with the interventions. Two sets of questions were included. The first set of questions measured the perceived humanness and interactivity of the interventions. The second set of questions collected respondents' opinions about the female faces (e.g., attractiveness); they were only administered to the respondents who received interventions in the two human-like conditions.

An overview of the design of this experiment is presented in Figure 3.1. The wording of all the questions in this experiment is presented in Appendix 3.3.

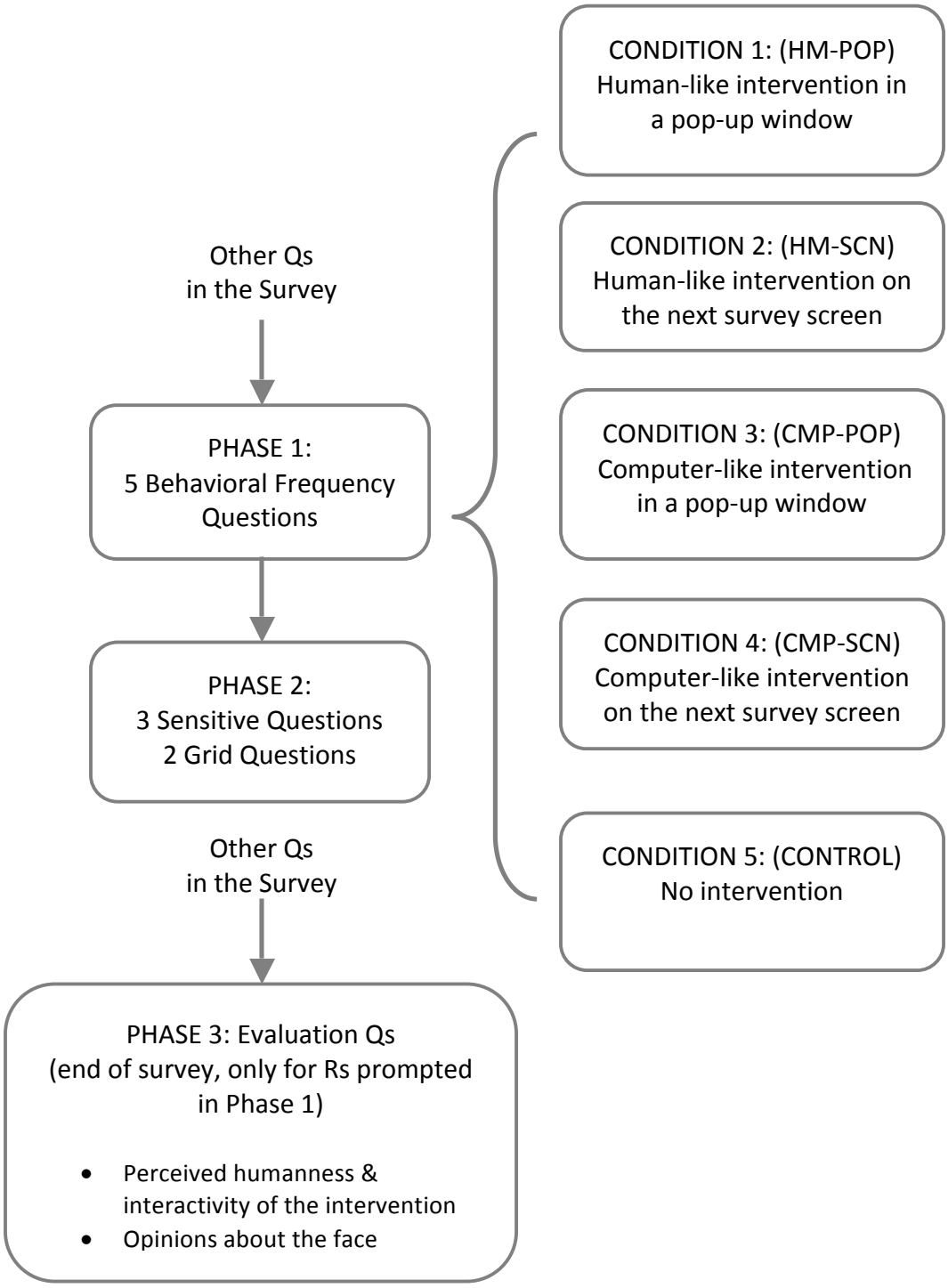


Figure 3. 1. Experimental Design

Predictions

1. **Difference between control and the four intervention conditions:** Four intervention conditions will increase the level of respondent effort compared to the control condition. In this study, respondent effort will be measured using a variety of indictors, including instances of speeding, willingness to go back and re-consider the answers after having been prompted, and satisficing behaviors on the behavioral frequency questions (speeding and rounding) and the grid questions (speeding and straightlining).

2. **Differences among the four intervention conditions:** Both humanized cues and perceived interactivity of the intervention will have to be present to trigger the illusion of interacting with a human agent. Therefore, HM-POP condition (with the presence of both stimuli) is expected to be more likely to trigger a greater sense of social presence than the other three conditions (i.e., HM-SCN, CMP-POP, and CMP-SCN). If this is the case, HM-POP condition is likely to yield more respondent effort (same measures as above) and more socially desirable answers to the sensitive questions compared to the other three intervention conditions.

These predictions are summarized in Table 3.1 below.

Table 3. 1. Predicted Differences Across Experimental Conditions

Respondent behaviors	Measures	Expected difference across conditions
Respondent effort	1. Speeding instances (behavioral frequency question, sensitive question, and grid questions) 2. Go-back after the intervention prompt, 3. Rounding (behavioral frequency questions) 4. Straightlining (grid questions)	HM-POP > HM-SCN CMP-POP CMP-SCN > CONTROL
Socially desirable Answers	3 sensitive questions: marijuana use, exercise, binge drinking	Same as above

Note:
HM-POP: Human-like intervention in a pop-up window
HM-SCN: Human-like intervention on the next survey screen
CMP-POP: Computer-like intervention in a pop-up window
CMP-SCN: Computer-like intervention on the next survey screen

Data Collection

This experiment was embedded in a Web survey of opt-in panel members. The survey was conducted as part of an NIH funded study on Web survey methodology.[11] The sample was drawn from two sources, half from the panel of Authentic Response and the other half the panel of Survey Sampling International (SSI). The survey was conducted by Market Strategies International and was fielded from December 13, 2011 through December 23, 2011. A total of 3,274 respondents started the survey and among them 2,427 (74.1%) completed the questionnaire.

3.3 Results

Break-offs

I first examined the break-offs in the intervention phase (i.e., five behavioral frequency questions). As shown in Table 3.2, overall the break-offs were very rare in all the experimental conditions ranging between 0 and only 6 respondents. The HM-POP condition had slightly more break-offs (6 respondents) than other conditions (0 or 1 respondent), but the break-off rate was still very low (about 1%).

[11] The NIH grant (#R01HD041386-04A1) was awarded to Roger Tourangeau, Mick P. Couper, Frederick G. Conrad, and Reg Baker.

Table 3. 2. Break-offs in Each Experimental Condition

	Control	Human-like, pop-up window (HM-POP)	Human-like, survey screen (HM-SCN)	Computer-like, pop-up window (CMP-POP)	Computer-like, survey screen (CMP-SCN)
n	515	510	529	451	462
break-off (%)	1 (0.2%)	6 (1.2%)	0 (0%)	1 (0.2%)	1 (0.2%)

Effect of Intervention on Respondent Effort

Next, I evaluated whether the intervention improved respondent effort and whether this differed across different designs of the intervention. Respondent effort was measured by (1) instances of speeding, (2) willingness to go back after prompted, and (3) other indicators.

Speeding

Table 3.3 presents the average number of speeding instances for the three types of questions in this experiment, as well as the average total number of speeding instances. (Because the prompts could only occur after respondents answered the first behavioral frequency question, the analysis of speeding excluded the first question.) The right-most column of Table 3.3 shows the comparison between the control condition and the four intervention conditions combined. For all question types, all four intervention conditions consistently yielded fewer speeding instances than the control condition. The effectiveness of the interventions at curtailing speeding seems to decline as the distance between the intervention prompts and the question increases. The reduction of speeding is most noticeable for the behavioral frequency questions (where speeding was prompted); for the sensitive questions (which followed the behavioral frequency questions), the speeding reduction is still significant, but the effects seem to be smaller; for the grid questions (which came after the sensitive questions), the intervention seems to have little effect.

Table 3. 3. Speeding Instances in Each Experimental Condition

Instances of Speeding	Control	Human-like, pop-up window (HM-POP)	Human-like, survey screen (HM-SCN)	Computer-like, pop-up window (CMP-POP)	Computer-like, survey screen (CMP-SCN)	Control vs. Intervention (*t*-test)
4 Behavioral frequency Qs	1.47	0.95	1.08	1.02	1.24	t=6.95, p<.0001
3 Sensitive Qs	1.54	1.38	1.49	1.43	1.48	t= 2.51, p=0.012
2 Grid Qs	0.83	0.74	0.75	0.79	0.80	t=1.37, p=0.171
Total	**3.85**	**3.07**	**3.32**	**3.26**	**3.53**	***t*=4.86, *p*<.0001**

Next, I focused on the behavioral frequency questions to investigate which intervention design is more effective in reducing the speeding. I also examined whether the effect of different intervention designs depends on respondent characteristics. The respondent characteristics consist of both demographic and survey-related variables, including age (18-24, 25-34, 35-44, 45-54, 55-64, 65 and older), gender, education levels (HS or less, some college/associate degree, BA, post-graduate degrees), the number of joined online panels (1, 2-3, >=4 panels), and reason for joining online panels (voice to be heard, for fun, for money, and "other" reason). I used ANOVA to analyze the instances of speeding, including the main effects of the experimental conditions and respondent characteristics, as well as the interactions between the two. The results from the final ANOVA model are presented in Table 3.4 where the overall effects not significant at level of 0.05 were removed.

Table 3. 4. ANOVA Analysis of the Speeding Instances during the Behavioral Frequency Questions

Source	df	F value	*p*-value
Exper conditions	4	13.48	<.0001
Age	5	51.59	<.0001
Gender	1	7.05	0.008
Education	3	7.89	<.0001
# Panel joined	2	8.96	0.0001
Reasons for joining panels	3	9.41	<.0001
Exper conditions × age	20	2.23	0.0013

I was particularly interested in the impact of experimental conditions on speeding, and whether the effects depend on respondent demographics. As can be seen in Table 3.4, there was a significant main effect of experimental conditions and significant interactions between the experimental condition and age. To show these interactions, Figure 3.2

presents the average number of speeding instances by experimental condition and by age. Figure 3.2 shows different patterns for the younger and older age groups. For the two younger age groups (18-24 and 25-34), speeding was noticeably less with the intervention than in the control condition. Further analysis shows that the speeding reduction did not differ across the four intervention conditions.[12] The effect of the interventions is measured against the amount of speeding without the intervention. If speeding is rare in the control condition, the interventions cannot reduce speeding beyond its already low level – a floor effect. This is exactly what was observed among older respondents – i.e., little speeding in the control condition and therefore no further reduction in speeding due to the experimental conditions.

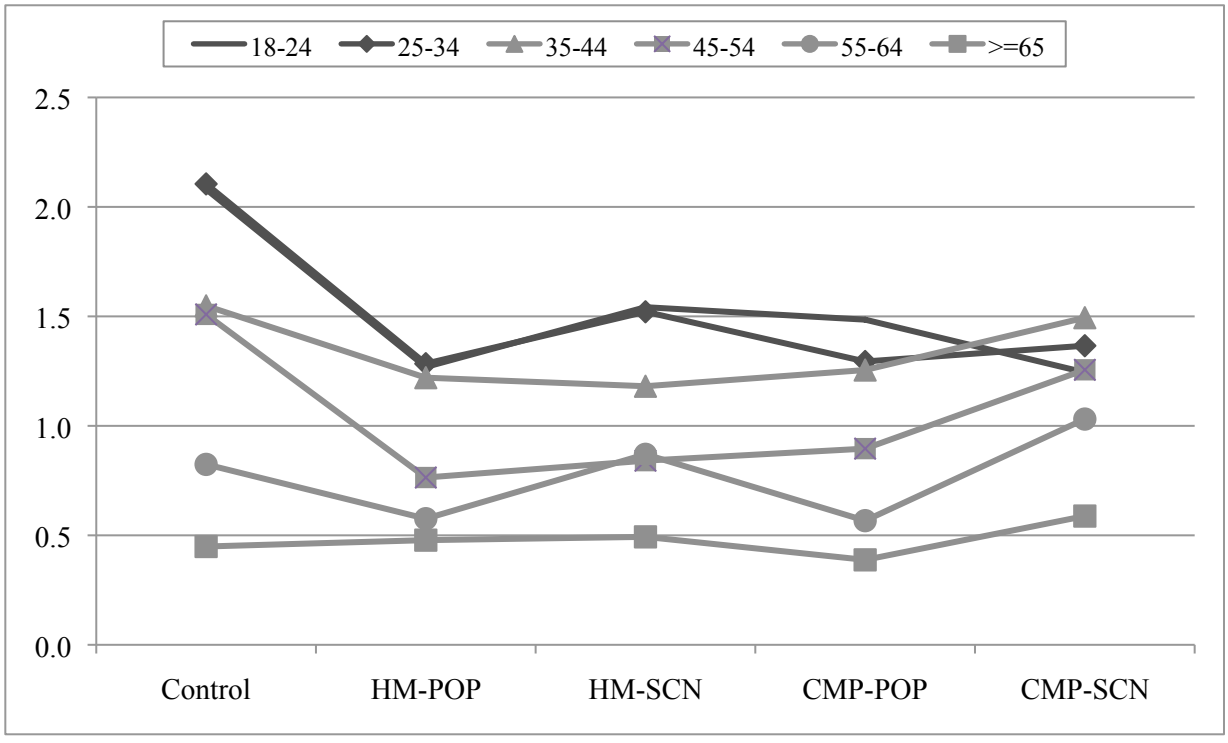


Figure 3. 2. Speeding on the Behavioral Frequency Questions by Experimental Conditions and Age

[12] I conducted ANOVA of speeding instances in the four intervention conditions among the two younger age groups. The effects of the intervention designs were not significant (F=1.40, df=3, p=0.2403).

Table 3.4 also reveals that a few respondent characteristics, including gender, education level, number of panels joined, and reasons for joining panels, affected speeding, regardless of the experimental conditions. The bivariate analyses show that (1) male respondents sped more often than female respondents; (2) more educated respondents sped more often than the less educated; (3) respondents who joined only one panel or many panels (>=4 panels) sped more often than the respondents who joined a few panels (2-3 panels); and (4) surprisingly, respondents who reported "voice to be heard" as their primary reason for joining the panel actually sped more often compared to the respondents who reported other reasons.

Backing up

The speeding intervention, if triggered, offers respondents the option of going back to the question on which speeding is detected and re-considering their answers. Thus, another measure of effort is whether the fast respondents are willing to go back after the intervention prompts. Table 3.5 presents the percentages of respondents who went back when prompted in each of four intervention conditions. The comparisons show that computer-like intervention in pop-up windows (CMP-POP) yielded the highest percentage of going-back across all four designs of the intervention. I also analyzed whether the tendency to go back depends on respondent characteristics. I found no significant impact of respondent characteristics, or their interactions with the experimental conditions on the decisions to back up.

Table 3. 5. Percentages of Backing Up by the Four Intervention Conditions

	(1) HM-POP	(2) HM-SCN	(3) CMP-POP	(4) CMP-SCN
% Backing Up	24.6%	15.5%	46.0%	24.4%
Comparisons	(3) vs. (1): $p<.0001$; (3) vs. (2): $p<.0001$; (3) vs. (4): $p<.0001$;			

Other indicators

In addition to speeding and backing up, I also examined other indicators of response quality, including rounding in the behavioral frequency questions (i.e., answers are multiple of five), and answer consistency in the grid questions. Overall, there was limited evidence that these behaviors were affected by the intervention prompts nor by the different prompt designs. For rounding, I examined both the total instances of rounding and the occurrence of any rounding during the behavioral frequency questions. Both measures indicated that rounding was not significantly different across the experimental conditions. Both of the grid questions in this study contain items that are somewhat reverse-worded (see Appendix 3.3 for the exact question wording). In this case, straightlining implies inconsistency of the responses. The findings on straightlining depend on the measure: the analysis of straightlining on both the grid questions showed no significant difference across the experimental conditions (χ^2=2.08, df=4, p=0.7206); but the percentage of straightlining on either of the questions was significant reduced in CMP-POP condition compared to other four conditions ((9.7% vs. 14.0%, χ^2=5.73, df=1, p=0.0167).

Effects of intervention on answers to the sensitive questions

Following the behavioral frequency questions, three sensitive questions asked about marijuana use, exercise, and binge drinking. Table 3.6 compares the proportions of undesirable answers to each of these questions across the experimental conditions. For the questions on marijuana use and exercise, the four intervention conditions yielded fewer undesirable answers, although the difference is only significant for the exercise

question. For the binge drinking question, the answers appear to be very similar across all the condition.

I also examined whether the impact of intervention on the sensitive questions further depended on respondent characteristics. For all three sensitive questions, respondents who reported they joined panels because they wanted "voice to be heard" seem to be particularly affected by the experimental conditions. Table 3.7 presents the regression analysis for all three questions; The dependent variable was whether respondents reported any undesirable answers to the three sensitive questions. I examined the effects of experimental conditions, respondent characteristics, as well as the interactions between the two. (Table 3.7 only includes the effects related to the experimental conditions. Appendix 3.4 shows the regression results for all the parameters.) As can be seen in Table 3.7, there were significant effects of the experimental conditions as well as interactions between the experimental conditions and respondent reasons for joining the panel(s). The negative and significant coefficients for the four intervention conditions indicate that the reports of undesirable behaviors were reduced by the prompts among the "voice to be heard" respondents (the reference category for the reason for joining panels). The additional test shows that the four coefficients are not significantly different from each other, suggesting the reduction of socially undesirable answers among the "voice to be heard" respondents did not differ across the four intervention conditions. The regression also reveals significant interactions between experimental conditions and reasons for joining panels. The positive coefficients of the interaction terms indicate that for respondents who reported other reasons for joining panels, the reduction of socially undesirable answers by the

intervention was less than that for “voice to be heard” respondents. To show those interactions, Table 3.8 presents the percentage of respondents reporting any socially undesirable answers by experimental conditions and reasons for joining online panels. For the “voice to be heard” respondents, about 80% reported at least one socially undesirable behavior in the control condition, while the percentage was reduced by about 15 to 20 percentage points in the four intervention conditions. For the respondents who reported they joined panels “for fun”, the percentage of reporting any socially undesirable answers appear to be also reduced, but to a lesser extent, in the intervention conditions; and the differences are not significant. For the “for money” respondents, the variation is even smaller and not significant.

Table 3. 6. Comparisons of Undesirable Answers to the Three Sensitive Questions across the Experimental Conditions

3 sensitive questions	Control	HM-POP	HM-SCN	CMP-POP	CMP-SCN	χ^2 test
Marijuana (yes)	44.5%	41.1%	40.5%	41.1%	37.9%	p=0.3507
Exercise (less than once a week)	40.8%	39.3%	38.7%	34.4%	33.4%	p=0.0776
Binge drinking (>=1 day during the past month)	33.3%	31.0%	32.8%	33.0%	33.1%	p=0.9428

Table 3. 7. Logistic Regression of Report of Any Undesirable Answers to the Sensitive questions*

Parameter	Estimate	SE	Odds ratio	p-value of χ^2
Effect of experimental conditions (ref=control)				
HM-POP	-0.72	0.22	0.49	0.0010
HM-SCN	-0.60	0.22	0.55	0.0058
CMP-POP	-0.71	0.23	0.49	0.0016
CMP-SCN	-0.53	0.23	0.59	0.0218
Interactions between experimental conditions (ref=control) and reasons for joining panels (voice to be heard)				
HM-POP × for fun	0.34	0.30	1.41	0.2505
HM-SCN × for fun	0.51	0.31	1.67	0.0974
CMP-POP × for fun	0.43	0.33	1.54	0.1877
CMP-SCN × for fun	0.30	0.34	1.35	0.3813
HM-POP × for money	0.71	0.25	2.03	0.0041
HM-SCN × for money	0.50	0.24	1.65	0.0353
CMP-POP × for money	0.54	0.25	1.71	0.0329
CMP-SCN × for money	0.28	0.25	1.32	0.2721

*****This table only includes the effects related to the experimental conditions. See Appendix 3.4 for the regression results for all the parameters.*

Table 3. 8. Report of Any Undesirable Answers to the Sensitive Questions by Experimental Conditions and Reasons for Joining Panels

Reason for panel(s)	Control	HM-POP	HM-SCN	CMP-POP	CMP-SCN	χ^2 test
Voice to be heard	81.3%	62.5%	66.2%	61.9%	65.8%	p=0.0032
For fun	81.1%	68.9%	75.3%	72.4%	70.4%	p=0.4464
For money	73.8%	77.8%	74.8%	74.7%	71.9%	p=0.6817

Respondent evaluation of the intervention design

At the end of the survey, people who received the intervention were asked to evaluate it. In addition, those in the two human-like conditions were asked to evaluate the facial image displayed in the intervention. The findings are presented in Table 3.9 and Table 3.10. As can be seen in Table 3.9, HM-POP design of the intervention reduced the feeling of completing a paper questionnaire compared to the other three designs of the intervention. However, when asked how much the intervention feels like interacting with a person, respondents rated the four intervention conditions simiarly. Table 3.9 also shows that respondents in the CMP-POP condition reported being interrupted more by the intervention compared to those in the other intervention conditions. (The difference is marginally significant). This may be somewhat related to the finding that respondents in CMP-POP condition were more likely to go back after the prompts.

For all the three questions evaluating the facial images, Table 3.10 shows puzzling interactions between the facial image and how it was presented in the intervention (in a pop-up window vs. on the next survey screen). Specifically, for one facial image (face 1 in Table 3.10), respondents reported paying less attention, liking the face less, and finding the face less attractive when the face was presented in the pop-up window compared to when it was on the next survey screen. However, such difference was reversed for the other two facial images. The ANOVA analyses show that these interactions were significant for all the three evaluation questions (results not reported here).

Table 3. 9. Respondent evaluation of the intervention*

Evaluation of the intervention	HM-POP	HM-SCN	CMP-POP	CMP-SCN	Diff
How much did the intervention … (1=not at all; 9= extremely)					
feel like completing a paper questionnaire	4.15	4.78	4.67	4.51	HM-POP vs. others: t=2.91, p=0.0037
feel like interacting with a person	4.13	4.12	4.07	3.99	Any diff among the four conditions: F=0.2, df=3, p=0.8997
interrupt the respondent	3.86	3.84	4.12	3.83	CMP-POP vs. others: t= -1.63, p=0.1032

* *The questions were only administered to the respondents who received any intervention.*

Table 3. 10. Respondent evaluation of the facial image in the intervention*

Evaluation of the facial image in the intervention	Face 1		Face 2		Face 3	
	HM-POP	HM-SCN	HM-POP	HM-SCN	HM-POP	HM-SCN
How much did you … (1= not at all; 9=extremely)						
pay attention to the face	4.64	4.82	5.01	3.97	5.20	4.00
like the face	5.15	5.33	5.47	4.56	5.23	4.60
find the face attractive	5.15	5.31	5.67	4.53	5.35	4.59

* *The questions were only administered to the respondents who received any intervention in the two human-like conditions.*

3.4 Discussion and Conclusions

This study explored whether interactive interventions in Web surveys can be designed to induce the illusion of interacting with a human agent. The sense of interacting with a person might lead respondents to be more cooperative with the intervention, thus making the intervention more effective compared to when respondents simply treat interventions as automatic computer feedback. In addition, this feeling of interacting with another person, if triggered, is likely to reduce the reports of socially undesirable answers to the sensitive questions.

Specifically, this study manipulated the picture and the text of the intervention so that the intervention would seem more like either a message from a human or from a computer program. The experiment also contrasted how the intervention was presented – either on the subsequent survey screen or in a pop-up window – to test the hypothesis that the latter might help create the sense of an agent stepping into the data collection process. In addition, the experiment also included a control condition with no intervention.

Overall, the effects of the intervention were similar across different designs. As the intervention in this study targeted speeding, the most direct impact of intervention should be on speeding. The findings show that speeding was equally reduced by the four intervention designs. In addition, the different designs for intervention seem to equally reduce the report of socially undesirable answers in the later sensitive questions. The only evidence that design of the intervention matters is that the computer-like intervention presented in a pop-up window yielded much more backing up compared to other designs of the intervention. My explanation is that such an intervention design (i.e., yellow error sign in a pop-up window with “yes/no” option for respondents to go back to the previous

question) resembles the traditional design of computer prompts where "yes" is often the recommended and correct action, therefore, such design may lead respondents to more likely to click on "yes".

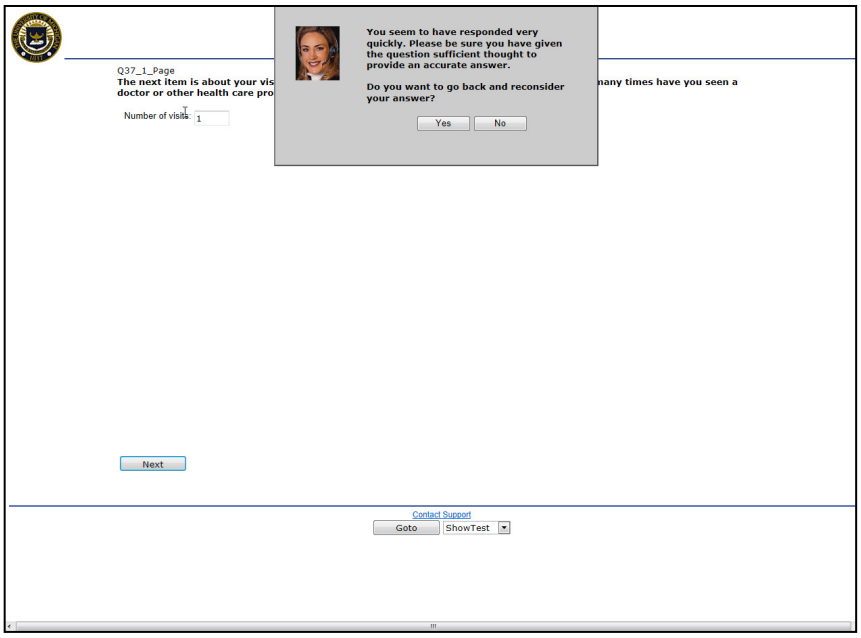
In summary, the manipulation of the intervention design in this study seems to produce limited impact on how respondents react to the intervention. Specifically, respondents who received human-like intervention in a pop-up window, which I thought would be most likely to induce the sense of being interacting with a human agent, did not expend more effort nor report more socially desirable answers compared to those in the conditions where the intervention was more obviously computer program. This suggests that this study did not identify an intervention design that can successfully trigger the illusion of being interaction with a person. This failure may be due to the cues of humanness (i.e., picture and text) being too subtle. The unsuccessful attempts of both this study and Tourangeau, Couper and Steiger (2003) seem to suggest that in real survey settings (vs. laboratory-based experiments) more blatant cues of humanness are necessary to induce the illusion of interacting with a human.

A finding not central to the original purpose of this study is that the intervention, regardless the design, seems to yield fewer reports of socially undesirable answers compared to the condition with no intervention. While all the previous studies on intervention (including the previous chapter of this dissertation) have focused on the impact on respondent effort, this current study shows that intervention has another consequence – more socially desirable answers. So, what is it about intervention that can motivate respondents to work harder and provide more socially desirable answers at the same time? This may be rooted in the fact that survey taking, even self-administered, is

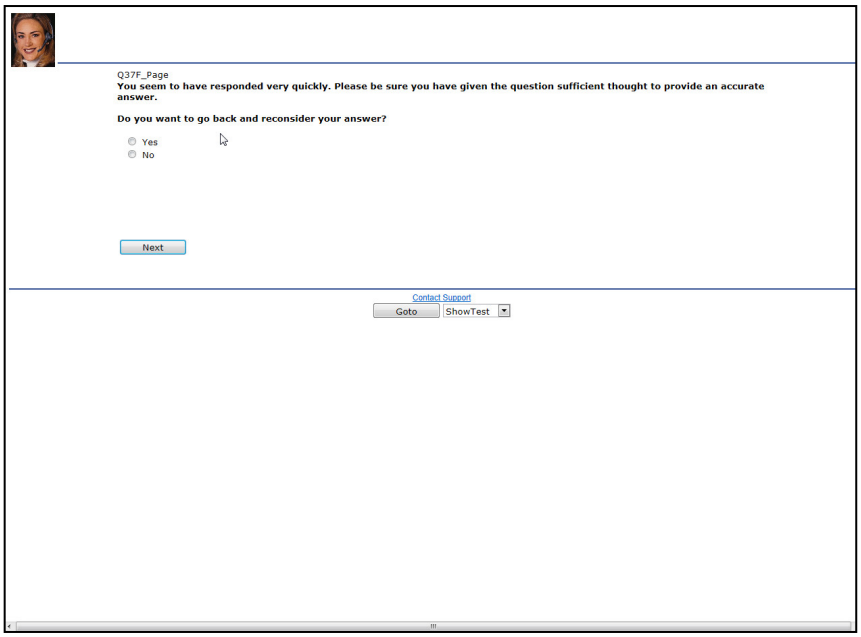
still a social interaction between respondents and the survey organization. Although Web surveys typically feel very private, respondents' data (more and more this includes paradata regarding respondent behaviors) are analyzed and examined by other people. (This contrasts with a complete private activity, such as writing a diary.) This sense of social interaction and consequently, the sense of social presence, can be triggered or enhanced by the interactive intervention. Specifically, the speeding intervention in this study may make it salient to respondents that their behaviors are captured and monitored by the survey organization. As a Web survey feels more social and as the presence of the survey organization becomes more salient, respondents may feel more obliged to present themselves in a positive light, thus motivating them to work hard and report in a socially desirable way. This effect could be particularly strong among respondents with higher self-presentation concerns. This may help explain the finding that the social desirability bias is larger among "voice to be heard" respondents, assuming that their reported reason for joining online panels somewhat reflects greater need for impression management compared to respondents who reported joining panels for fun or for money.

Appendix 3. 1. Screen captures of interactive prompts in the four intervention conditions

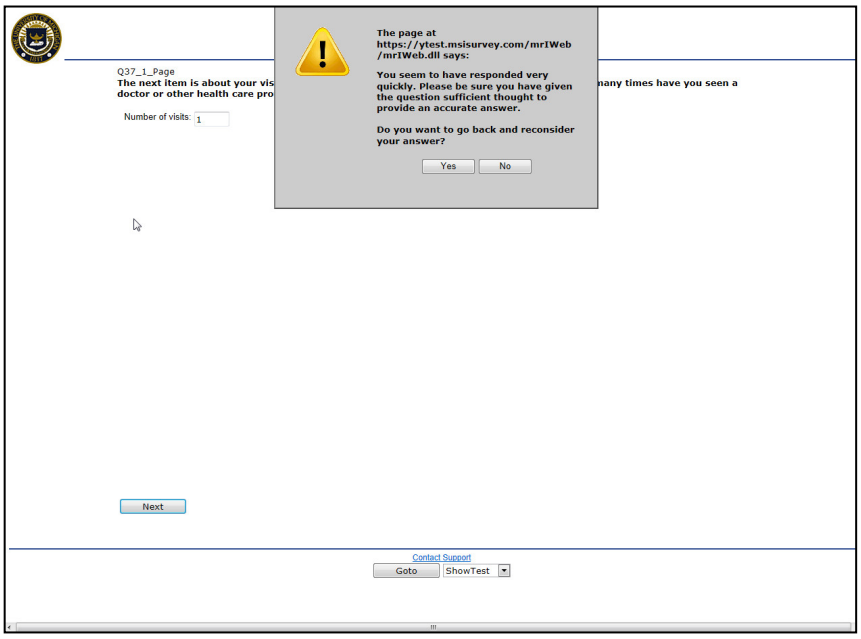
1. Human-like intervention in a pop-up window (HM-POP)



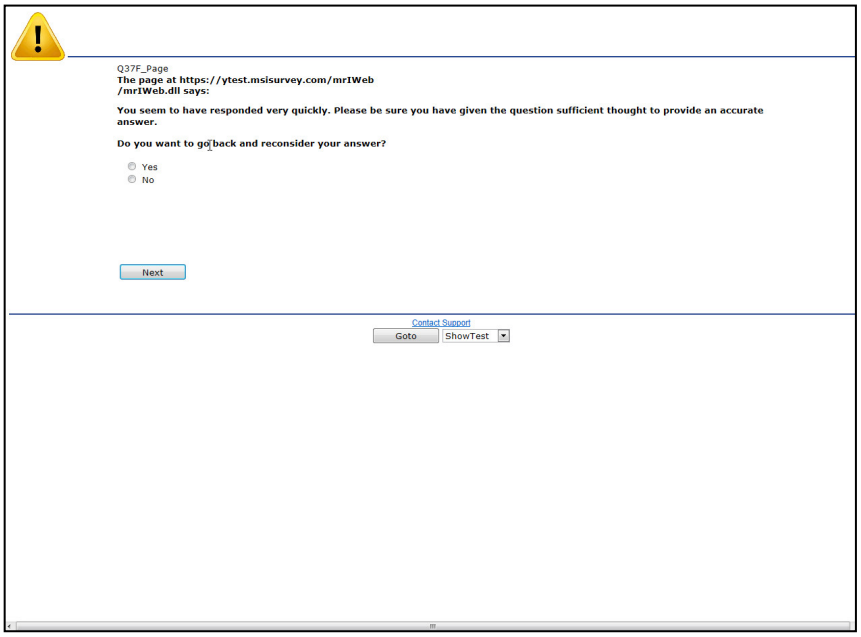
2. Human-like intervention in the next survey screen (HM-SCN)



3. Computer-like intervention in a pop-up window (HM-POP)



4. Computer-like intervention in the next survey screen (HM-SCN)



Appendix 3. 2. Pictures of three faces used in the two human-like intervention conditions (HM-POP and HM-SCN)



Appendix 3. 3. Question wording

Phase 1. Intervention

FREQ_1. The next item is about your visits to health care providers. During the PAST 12 MONTHS, how many times have you seen a doctor or other health care professional about your own health?

Number of visits: [RECORD NUMBER]

FREQ_2. During the PAST MONTH, how many days have you felt you did not get enough rest or sleep?

Days: [RECORD NUMBER]

FREQ_3. During the PAST MONTH, how many times have you eaten in restaurants? Please include both full-service and fast-food restaurants.

Times: [RECORD NUMBER]

FREQ_4. During the past month, how many times have you shopped in a grocery store? If you shopped at more than one grocery store on a single trip, please count them separately.

Times: [RECORD NUMBER]

FREQ_5. During the PAST 2 YEARS, how many overnight trips have you taken?

Number of trips: [RECORD NUMBER]

Phase 2. Intervention Evaluation

SENQ_1. Have you ever, even once, used marijuana or hashish?

Yes
No

SENQ_2. In a typical week, about how often do you exercise?

Less than 1 time per week
1 or 2 times per week
3 times per week
4 or more times per week

SENQ_3. During the PAST MONTH, how many times did you have 5 or more drinks of alcoholic beverage on one occasion?

Times: [RECORD NUMBER]

GRIDQ_1. Please indicate how much you favor or oppose each of the following statements.

1 Strongly oppose
2 Somewhat oppose
3 Neither favor nor oppose
4 Somewhat favor
5 Strongly favor

Avoiding "fast food"
Maintaining a healthy diet
Monitoring cholesterol levels closely
Emphasizing the taste of food rather than its nutritional value
Paying close attention to the nutritional information on food packaging
Limiting the amount of red meat in your diet
Balancing one's diet across the key food groups

GIRDQ_2. For each of the following statements, please rate how much it characterizes you.

1 Extremely uncharacteristic
2 Somewhat uncharacteristic
3 Uncertain
4 Somewhat characteristic
5 Extremely characteristic

I prefer to avoid taking extreme positions
I want to know exactly what is good and bad about everything
If something does not affect me, I do not usually determine if it is good or bad
There are many things for which I do not have a preference
I like to have strong opinions even when I am not personally involved
I would rather have a strong opinion than no opinion at all
I only form strong opinions when I have to
I am pretty much indifferent to many important issues

Phase 3. Design Assessment

EVAQ_1. During the survey, you were given feedback indicating that you responded quickly. We would like to ask a few questions about this so we can improve the design of this kind of feedback.

1	2	3	4	5	6	7	8	9
Not at all								Extremely

When you received this feedback…

a. How much did it feel like you were completing a paper questionnaire?
b. How much did it feel like you were interacting with a person?
c. How much did the feedback interrupt you as you were answering the questions?

EVAQ_2. When you were prompted for answering quickly, a picture of a human face appeared along with the prompt. We would like to get your views about this face.

1	2	3	4	5	6	7	8	9
Not at all								Extremely

a. How much did you pay attention to the face?
b. How much did you like the face?
c. How attractive did you find the face?

Appendix 3. 4. Logistic regression of report of any undesirable answers to the sensitive questions (including all the regression parameters)

Parameter	Estimate	SE	Odds ratio	*p*-value of χ^2
Effect of experimental conditions (ref=control)				
HM-POP	-0.72	0.22	0.49	0.0010
HM-SCN	-0.60	0.22	0.55	0.0058
CMP-POP	-0.71	0.23	0.49	0.0016
CMP-SCN	-0.53	0.23	0.59	0.0218
Interactions between experimental conditions (ref=control) and reasons for joining panels (voice to be heard)				
HM-POP × for fun	0.34	0.30	1.41	0.2505
HM-SCN × for fun	0.51	0.31	1.67	0.0974
CMP-POP × for fun	0.43	0.33	1.54	0.1877
CMP-SCN × for fun	0.30	0.34	1.35	0.3813
HM-POP × for money	0.71	0.25	2.03	0.0041
HM-SCN × for money	0.50	0.24	1.65	0.0353
CMP-POP × for money	0.54	0.25	1.71	0.0329
CMP-SCN × for money	0.28	0.25	1.32	0.2721
Effect of age (ref=18-24 years old)				
25-34	0.83	0.19	2.29	<.0001
35-44	0.40	0.19	1.50	0.0360
45-54	0.40	0.20	1.50	0.0402
55-64	0.33	0.21	1.39	0.1141
>=65	-0.17	0.21	0.84	0.4080
Effect of education (ref=HS or less)				
Some college/associate degree	-0.09	0.14	0.91	0.5004
BA	-0.41	0.15	0.66	0.0056
post-graduate degrees	-0.87	0.17	0.42	<.0001
Female (ref=Male)	-0.32	0.10	0.73	0.0016
Reasons for joining online panels (ref=voice to be heard)				
For fun	-0.31	0.13	2.29	0.0175
For money	0.01	0.12	1.01	0.9449

References

Conrad, F.G., Schober, M. F., Nielsen, D. (2011). Race-of-Virtual-Interviewer Effects. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, AZ, May.

Conrad, F. G., Tourangeau, R., Couper, M. P., & Kennedy, C. (2009). Interactive Interventions in Web Surveys Can Make Respondents More Conscientious. Paper presented at the annual meeting of the American Association for Public Opinion Research, Hollywood, FL, May.

Conrad, F. G., Tourangeau, R., Couper, M. P., & Zhang, C. (2011). Interactive Interventions in Web Surveys Can Increase Response Accuracy. Paper presented at the annual meeting of the American Association for Public Opinion Research, Phoenix, AZ, May.

Fuchs, M. (2009). Gender-of-interviewer effects in a video-enhanced web survey. *Social Psychology, 40*(1), 37-42.

Holland, J. L., & Christian, L. M. (2009). The influence of topic interest and interactive probing on responses to open-ended questions in web surveys. *Social Science Computer Review, 27*(2), 196-212.

Krysan, M., & Couper, M. P. (2003). Race in the live and the virtual interview: Racial deference, social desirability, and activation effects in attitude surveys. *Social Psychology Quarterly, 66*(4), 364-383.

Lee, K., & Nass, C. (2005). Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology, 7*(1), 31-45.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2). Available: http://jcmc.indiana.edu/vol3/issue2/lombard.html

Moon, Y., & Nass, C. (1996). How "real" are computer personalities? *Communication Research, 23*(6), 651-674.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply wen they interact with animated pedagogical agents? *Cognition and Instruction, 19*(2), 177-213.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*(6), 669-678.

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? responses to computer-based interviewing systems. *Journal of Applied Social Psychology, 29*(5), 1093-1110.

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, C. D. (1995). Can computer personalities be human personalities? *InternationalJournal of Human-Computer Studies, 43*, 223-239.

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27*(10), 864-876.

Pratt, J. A., Hauser, K., Ugray, Z., & Patterson, O. (2007). Looking at human-computer interface design: Effects of ethnicity in computer agents. *Interacting with Computers, 19*(4), 512-523.

Sproull, L., Subramani, M., Kiesler, S., Walker, J., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11*(2), 97-124.

Tourangeau, R., Couper, M. P., & Steiger, D. M. (2003). Humanizing self-administered surveys: Experiments on social presence in web and IVR surveys. *Computers in Human Behavior, 19*(1), 1-24.

Overall Conclusions

Web surveys are known for eliciting honest responses. However, without interacting with a human interviewer, respondents in Web surveys may have little motivation to expend sufficient effort to provide accurate responses – a phenomenon often referred to as satisficing. This dissertation consists of three studies that examine satisficing and the use of interactive intervention to reduce satisficing in Web surveys.

The first study evaluates speeding (or very fast responding) as an indicator for satisficing. By investigating the association between speeding and straightlining, the study demonstrates that speeding is indeed related to satisficing. This is true even for young respondents who are often thought to be able to answer both quickly and thoughtfully.

While a small number of studies have shown that interactive intervention in Web surveys can help reduce satisficing behaviors such as speeding, in general it is still largely unknown as to how such interventions work. In particular, the findings from the first study indicate that unmotivated respondents are likely to perform more than one satisficing behavior. This raises the question as to whether the intervention for different satisficing behaviors can differently affect overall response quality. This is examined in Study 2, which compares intervention for speeding and non-differentiation in grid questions. The findings show that both interventions can improve the overall response

quality – i.e., the positive effect of intervention is not restricted to the behavior targeted in the intervention nor the type of questions where the intervention occurred. These results have several important implications. First, interactive interventions of this sort seem to draw respondents' attention to their reporting behaviors in general, not just the one explicitly mentioned in the intervention. Second, the improvement in various aspects of respondent behaviors also suggests that the changes in respondent behaviors are genuine. (If the effect of interventions were superficial, there would be at most a reduction in the targeted behavior, but not in the other behaviors.)

The third study also examines the design of the intervention. The original intention was to explore interventions that created the illusion among respondents that they were interacting with a human agent. The experiment manipulated the image and the text displayed in the intervention, as well as how the intervention was presented. Overall, the manipulation seems to have limited impact on how respondents react to the intervention. Specifically, respondents in the condition where the intervention was designed to feel more like a message from a human did not expend more effort nor did they report more socially desirable answers than respondents in the conditions where the intervention was more obviously a computer program. As a result, there was no evidence of social presence arising from the illusion of the intervention from a human agent. However, this study does find evidence of social presence as a result of the intervention itself. This social presence is not due to any sort of illusion, but likely a recognition that Web survey respondents are interacting with the survey organization – a real social entity, althougth the interaction takes place in a remote and non-synchronous manner. The results

demonstrate that social presence can exist in Web surveys and be made salient through the interactions with the respondents.