

Understanding Sources of Social Desirability Bias in Different Modes: Evidence from Eye-tracking



Olena Kaminska

Institute for Social and Economic Research
University of Essex

Tom Foulsham

Department of Psychology
University of Essex

No. 2013-04
March 2013

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

Non-technical summary

Among many questions asked in surveys some questions ask people to reveal unpleasant sides of themselves. These questions may ask about behaviours that are not socially accepted (e.g. drug use) or not socially approved (e.g. not voting in elections). A challenge for survey design has been to improve survey reports of unpleasant behaviours or attitudes, which tend to be underreported.

Embarrassment is currently the main explanation for underreporting of unpleasant behaviours and attitudes in the research literature. Respondents are believed to understand the question well, come up with the correct answer, but report another answer that makes them look better. This is done by respondents on purpose in order to avoid embarrassment. This theory has received some empirical evidence and much acceptance among researchers.

The current paper explores whether another source of underreporting of unpleasant behaviours and attitudes may be present. Specifically, respondents may misreport the answer subconsciously, mainly as a result of lack of effort given to answering the question. Effortless answering is likely to lead to reporting pleasant behaviours and attitudes because such reports are easier.

This alternative explanation is tested by tracing respondents' eye movements when they read questions on paper or on a PC screen without an interviewer present, and comparing such movements to the ones when respondents read response options on a card in an interviewer-administered interview. We collected eye-movement and response information from 14 volunteers in a Psychology Laboratory at the University of Essex.

We found support for the theory that some reporting of unpleasant behaviours results from a lack of effort put in answering survey questions in the interviews where an interviewer is not present. Specifically, in such interviews reading survey questions fast leads to a lower chance of providing unpleasant answers. But this theory does not explain why interviewer presence leads to fewer unpleasant answers even if respondents pay more attention when an interviewer is present. This suggests that both sources of error are present: embarrassment and lack of effort.

In addition to the contribution to theoretical discussion, the paper also opens new avenues for improving survey design for surveys with unpleasant questions. By motivating respondents or by communicating the importance of their answers, we may be able to achieve even better quality of responses to such questions.

Understanding sources of social desirability bias in different modes: evidence from eye-tracking

Olena Kaminska⁺

Institute for Social and Economic Research, University of Essex

Tom Foulsham⁺⁺

Psychology Department, University of Essex

⁺ Corresponding author: Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK (email: olena@essex.ac.uk)

⁺⁺ Psychology Department, University of Essex, UK (email: foulsham@essex.ac.uk)

Abstract: Our paper provides an empirical study of whether satisficing is related to social desirability bias using real-world eye-tracking. The method enables detecting latency of eye gazes in web, face-to-face and paper and pencil self-administered (SAQ) modes. Through these we infer respondents' attention to question wording, socially desirable and socially undesirable response scale points. We link the gaze latency measures to responses to understand how respondents arrive at socially desirable or undesirable answers. We find that satisficing is related to social desirability bias in self-completion modes. Yet it does not explain the higher incidence of socially desirable responses in face-to-face mode.

Key words: social desirability, mode effect, satisficing, eye-tracking, GHQ-12

JEL Codes: C81, C83

Acknowledgements: This study was funded by the University of Essex Interdisciplinary Research Fund for Social Sciences "Understanding Survey Responses through Eye-Tracking: Mixed-Mode Comparison" granted in 2011 (Principal Investigator: Olena Kaminska; co-PI: Tom Foulsham). The authors are grateful to Gemma Luxton for hours of diligent coding of eye-tracking data.

Introduction

Social desirability bias is one of the recognized types of measurement error and occurs when a respondent provides an answer which is more socially acceptable than his / her true attitude or behaviour. Studies that have compared survey reports to external records consistently found underreporting of socially undesirable behaviour, e.g. underreporting abortions or drug use (Schaeffer 2000; and Tourangeau et al. 2007) or overreporting of socially desirable behaviour, e.g. overreporting voting (Anderson and Silver 1986; Belli et al., 2001). This poses a challenge for survey methodologies first to understand the causes of social desirability bias, and second to develop the best design methods to improve the reporting of socially undesirable attitudes and behaviours.

The most common source of social desirability bias mentioned in the literature is the respondent's lack of comfort to reveal his or her true attitudes (Tourangeau et al., 2000; Tourangeau et al., 2007; Groves et al., 2009, Holgraves 2004), also called 'impression management' (Holgraves, 2004). Respondents are thought to act so as to avoid the embarrassment, unease and distress that revealing socially undesirable answers may bring. In describing four stages of cognitive processing (comprehension, retrieval, judgement and response) Tourangeau, Rips and Rasinski (2000) suggest that social desirability bias occurs at the last stage of this process: response. This is the stage when respondents map their response onto the scale and may 'edit' it in order to make themselves look more favourable. The theory implies that the respondent has understood the question correctly, retrieved all the relevant information from his/her memory adequately, and has integrated this information well. In other words, the assumption is that the respondent has arrived at the correct answer and is aware of it. Social desirability bias occurs only due to purposeful misreporting of this answer at the last stage of the cognitive process.

Much empirical evidence has been gathered which supports the above theory. First of all, a number of experimental studies have consistently shown higher chance of revealing socially undesirable answers in self-completion modes (either SAQ or web) in comparison to interviewer administered modes, especially a face-to-face mode (Holbrook and Krosnick 2010; de Leeuw 2005; Tourangeau et al., 2000; Schwarz et al., 1991). It is understood that respondents feel high unease in front of interviewers if their answers do not conform to norms or social expectations, and therefore alter their answer in order to look better in the eyes of the interviewer.

Another set of evidence which may be thought to support the theory of social desirability bias being due to embarrassment comes from studies on questionnaire design. Wording questions in less threatening way has been shown to increase socially undesirable answers (Holbrook and Krosnick 2010; Peter and Valkenburg 2011; Malvin and Moskowitz 1983). Such wording may imply that socially undesirable behaviour or attitude exists, and is common, may give excuses for why one may have socially undesirable attitudes or may simply ask more embarrassing questions first such that in light of them the question of interest is perceived as less embarrassing (Barton, 1958). Thus, decreasing embarrassment of a question through wording or question order seems to influence the conscious reporting of true attitudes or behaviour as well.

In an experimental study using web mode Holgraves (2004) found that longer response latencies are related to socially desirable answers, especially in the non-anonymous condition. The finding supports the idea that providing a socially desirable answer takes more effort, and therefore is a result of conscious editing.

The intuitive embarrassment theory of social desirability bias and the above experimental empirical studies have led to clear improvements in how we collect socially sensitive data. It has become common in survey

practice to use a careful nonthreatening wording when asking sensitive questions, and in surveys conducted face-to-face to collect socially undesirable data via self-administered mode (CASI, ACASI or SAQ).

Yet, the embarrassment source is not the only source of social desirability bias mentioned in the literature. In a much less well-known theoretical development of causes for social desirability bias, Schaeffer (2000) and, independently, Holgraves (2004) discuss heuristics. A heuristic process refers to a cognitive process that is superficial or not complete. The term more commonly used in the survey methodology literature to describe this process is satisficing (Krosnick, 1991). Schaeffer suggests that heuristic process can lead to socially desirable answers through a number of ways. First, respondents may retrieve a socially desirable value and use it as an anchor in 'an anchor and adjust response strategy'. Second, respondents may use 'a routine denial' instead of refusing to answer the question, something that can be perceived as socially appropriate in the situation when the person asking an uncomfortable question does not have any right to do so. The author writes (Schaeffer, 2000, p.119): "I would speculate that recipients of intrusive questions routinely use denial as an automatic, socially routine, defensive strategy. Such conventional denials could be offered simply because an intrusive question was asked and without any calculation of risk." Third, the answer may be effortless and ready as a 'public' answer to the question which is part of the respondent's public biography. Schaeffer suggests that respondents may store information at two layers: public information which creates a public image and is easily available; and private information which reflects the truth.

Similarly, Holgraves (2004) outlines two possible ways how responding in a socially desirable way may result from heuristic processing. First, a respondent may simply skip the retrieval stage and provide a response that is only based on socially desirable implications. Second, a respondent may use heuristics during the retrieval stage, mainly by recalling positive information and neglecting negative information. Similarly, Belli et al. (2001) mentions that in recalling factual episodes social desirability bias may result from 'source monitoring', when a respondent makes inferences about his/her behaviour during the retrieval process by confusing other similar events with the episode in question. Such selective or incomplete retrieval could also result in a biased response. Overall, the satisficing theory suggests that producing a socially desirable answer is cognitively easier, and that such a response may be selected when satisficing takes place.

The important distinction between the two theories described here is that the embarrassment theory suggests that the respondent has arrived at a correct answer, is aware of it and purposely misreports the answer. In contrast, the satisficing theory suggests that a respondent has not completed adequate cognitive processing, and therefore arrived at a wrong answer. Importantly, while both causes of social desirability bias cannot theoretically occur simultaneously, they do not contradict each other and it is possible that both are the causes of social desirability bias in different situations.

The current study provides empirical evidence for each theory. We first test whether satisficing leads to a higher likelihood of providing socially desirable answers in self-completion mode. After this we test whether satisficing can explain higher social desirability bias in face-to-face mode in comparison to self-completion mode. The evidence comes from a small-scale laboratory experiment where respondents were randomly assigned to face-to-face, and self-completion modes (SAQ and web). All respondents had their eye gazes recorded via a real-world eye-tracker. Given the innovative nature of real-world eye-tracking in a survey methods context we describe the method in detail in the following section.

Real-world eye-tracking

Eye-tracking is a relatively new method for studying the survey response process. The method involves measuring visual attention by following respondents' eye movements while they respond to survey questions (see Galesic and Yan (2011) for a technical description of the method). Since its introduction to survey methodology by Redline and Lankford (2001), followed by Graesser et al (2006) and Galesic et al. (2008), eye-tracking has gained popularity as a tool to explore visual attention when responding to a web survey (e.g. Ashenfelter and Hughes, 2011; Kamoen et al., 2011; Lenzner et al., 2011a; Lenzner et al., 2011b; Menold et al., 2011; Romano, and Chen 2011). Eye-tracking has been so successful in detecting issues with web questionnaires that it is now used by the United States Census Bureau (Ashenfelter and Hughes, 2011; Romano, and Chen 2011) and the German Federal Statistical Office (Federal Statistical Office, 2011) as a pre-testing tool.

The eye-trackers used in previous web studies of survey responding (e.g. the TOBII X500) are not suitable for studying visual attention in modes other than on computer, as they only detect eye movements within a PC monitor. Eye-movements are recorded by an infrared camera which is synchronized with a stimulus on the PC screen via specialized software such that the screen view is known at any point of time when the respondent is looking at it. This requirement, and the restricted range of such eye-trackers, means that they are limited to tracing gazes on the PC screen only. The only two previous attempts to extend eye-tracking to SAQ mode (Redline and Lankford, 2001; Potaka, 2007) mounted the questionnaire vertically in front of the participant, as if on a computer monitor. However, the set-up and desk-mounted eye-tracker used in these studies were inflexible (because the camera and participant could not move) and so responses may have been different from a real SAQ mode.

To track eye-movements in survey modes which do not involve a PC and have a natural set up we use an innovative real-world eye-tracker, the SMI HED system (Sensorimotoric Instruments; Teltow, Germany). In essence, this eye-tracker follows a similar principle to the one used with the PC: a camera records eye movements, and a calibration routine (where the respondent looks at a series of specified points on a wall or desk) ensures that features from the eye gaze are mapped onto points in the visual field. The respondent's view is not limited this time, and eye movements can be recorded wherever a respondent may look via a system mounted on a headset that is worn on respondent's head. Simultaneously another "scene" camera records the view in front of the respondent, and software maps the eye position onto this video at each point in time. Because of the flexibility of this system the exact view in front of each respondent differs and changes as they move their head. Additional coding is therefore required if a researcher wants to standardize and compare eye movements and fixations across respondents.

Real-world, mobile eye-trackers are a new tool for survey methodology field. Previously, this tool has been used successfully in psychology, sports science and ergonomics. For example, Hayhoe and Ballard (2005) review studies of eye movements during natural behaviours such as walking and even while making tea and sandwiches. The results of these studies emphasize that visual attention is deployed in a highly systematic way and as part of a routine of actions, coordinated to select the most important visual information for each moment in time. Real-world eye-tracking proved a useful tool in studying eye movements in car drivers—who fixate a particular point when steering round a bend (Land, and Lee 1994) —and in sportsmen such as cricketers who are highly skilled at predicting where the ball will bounce (Land and McLeod, 2000).

Foulsham, Walker and Kingstone (2011) compared the eye movements of people walking around the real environment to those of people watching a video on computer and found some important differences. Given

the differences between attention on computer and in the real world, one may expect differences in visual attention across survey modes as well.

Data collection

The data collection was set up in a Psychology laboratory as part of collaboration between a survey methods team and a Psychology department in a UK University. A number of meetings between experts in Psychology and Survey Methods took place to find the closest set up to a real survey interview while obtaining good-quality eye tracking information. While wearing headsets with the real world eye-tracker, respondents were not restricted in their posture, eye and head movement. No instructions on where to look or how to move / sit were given, except during the set up of the equipment. No person (interviewer or instructor) was present when respondents completed the survey in web or SAQ modes: the respondents were left on their own in the lab after the equipment was set up. Only the interviewer was present during the face-to-face interview.

In total, 20 participants were invited from a list of volunteers for Psychology experiments, with the requirement of having normal vision. Because of the complexity of setting up the eye-tracker in SAQ mode, 6 interviews in SAQ mode were uncodable. This resulted in only 2 good quality interviews in SAQ mode. In our analysis we do not study this mode separately from web mode, but combine the information together. This is also in part due to theoretical expectations of these modes being similar to each other with respect to social desirability bias. Overall, the analysis is based on 9 self-completion interviews (7 web, 2 SAQ) and 5 face-to-face interviews.

Of 14 participants, for whom the data were codable, 3 were male (2 in face-to-face mode and 1 in web mode), 3 were not native English speakers (1 in each mode) and all were between 20 and 26 years old. This is not a representative study and the results should not be treated as such. The strength of the study is in its experimental design in which respondents were randomly assigned to modes, and therefore the measurement error observed is not related to mode choice or self selection into mode.

All questions in face-to-face mode were asked with show cards. The response categories were not read out by an interviewer. In this situation, the respondent's awareness of response options is reliant on their visual perception: seeing and reading each option. The interviewer instructed the respondent to turn over each show card before starting to read the question. No instructions were given on how to hold show cards – respondents were free to hold them, place them on laps or on a desk, keep them sideways or wave them around. While eye-tracking in a web mode is possible with a desk-mounted eye-tracker (e.g. the TOBII X500), which codes visual gazes automatically, we decided to use the real-world eye-tracker for web mode for comparability with other modes.

The eye-tracking information was then manually coded by a coder who watched the resulting videos at a slow speed using custom annotation ChronoViz software (Fouse et al., 2011). The eye tracker recorded eye position at 50Hz. The general health questionnaire (GHQ-12) (Goldberg and Williams, 1991), which is used broadly in major and small-scale studies to measure general well-being and distress, was selected for coding and analysis. All 12 questions have a 4-point scale where socially desirable response points are presented at the top of the list. The scale is of ordinal nature, and its wording changes from questions to question with only three questions having the same scale options as the previous one. We coded the beginning and end time points when the interviewer or respondent were reading any instructions; when the interviewer or respondent were

reading the question; when a respondent was looking at either of the top two (socially desirable) response options; when a respondent was looking at the bottom two (socially undesirable) response options; when a respondent was looking away (at an interviewer / or something else e.g., the wall); and when the respondent's gazes were uncodable. Of all gazes 13-14% were uncodable in each mode, due to natural eye blinks and tracker data loss. Overall, we obtained time information for 1847 gazes (eye-fixations lasting over 100 milliseconds). For analysis, we combine much of this information to obtain question-level data.

The analysis is conducted at the question level. Given that questions are nested within respondents and respondents may vary in average speed reading and their tendency to look at particular response options, we control for clustering within respondents in all models presented in this paper. Logistic regression with the `svyset` option in Stata 12.1 is used.

Analysis and results

1 Satisficing as a cause of social desirability bias in self-completion mode

To test whether satisficing is related to the tendency to give socially desirable responses we first explore self-completion modes. Here respondents are free to read and process questions carefully or superficially. Our expectation is that a lack of careful attention to a question will be related to a higher chance of giving socially desirable answers.

From eye-tracking a number of different measures can be obtained, including the time spent reading the question, the time spent looking at socially desirable or undesirable response options, or the time spent on the question overall. Given that time looking at response options may reflect both time reading and selecting a response option, and therefore may be confounded with the option selected, we use the time respondent spent reading the question. Question reading usually depends on its number of words. To control for this, we divide the time spent on the question by the number of words, thus reflecting the average speed of reading per question word. We do not believe that satisficing is linearly related to question reading speed. It is more likely that those questions which are skimmed faster than usual are read with satisficing. We therefore create an indicator of whether a question is read faster than average within self-completion modes or not, and use this as a proxy measure for satisficing. In our dataset this corresponds to a reading speed of 400 milliseconds or less per word. Readings were not coded for 11 questions, which were excluded from the analysis in this section.

It is of interest to know whether fast reading is related to some particular questions or particular respondents. We find that each question got at least one fast reading, and the number of fast readings ranges between 2 and 6 of 9 readings per question. More variability is observed among respondents. Of the 9 respondents, 2 read all but 1 question carefully, and 2 read over 10 of 12 questions fast, while others read between 3 and 9 questions fast. The two respondents who are non-native English speakers did not read slower than others (with 3 and 9 questions read fast); and t-test of average word speed within a question showed no significant difference between native and non-native speakers. Fast reading is therefore not restricted to particular questions or particular respondents and we proceed with the analysis of interest.

We use logistic regression with socially desirable responses as an outcome (socially desirable answer=1; socially undesirable answer=0), and fast reading as an explanatory variable. Results, presented in Table 1 (Model 1), indicate that the relationship between fast reading and socially desirable responses is significant.

Questions read faster than average have 4.2 times higher odds of getting socially desirable answer than questions read slower.

Table 1. Socially desirable answers as a function of fast question reading in self-completion mode

	Model 1: all questions				Model 2: questions with gazes at soc. undes. options			
	Parameter	CI low	CI up	p-val	Parameter	CI low	CI up	p-val
Intercept	0.70	0.21	2.38	ns	0.39	0.16	0.98	0.05
fast reading	4.15	1.11	15.54	0.04	1.86	0.47	7.43	ns
N questions	97				51			
N respondents	9				8			
Model fit	F(1,8)=6.19, p=0.04				F(1,7)=1.12 p=0.33			

*the parameters reported are odds for the intercept and odds ratios for the explanatory variable

One may argue that the relationship between satisficing and selecting socially desirable responses may be mediated by a primacy effect. In other words, in self-completion mode socially desirable answers may have higher selection chance only because they are presented at the top of the scale. To test this we next restrict our analysis to only those questions during which a socially undesirable answer is fixated at least once. The fixation time is selected to be at least 200ms, which is a low estimate for the average fixation duration during reading (Rayner, 1998). Of 97 questions 51 got fixations on socially undesirable options.

When analysis is restricted to these, the relationship between fast question reading and selecting socially desirable responses becomes nonsignificant (Table1, Model 2). This may indicate that satisficing effect on social desirability bias is mediated by primacy effect: a respondent reads a question fast, reads only first options, skipping the bottom ones, and therefore selects socially desirable answer. An alternative is that satisficing leads to both, primacy effect and to higher chance of socially desirable answers; and by selecting those respondents which have read bottom categories we limited our analysis to optimizers. Our data does not allow us to answer which of the two mechanisms takes place.

2 Satisficing as a cause of social desirability bias in different modes

Similar to previous studies (see reviews in de Leeuw 2005; Tourangeau et al., 2000) we find significantly higher levels of socially desirable responses in face-to-face mode than in self-completion modes (Table 2, Model 3). One expectation is that this discrepancy is due to higher embarrassment in face-to-face mode. An alternative explanation, given the results from the previous section, is that the difference may be caused by higher satisficing in face-to-face mode. This explanation, although unlikely, has not been tested empirically before. This section explores whether the difference in social desirability bias between face-to-face and self-completion modes can be explained, at least in part, by satisficing theory.

Table 2. Socially desirable responses depending on the mode of interview

	Model 3: all questions				Model 4: questions with gazes at soc undes. options			
	Parameter	CI low	CI up	p-val	Parameter	CI low	CI up	p-val
Intercept	1.63	0.56	4.74	ns	0.74	0.21	2.66	ns
face-to-face	6.73	1.35	33.46	0.02	12.38	2.31	66.48	0.01
N questions	168				112			
N respondents	14				13			
Model fit	F(1,13)=6.6, p=0.02				F(1,12)=10.65, p=0.01			

*the parameters reported are odds for the intercept and odds ratios for the explanatory variable

Unlike question reading in self-completion mode, the time during which the question is read in face-to-face mode does not depend only on the respondent, but on the respondent-interviewer interaction. This is therefore a less appropriate measure of respondent's satisficing. Yet, if satisficing is higher in face-to-face mode we would observe fewer gazes at the options in the end of the response list. Given that these options are socially undesirable, less attention to them could explain less likelihood of their selection. Recall, interviewers did not read out response options – response options were available only on show cards. Thus respondents in both modes were relying only on visual communication channels for response options. We find that socially undesirable options are looked at in more questions in face-to-face mode than in self-completion modes (Table 3).

Table 3 Questions with gazes on socially undesirable answers as a function of mode

	Parameter	CI lower	CI upper	p-value
Intercept	1.30	0.45	3.72	ns
face-to-face	4.37	1.05	18.19	0.04
N questions	168			
N respondents	14			
Model fit	F(1,13)=4.98, p=0.04			

*the parameters reported are odds for the intercept and odds ratios for the explanatory variable

And restricting the analysis to those questions that feature a gaze of at least 200ms on socially undesirable answers does not change the conclusion: selecting socially desirable answers is still significantly more likely in face-to-face mode than in self-completion modes (Table 2, Model 4). This contradicts the suggestion that face-to-face mode achieves more social desirable answers because of satisficing. In fact, the results support the alternative theory of embarrassment as a dominant source of social desirability bias in a face-to-face mode.

Finally, we test whether the latency of gazes on socially undesirable response options may explain the difference in responses between modes. Our findings indicate no significant difference in average time respondents fixate at socially undesirable options between face-to-face and self-completion modes among the questions that receive such gazes.

Table 4 Total gaze time on socially undesirable options (in milliseconds) as a function of mode among question that received such gazes

	Gaze time	CI lower	CI upper	p-value
Intercept	1614.76	647.60	2581.91	0.003
face-to-face	279.36	-766.49	1325.21	ns
N questions	112			
N respondents	12			
Model fit	F(1,12)=0.34, p=0.57			

Overall, we do not seem to find any evidence of higher satisficing in face-to-face mode than in self-completion modes. If anything, we find evidence of the opposite: respondents are more likely to read socially undesirable answers in face-to-face mode than in other modes. While we cannot show that social desirability bias observed in face-to-face mode is due to embarrassment, we can conclude that it is unlikely a result of satisficing. And therefore, another source, very possibly embarrassment, must be causing the difference in social desirability bias between interviewer- and self-administered modes.

Discussion

For many years social desirability bias has been thought to occur as a result of respondent's embarrassment and conscious misreporting. Much research has explored how to decrease embarrassment through either changing question wording or question context, or via switching to self-administered mode. Yet, causes of social desirability bias, aside of embarrassment, have not received much research attention until now.

Our paper provides first empirical findings suggesting that social desirability bias may occur as a result of satisficing. Specifically, we find that in a self-completion mode faster question reading is linked to a higher chance of giving socially desirable answers. We observe that such questions tend also to receive fewer readings of answer options at the bottom of the scale (socially undesirable answers in our study). In our opinion this may be a result of a primacy effect only, of arriving at socially undesirably answer via satisficing before looking at the response scale, or via the combination of the two.

Primacy effects have been previously linked to satisficing (Krosnick, 1991; Galesic et al., 2008). It is possible that satisficing respondents simply do not look at bottom response options and as a result select earlier options, which happen to be socially desirable in our study. Primacy effects tend to be discussed in the context of long response option lists of categorical nature, where many categories may be suitable as an answer (Tourangeau et al., 2000; Schwartz et al., 1991). The questions in this study are of ordinal nature and are mutually exhaustive, where only one option is theoretically closest to the 'true' value. The usual argument for primacy effects that a respondent selects the first suitable answer seen cannot explain primacy effect in the situation of an ordinal scale of sensitive nature – as such an approach theoretically would lead the respondent to read the response scale until they found the 'best matching' answer. We suspect that in the situation of sensitive questions with an ordinal scale primacy effect may work via 'an anchor and adjust response strategy' mentioned by Schaeffer (2000). Respondents may use the first categories seen as an anchor in thinking about their answers to the question. If this is true, sensitive questions asked in a self-completion mode should

reserve the top of an ordinal scale for socially undesirable answers. We hope future research will experiment whether such question design improves revealing of socially undesirable behaviours and attitudes.

Alternatively, through engaging in satisficing a respondent may arrive at a “routine denial” or a “public answer” (Schaeffer, 2000) first, then search the response scale to find the appropriate answer, and thus without looking at later response options, provide a socially desirable answer. In this situation an increase in motivation may improve the quality of answers, for example suggesting that the following set of questions may influence policy decisions. This is another area where future research should implement experimental testing.

The set of results from mode comparison confirms previous thinking in survey methods literature: less satisficing is found in face-to-face mode; and social desirability bias which occurs in face-to-face mode in higher rate than in self-completion mode is found to not be caused by satisficing. Face-to-face mode is often described as a high quality mode due, in part, to the presence of many communication channels: verbal and nonverbal, aural and visual. Interestingly, even when respondents rely only on visual channel for perceiving response options face-to-face mode achieves more looks at bottom categories than self-completion mode. Even more interestingly, once a respondent fixates a bottom category the look seems to have a similar latency in face-to-face mode as in self-completion modes. But full perception of a scale in face-to-face mode, and as a result more frequent looks at socially undesirable options, does not explain higher social desirability bias in this mode. An alternative cause, most likely embarrassment, must be present in the face-to-face mode to explain the mode difference.

We would like to caution the reader from simplifying our results in a way that self-completion modes receive socially desirable answers because of satisficing and face-to-face mode because of embarrassment. This is unlikely to reflect the full process leading to socially desirable responses. While we are limited to showing satisficing effects on social desirability bias within self-completion mode, and face-to-face mode featuring less satisficing than self-completion modes, this does not mean that only one source is responsible for social desirability bias in each mode. It is more likely that social desirability bias is influenced by embarrassment and satisficing in both face-to-face and self-completion modes, but this influence may vary in its extent: face-to-face mode is likely to have less social desirability bias due to satisficing but more social desirability bias due to embarrassment in comparison to self-completion modes.

Overall, the results from our study should be treated with caution due to its lack of representation of the general population. The participants were members of a volunteer list for psychological experiments in one university. Although much was done to make the survey situation realistic, the interviews were conducted in a Psychology lab with respondents wearing a helmet with an eye-tracker. This all calls for further research into the topic, and only through consistency of the findings can we extrapolate the results more generally.

Finally, our paper provides an example of real-world eye-tracker usage in the context of survey methodology. This is the first usage of such method in this field, which proved successful and useful, especially in the mode comparison area. Future research can study the use of show cards, and visual attention to an interviewer depending on interviewer behaviour / training in a face-to-face mode; the effect of visual lay-out of questions, and response options and question sets in SAQ mode; and off-monitor attention in web mode. Understanding the causes of differences in attention across modes may prove exceptionally valuable for developing mode-proof survey questions in the future. We envisage and encourage further usage of real-world eye-tracking to improve survey quality.

References

- Anderson, Barbara A. and Brian D. Silver. 1986 "Measurement and Mismeasurement of the Validity of the Self-Reported Vote." *American Journal of Political Science* 30: 771-785.
- Ashenfelter, Kathleen T., and Todd R. Hughes. 2011. "Results from Iterative Usability Testing of the American Community Survey (ACS)." *Paper presented at European Survey Research Association Conference, Lausanne, Switzerland*
- Barton, Allen H. 1958. "Asking the Embarrassing Questions." *Public Opinion Quarterly*. 22: 67-68.
- Belli, Robert F., Michael W. Traugott, and Matthew N Beckmann. 2001. "What Leads to Voting Overreports? Contrasts of Overreporters to Validated Voters and Admitted Nonvoters in the American National Election Studies." *Journal of Official Statistics*. 17: 479-498.
- Federal Statistical Office. 2011. *Methods – Approaches – Developments: Information of the German Federal Statistical Office*. Wiesbaden: Federal Statistical Office.
- Foulsham, T., Walker, E. & Kingstone, A. 2011. "The where, what and when of gaze allocation in the lab and the natural environment." *Vision Research*. 51(17): 1920-1931.
- Fouse, A., N. Weibel, E. Hutchins and J. Hollan. 2011. "ChronoViz: A system for supporting navigation of time-coded data" In *Extended Abstracts of CHI 2011*, SIGCHI Conference on Human Factors in Computing Systems, Vancouver, Canada, May 2011.
- Galesic, Mirta, Roger Tourangeau, Mick P. Couper, Frederick G. Conrad. 2008. "Eye-tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding." *Public Opinion Quarterly*, 72(5): 892–913
- Galesic, Mirta and Ting Yan. 2011. "Use of Eye Tracking for Studying Survey Response Processes." In *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*, eds. Marcel Das, Peter Ester and Lars Kaczmirek. London: Routledge Taylor and Francis Group.
- Goldberg D, Williams P. 1991. *A User's Guide to the General Health Questionnaire*. London: NFER-Nelson.
- Graesser, Arthur C., Zhiqiang Cai, Max M. Louwerse, and Frances Daniel. 2006. "Question Understanding Aid (QUAID): A Web Facility that Tests Question Comprehensibility." *Public Opinion Quarterly*. 70:1–20
- Groves, Robert M., Floyd J. Fowler, Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. Hoboken, New Jersey: John Wiley and Sons, Inc.
- Hayhoe, Mary, and Dana Ballard. 2005. "Eye movements in natural behaviour." *Trends in Cognitive Sciences*. 9(4): 188-194
- Holbrook, Allyson, and Jon. A. Krosnick. 2010. "Social Desirability Bias in Voter Turnout Reports: Tests using the Item Count Technique." *Public Opinion Quarterly* 74:37-67.
- Holgraves, Thomas. 2004. "Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding." *Personality and Social Psychology Bulletin*. 30: 161-172.
- Kamoen, Naomi, Bregje Holleman, Pim Mak, and Ted Sanders. 2011. "Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions." *Paper presented at European Survey Research Association Conference, Lausanne, Switzerland*
- Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology*. 5: 231-236.
- Land M F, Lee D N. 1994. "Where we look when we steer." *Nature (London)* 369, 742-744
- Land, M.F., and P. McLeod. 2000. "From eye movements to actions: how batsmen hit the ball." *Nat. Neurosci*. 3: 1340–1345.

- Lenzner, Timo, Lars Kaczmirek, and Mirta Galesic. 2011a. "Seeing Through the Eyes of the Respondent: An Eye-tracking Study on Survey Question Comprehension." *International Journal of Public Opinion Research*: 23(3): 361-373.
- Lenzner, Timo, Lars Kaczmirek, Mirta Galesic, and Mona Merkert. 2011b. "Left feels right! The optimal position of answer boxes in Web surveys." *Paper presented at European Survey Research Association Conference*, Lausanne, Switzerland
- de Leeuw, Edith D. 2005. "To Mix or Not to Mix Data Collections Modes in Surveys." *Journal of Official Statistics*. 21: 233-255.
- Malvin, Janet H. and Joel Moskowitz. 1983 "Anonymous Versus Identifiable Self-Reports of Adolescent Drug Attitudes, Intentions, and Use." *Public Opinion Quarterly* 47: 557-566.
- Menold, Natalja, Lars Kaczmirek, and Timo Lenzner. 2011. "Cognitive Processes in Answering Questions in Web surveys: How do Respondents Attend to Verbal Labels in Rating Scales?" *Paper presented at European Survey Research Association Conference*, Lausanne, Switzerland
- Peter, Jochen and Patti M. Valkenburg. 2011 "The Impact of 'Forgiving' Introductions on the Reporting of Sensitive Behaviour in Surveys." *Public Opinion Quarterly* 75: 779-787.
- Potaka, Lyn. 2007. "What the Eye doesn't see: A Feasibility Study to Evaluate Eye-Tracking Technology as a Tool for Paper-Based Questionnaire Development." *Proceedings from Quest 2007 Statistics Canada*, Ottawa, Canada: 162-171.
- Rayner, K. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin*. 124: 372-422.
- Redline, Cleo D., and Christopher P. Lankford. 2001. "Eye-Movement Analysis: A New Tool for Evaluating the Design of Visually Administered Instruments (Paper and Web)." *Proceedings of the Survey Research Methods Section of the American Statistical Association*, available at www.amstat.org/Sections/Srms/Proceedings/y2001/Proceed/00248.pdf.
- Romano, Jennifer C. and Jennifer M. Chen. 2011. "A Usability and Eye-tracking Evaluation of Four Versions of the Online National Survey of College Graduates (NSCG): Iteration 2." *Study Series: Survey Methodology 2011-01*. Washington D.C.: U.S. Census Bureau.
- Schaeffer, Nora Cate. 2000. "Asking Questions About Threatening Topics: A Selective Overview." In *The Science of Self-Report: Implications for Research and Practice*, eds. Arthur A. Stone, Jaylan S. Turkkan, Christine A. Bachrach, Jared B. Jobe, Howard S. Kurtzman and Virginia S. Cain. London: Lawrence Erlbaum Associates, Publishers.
- Schwarz, Norbert, Fritz Strack, Hans-J. Hippler, and George Bishop. 1991. "The Impact of Administration Mode on Response Effects in Survey Measurement." *Applied Cognitive Psychology*. 5: 193-212.
- Tourangeau, R., Rips, L. J., & Rasinski, K. A. 2000. *The psychology of survey response*. Cambridge, United Kingdom: Cambridge University Press.
- Tourangeau, Roger and Ting Yan. 2007 "Sensitive Questions in Surveys." *Psychological Bulletin*. 133: 859-883.