Real-world Eye-tracking in Face-to-face and Web Modes

To appear in *Journal of Survey Statistics and Methodology*, 2014

Olena Kaminska*, Tom Foulsham**

* 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

Eye-tracking is becoming a popular tool for understanding how different forms of asking questions influence respondents' answers. Galesic et al. (2008) have successfully shown how primacy effects can be detected through an eye-tracker which measures the time of an eye fixation at different points of a question or response scale. Until now this method has almost exclusively been used to test questions on a computer. Our paper extends the application of eye-tracking to face-to-face mode with show-cards (PAPI). Unlike eye-tracking on a computer screen, tracking eye movements in different modes requires using an innovative real-world eye-tracker which enables following eye movements anywhere the respondent may look. The current article reports on using a real-world eye-tracker to measure visual attention in two modes with visual materials—web and PAPI. We discuss the potential and limitations of the technique, provide successful examples of measuring and comparing visual attention in the two modes and conclude with suggestions for new avenues which can be studied using this new tool.

Key words: eye-tracking, mode effect, primacy effect, show cards.

Acknowledgements: This study was supported by the University of Essex Interdisciplinary Research Fund for Social Sciences "Understanding Survey Responses through Eye-Tracking: Mixed-Mode Comparison" granted in 2011. The authors are grateful to Gemma Luxton for hours of diligent coding of eye-tracking data, and for comments of participants of 2012 RC33 8th International Conference on Social Science Methodology and 2013 European Survey Research Organization conference.

Word count: 6,153

1. Introduction

In general, the survey response process consists of four main stages: comprehension of the question and question instructions, retrieval of relevant information, judgment, and reporting of the survey answer (Tourangeau, Rips, and Rasinski 2000). However, respondents will not always go through the same response process in all conditions. Survey modes have been known to influence survey answers even if the wording of the question, question instruction and response scales are the same (Bowling 2005; de Leeuw 2005).

While mixed mode studies can produce gains in response rate, survey coverage, and cost (Groves et al. 2004; de Leeuw 2005), there is a resistance to using mixed mode studies because of mode effects. Their presence means that different sample members are measured with different quality. Many studies have explored the differences between modes (for overview see Bowling 2005; de Leeuw 2005), and Dillman, Smyth, and Christian (2009) have also explored questionnaire design features which decrease these differences. For example, as part of a 'unimode' design (the design tailored to be suitable for an interview in different modes), Dillman et al. (2009) suggest consistent use of 'don't know' and refusal options, moving from check-all to forced-choice options, using fully labelled response categories, and making the response options the same while incorporating them into the question wording. Nevertheless, some mode differences and their causes remain unexplained, which opens an avenue for innovative survey response testing tools.

To better understand the causes of survey error during the response process a number of methods are used. Cognitive interviewing, when a respondent reports himself how he arrived at an answer, is a common questionnaire pretesting tool. Such techniques as concurrent and retrospective verbal probing and think aloud are used to understand the nature of measurement error (Willis, 2005). Issues with misinterpreting question instructions and question wording, recalling response-relevant information, mapping onto a response scale and visual layout are often revealed in cognitive interviews, and, as a result, a questionnaire is improved before it is fielded. Audio or video recording of survey interviews and subsequent coding of interviewer-respondent communication have also been used successfully in studying interviewer effects on the response process (Jans 2009) or comparing response process in different ways of question administration (Bilgen and Belli 2010). An unobtrusive computer audio-recording (CARI) has also been used to detect field-interview fabrication and has been recently shown to be useful for quickly identifying potential for measurement error (Hicks et al. 2010).

Eye-tracking allows the researcher to record eye movements and identify what the respondent is looking at, which can provide insights into the survey response process. Following early work by Redline and Lankford (2001) and Graesser et al. (2006), and a more recent publication by Galesic et al. (2008), which studied the nature of primacy effects in web mode, and developments in eye-tracking technology have led to a rise in the popularity of eye-tracking as a tool to explore visual attention in surveys (e.g. Ashenfelter and Hughes 2011; Kamoen et al. 2011; Lenzner et al, 2011; Menold et al. 2011; Romano and Chen 2011). Eye-tracking has now received recognition as a successful pre-testing tool, with its current regular use by United States Census Bureau (Ashenfelter and Hughes 2011; Romano and Chen 2011) and the German Federal Statistical Office (Federal Statistical Office, 2011).

To date, due to technology and the ease of using eye-trackers on a computer screen (Galesic and Yan 2011), eye-tracking has been almost exclusively used with web surveys. Two notable exceptions include feasibility studies of eye-tracking in SAQ mode (Redline and Lankford 2001; Potaka 2007). While these studies demonstrated the advantages of tracing eye-movements for SAQ questionnaire development, they lacked a realistic set-up for the interview environment: the paper questionnaire was placed vertically on a clipboard and respondents had to be seated at an unnatural distance from the questionnaire. To our knowledge, ours is the first study to use eye-tracking to study PAPI mode.

We describe our method and a study where we investigated the feasibility of using eye-tracking methodology for untangling differences in response process across survey modes. Recognizing that survey context has the potential to influence respondents' answers (Schwarz, 1996), we pay attention to keeping the survey response environment as close to the real survey situation as possible; and we demonstrate that the new real-world eye-tracking technology enables tracking eye movements without restrictions on respondents' posture or the position of a survey instrument.

The next section provides an overview of the use of eye-tracking in survey methods research, followed by an introduction to real-world eye-tracking. After a short description of the data collection for our study, we report on the successes and challenges of using a real-world eye-tracker in web and PAPI modes (and also in SAQ mode), and provide recommendations for future studies using this technique. We then provide an example of how eye-tracking information can be used to compare response process across modes. Specifically, we explore differences in time spent on question wording and response options and the order of their viewings across modes. We end by discussing the findings and the new avenues of research which can be explored with this technique.

1.1 Eye-tracking in Survey Research

In designing the visual materials for a survey, we often expect that a respondent will make the needed cognitive effort and pay close attention to each word of a question and each scale point. Such expectations may not represent the reality of how respondents read text or perceive questions and response scales.

Previous research in psychology has confirmed that some visual items attract our attention more than others. For example, bright or unusual objects receive longer viewer fixations (Underwood & Foulsham 2006). In the context of text, respondents read progressively, skipping highly predictable words, fixating longer on unusual or low-frequency words, and rereading ambiguous or unfamiliar words (Rayner 2009). Our visual attention may vary, with more words being skipped when subjects admit paying little attention to the task (Reiche et al. 2010). Respondents also pay little or no attention to advertisement banners on web pages, and eye-tracking has been used to study how changing the location and look of web page banners can attract the visual attention of viewers (e.g. Burke et al., 2005).

A consistent pattern of eye-movements has been observed when respondents read survey questions in web mode. Respondents' eye fixations are prolonged by such linguistic characteristics of survey questions as low-frequency words, vague or imprecise relative terms, vague or ambiguous noun phrases, complex syntax, complex logical structures, and low syntactic redundancy (Lenzner et al. 2011a). Negative wording of a question or a response option increases the chance of rereading in comparison to positive wording of the same question or response option (Kamoen et al. 2011). Just as web banners are not attended to very often, in survey research respondents pay little attention to progress indicators, and help links (Ashenfelter and Hughes, 2011).

An important finding from eye-tracking research in surveys was reported by Galesic et al. (2008). In a small scale study, the authors found that around 10 percent of respondents never looked at the last two response options in a 12-category list and that respondents spent more time looking at response options in the first half of a list, resulting in selecting one of these options (i.e., a primacy effect). Across questions, 75-80 percent of participants skipped at least one response option. Understanding how visual attention varies when reading survey questions can help us to design better questions by avoiding word features that increase cognitive effort and by drawing attention to those question features and instructions that are most important.

Eye-tracking is especially useful in exploring how changes in the visual features of a questionnaire may improve respondent’s attention when reading survey questions. A few recent studies have compared eye-movements and timing of eye fixations with different layouts of the same questions. For example, Menold et al. (2011) found that there was no significant difference in overall fixation time between fully-labelled and end-labelled response scales. However, respondents spent longer looking at the rightmost label in end-labelled scales than in fully-labelled scales, where they were more likely to look at all the categories. Another experimental study compared eye-movements when response boxes are placed to the right versus to the left of the response scale options (Lenzner et al. 2011b). The authors discovered that, when the answer boxes are to the left of the response options, there are shorter response latencies, fewer fixations on the answer boxes, and fewer gaze switches between answer boxes and response options.

The only two studies which used eye-tracking with an SAQ questionnaire both explored eye-movements when respondents read Census forms. The study in the US by Redline and Lankford (2001) discovered that the visual design of routing instructions affected the likelihood that respondents would notice these instructions. This article encouraged further experiments which led to an improvement in routing instructions (Redline and Dillman 2002). A study from New Zealand by Potaka (2007) found that respondents notice only the first bubble instruction indicating how to mark answers and skip the later ones. In addition, the author showed that respondents do not look at some of the response options which appeared below alpha-numeric boxes, but are more likely to notice them if they actively search for such response option.

1.2 Real-World Eye Tracking

Recent eye-tracking studies are largely limited to exploring response processes on a computer screen. Tracking eye movements is easier when respondents look at a computer screen or another fixed plane. The lack of eye-tracking studies in SAQ may be explained by the use of an unnatural position of the paper questionnaire in previous studies (Redline and Lankford 2001; Potaka 2007). But just as the authors of the latter studies had hoped, the technology has developed since to allow tracking eye-movements in other modes in a more natural set up. Our study extends the use of eye-tracking to PAPI mode, and we also report on our experience with this technique in SAQ mode where the questionnaire is positioned naturally.

We therefore used an innovative setup with a real-world eye tracker (the SMI HED system) that allows for tracking in a non-lab setting. This equipment monitors eye movements using a camera worn on a headset, while a second camera captures the scene in front of the participant as well as any sounds. Participants are free to move around, their view is not limited to a computer screen, and their gazes can be recorded on anything in their field of vision. Because this field of view changes with each head movement and is not constant as with the computer monitors used in previous studies, additional coding is required to identify what is being looked at in different situations.

Real-world eye-tracking is a relatively new tool that has been used successfully to study vision and attention during such diverse behaviours as walking, driving and playing cricket (Hayhoe and Ballard 2005; Land and Lee 1994; Land and McLeod, 2000). Studying visual attention in these contexts has revealed the coordination and timing involved in real world actions—such as fixating on an object before reaching to pick it up. Moreover, while much can be learned with experiments on a computer screen, visual attention in realistic environments may be different. For example, Foulsham, Walker, and Kingstone (2011) found that people looked at different locations when watching a video of a street as opposed to walking down the same street in the real world. Freeth, Foulsham, and Kingstone (2013), meanwhile, found that attention to a video of an interviewer asking questions was different from when the same questions were asked face-to-face. These differences make it clear that it is important to apply eye tracking methods to a range of realistic survey modes.

1.3 The present study

This study was designed specifically to explore whether real-world eye tracking can be used to study the survey response process in web and PAPI and whether comparable fixation data can be obtained for comparison of visual attention across these modes. We therefore had two main aims.

First, we asked whether using this novel method was feasible and accurate enough for a systematic study of survey responding. This included determining which modes could be successfully tracked and identifying any limitations of this method. Our findings with regard to this question are described in Section 3.

Second, we carried out a detailed examination of the visual attention within the different modes, with the aim of comparing the response process in each case. To date there has been no study which compares the visual attention of respondents across PAPI, and web modes. We know little about whether and when respondents read show cards in a PAPI mode; and how the attention to survey questions and response categories differs across modes. These questions are addressed with data and analysis in Section 4. In particular, we considered the time spent fixating questions and response options and the perception order of response categories.

Do respondents take longer processing the question and rush through response options in one mode but not in another mode? Given previously reported mode effects, we expected that attention to different visual elements might differ. This is particularly interesting in the PAPI mode because there is an important difference here from self-completion modes: while in self-completion modes a respondent cannot perceive question wording simultaneously with response options, in PAPI mode respondents can look at response options while an interviewer is reading the question out loud. For consistency, interviewers in our experiment instructed respondents to turn to the relevant show card before reading the question.

Visual presentation of response scales is known to produce primacy effects (e.g., Krosnick 1991; Galesic et al. 2008), where respondents spend more time reading top options and may skip looking at the options presented at the bottom of the scale. Previous research has found consistent primacy effects when looking at the options chosen in self-completion modes and in face-to-face mode with show cards, whether read aloud by an interviewer or not (Schwarz et al. 1991). We therefore expected that we would find more fixations on the first response items in a list, but we were able to examine for the first time whether this happens equally in both modes. An additional reason why modes may affect responses may be that response options are read in a different order in different modes; the data in our study can be used to test this hypothesis.

2. Method

Twelve participants (seven in web and five in PAPI mode) took part in a small experiment, conducted between January and April 2012, where each respondent was randomly assigned to one of the three modes. Of these respondents17 percent were male, 17 percent were not native English speakers, and all were between 21 and 22 years old.

All participants were interviewed in a Psychology Lab using a real-world eye tracker (the SMI HED system; Sensorimotoric Instruments; Teltow, Germany – see Figure 1). We used the same real-world eye-tracker for all modes, including on the computer screen, in order to obtain comparable measures. Respondents were not restricted in their head movements or instrument position when wearing the real-world eye tracker and were not given any instructions on posture or placement of survey instruments. Importantly, in PAPI mode, interviewers read out loud only the question wording, while the response options were presented only visually on show cards. The eyetracker records eye positions at 50Hz, with a typical gaze position accuracy of 0.5-1 degrees of visual angle. The scene camera has a field-of-view of approximately 50 degrees vertically and 70 degrees horizontally. The system is calibrated by asking the respondent to look at a sequence of points on a wall or desk so that eye fixations can be mapped onto points in the real world.

The participants were volunteers from a pool interested in taking part in Psychology experiments, and were required to have normal vision. Each interview took around 15 minutes, with approximately 2-5 minutes of eye-tracker setup. All respondents were paid £6 for participation.

The survey questions selected for this study were previously administered to panel members of a large national survey panel in two modes, in SAQ and CASI (computer-assisted self-interviewing, similar to web). The question font and layout was not changed in web mode, and show cards (of A4 size, i.e. 210mm by 297mm) for the PAPI questionnaire were developed based on these questions. Additional demographic and personality questions were asked separately in web mode in the end of the experiment.

One section of the questionnaire, encompassing 12 general health questionnaire items (GHQ-12, see Goldberg and Williams 1991), was selected for manual coding. All questions have a four point fully-labelled response scale presented vertically and they were placed in the middle of the interview. Using the custom annotation software ChronoViz (Fouse et al. 2011), one coder noted the beginning and end time of gazes on the question wording and on each response category (or, where this was not possible, whether gaze was on the top two or bottom two categories). From these we calculated the duration of each fixation. Fixation durations used in the analysis were at least 100 milliseconds long due to low reliability of coding shorter fixations. Respondents also looked away from the survey materials (including at the interviewer in PAPI mode or at other objects in the room) which was coded as 'looking at other'. Finally, 9-10 percent of gazes across modes were uncodable. From audio and video recordings the start and end times of the interviewer reading the questions were also recorded.

3. The feasibility of using real world eye-tracking for mode comparison

In this section, we report on real-world eye-tracking in PAPI and web modes and also report on our experience with SAQ mode, which, though less successful, resulted in practical guidance for future studies. Unlike eye-tracking on a computer screen, the data consist of gaze positions overlaid on a video of the participant's field of view. Figure 1 presents example video shots from our study. Despite our progress in using this method, there were a number of difficulties that we describe below and consider in more detail in the discussion.

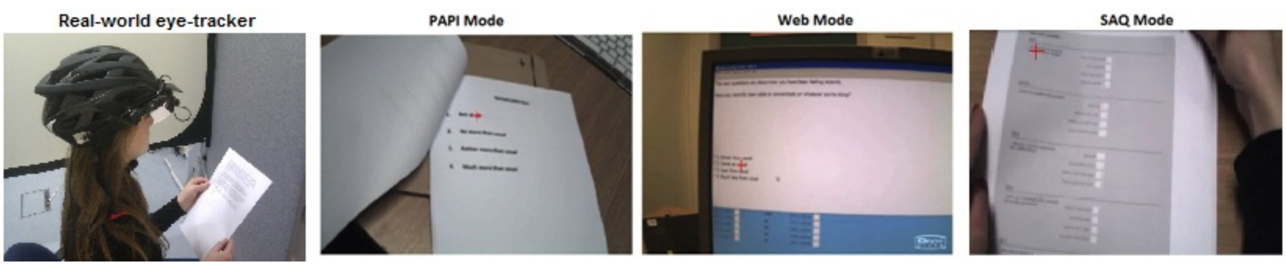

Figure 1. Examples of the real-world eye-tracker and eye-tracking in PAPI, web and SAQ modes

Web was probably the easiest and most consistent mode for tracking eye-movements using the real world eye tracker, even though the lab assistant left the room after calibrating the eye-tracker to ensure privacy. This is because the questionnaire was confined to a limited area in the centre of the equipment's field of view, and also to the occurrence of relatively few head movements. Such movements make recording and coding more difficult. The eye-tracker video also allowed the tracking of gazes away from the computer screen such as at the keyboard, mouse, mobile phone, or any other distractions.

We found tracking eye-movements in a PAPI mode to be the second most successful. Our original aim was to track eye-movements on both the show cards and on the interviewer so that we could differentiate when respondents look at an interviewer or the show card (including which category they look at). Because the eye-tracker has a smaller field-of-view than humans do, and because respondents usually switched gazes between the interviewer and a show card with an eye-movement without moving their heads, it was often not possible to track both simultaneously. Nevertheless, we could successfully set up the eye-tracker for show-cards or the interviewer, separately. Because we are mainly interested in eye-movements while respondents read response options, we tracked eye-movements on the show cards only. When respondents looked away from the show card and higher in the visual field (i.e., beyond the field of view of the camera), we can usually infer whether it was to look at the interviewer (e.g. when a respondent asks for clarification or when the interviewer reads a question).

Our original aim was to also study SAQ mode, which we found to be the most challenging of the three modes. Our aim was to track eye gaze in a natural SAQ interview setting, allowing the respondent full flexibility to hold, place, and fill in the questionnaire in any way that they preferred. We also wanted to keep the situation private by leaving the respondent alone in the room. This prevented us from interfering in order to adjust the eye-tracker. Providing full flexibility during the SAQ interview was challenging with the first six interviews in SAQ mode resulting in only partially codable data. The most common problem was that after the first few questions the quality of tracking decreased sharply due to changes in the position of the questionnaire or equipment. After careful review of videos, we realized that as soon as an interviewer left the room, respondents changed their posture drastically, usually by slouching. This meant that the field-of-view of the eye-tracker scene camera was no longer aligned with the questionnaire. A surprisingly simple solution was found. We asked respondents to start filling in the questionnaire during calibration, allowing respondents to take their natural posture. After the first two questions were completed, the eye-tracker setup was finalized and the interviewer would leave the room. Two interviews were conducted using this approach, both of which resulted in similar quality to other modes (9 percent of fixations were uncodable). Importantly, these interviews were successful in tracking eye gaze without setting any restrictions to respondents and by ensuring privacy as well. Because only two interviews resulted in acceptable quality data, we exclude SAQ mode from the analysis.

Overall, we found that in PAPI and web modes it was possible to measure the visual attention of respondents well enough to differentiate whether they are looking at an instruction, at the question wording, or at the top or bottom part of the response scale.

4. Empirical example: Comparing visual attention between PAPI and web modes

This section provides empirical insights into cognitive response process differences across modes using eye-tracking.

4.1 Analytical approach

From the detailed fixation-level dataset, we derived several measures for each item. First, we defined the "total time" per item as the time from being first exposed to the question materials until making the response (including giving the response). In the PAPI mode, this period began when the interviewer began reading the question, while in the web mode it began when the respondent moved to a new item. The total time per item will consist of reading or listening to the question, processing the options, and deciding on and giving a response. Eyetracking allows one to divide this process further. We therefore defined the "time on question" as the subset of total time when respondents were likely to be processing the question, i.e., while the question is being read aloud in the PAPI mode and while participants are looking at the question wording in web mode. The differences in total time on item and time on question between modes are described in Section 4.2. In Sections 4.3 and 4.4, we consider the "time on options": the period spent looking at the response options (which was evaluated in the same way in web mode and the showcards in PAPI mode). Finally, we discuss the effect of option order, by examining fixations at individual options, in Section 4.5.

Each respondent was administered 12 questions which are treated as separate units of analyses, called 'question-administrations.' In other words, our data has the following structure: each row represents a question-administration, for which we have a value of total time on item, time on question, time on each option, and so on. We compare modes using mixed model ANOVA where mode is a between-subject factor and questions are a repeated measure factor.

4.2 Attention to question wording

Looking at total time, time on question, and time on options (Table 1), one may observe that the difference between modes in mean time is considerably larger than standard errors within modes. Specifically, the difference between PAPI and Web is greater than twice the PAPI standard error for all time measures, except for time on question. For example, the difference in total time spent on each item between PAPI and Web was 2.56 seconds (8780-6220ms) compared to a standard error of 0.415 seconds. This suggests that the sample size in our study may be adequate to detect significant differences in time measures across modes.

Table 1. Average total time, time on question, and time on options, by mode of data collection

	Total time	Time on question	Time on questions read only once	Time on options	Time on options when bottom options viewed

	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
PAPI	8780	415	2872	149	2872	149	4059	284	4158	292
web	6220	463	2796	204	1915	134	2529	229	2665	271
All modes	7287	336	2830	131	2575	120	3154	189	3349	209
n of questions	144		134		87		142		120	
n of respondents	12		12		12		12		12	

*Time is presented in milliseconds

Looking at total time on item, we found that respondents spent significantly longer listening to each question and reading response options from show cards in PAPI mode than they did reading questions and response options in web mode ($F(1,125)=17.31$, $p<0.001$). On average, the mean total time per item in PAPI mode was 8.7 seconds, 2.6 seconds longer than in web. One might think that such a result was expected given that it takes longer to read a question out loud than to read a question silently. However, looking at the time spent on the question wording (in web mode, this included re-readings), PAPI mode was not significantly different from web mode ($F(1,115)=0.01$, $p=0.94$).

Does this mean that respondents read the questions in web mode at a speed similar to an interviewer reading them out loud? Interestingly, the effect appears to be due to the number of re-readings in web mode; we defined re-readings as occasions when the respondent looked at the question wording, then looked at something else (normally the options), before returning to look at the question again. In PAPI mode, interviewers read each question only once. In contrast, in web mode only 37 percent of questions were read just once, with another 39 percent read twice, and the rest read three times or more. In fact, in web mode, some questions were returned to seven times. When we restricted the analysis of time spent on question to question-administrations read only once, we indeed found that web mode have significantly shorter question reading time than interviewer reading time in PAPI mode ($F(1,68)=13.08$, $p=0.001$). Thus, questions were frequently re-read in web mode but the difference in total time between modes cannot be due to the time on question. We therefore expected some differences between modes in the time spent reading response options.

4.3 Attention to response options

Response options were presented visually in both modes. In PAPI mode, the interviewer was instructed to not read the response options out loud but to let respondents read them from the show cards. Each question had four response options displayed vertically. The response options were similar for all the questions, but varied in wording such that only three of the 12 questions had the same response scale as the preceding question. Because the presentation of the scale was visual in each mode, one might expect many similarities in reading the scale across modes. Yet looking at the overall time spent on response options, we found that reading them in a PAPI mode took 1.6 times as long as in web mode ($F(1,123)=21.9$, $p<0.001$).

Next, we considered primacy effects in eye gaze. We found that the odds of not looking at the bottom two response options were over 5 times higher in the web mode than in PAPI mode ($z=-2.59$, $p=0.01$). When we restricted our analysis to question-administrations in which the bottom two response options were looked at least once, time on options was still longer in PAPI mode than in web mode ($F(1,101)=20.2$, $p<0.001$).

4.4 Comparing attention to question wording versus response options

So far, we have noted that participants spent less total time on an item and less time on the response scales in web mode than in PAPI mode. Next, it is of interest to examine whether there is difference across modes in the proportion of time spent on the question wording rather than response options. On average this ratio is 1.07 for web mode and 1.49 for PAPI mode ($F(1,113)=10.79$, $p=0.001$),

Examining fixations on the response options in PAPI mode, we found that for more than a third (37 percent) of the time that the interviewer was reading a question out loud the participants were simultaneously looking at response options. This varied across question-administrations, with 95 percent of question-administrations having fixations on the response options between 26 and 50 percent of the time the interviewer was reading a question. Around half of the time spent looking at response options (48 percent, se=2 percent) occurred while the interviewer was reading a question out loud.

4.5 Order of fixations on the response options

Because the response options were always presented visually, we might expect that they would be read in the same order in both modes. Each question-administration was coded according to whether the first fixation on the uppermost response option was before the first fixation on the next response option (options 1,2); whether the first fixation on the second option was before the first fixation on the third option (options 2,3); and similarly for options 3 and 4 (options 3,4). If a response option was not viewed, the order measure involving this option was set to missing. Question-administrations where we could not differentiate between fixations on specific adjacent options were excluded from this analysis.

Overall, there was similarity across modes in the order of viewings for adjacent response options only for the last two categories. In each mode, the bottom two response options were viewed in sequential order in two-thirds of the question-administrations in which these options were viewed. The difference between modes was larger for the first and middle two categories. In PAPI mode, the top option was viewed before the second top option in only 48 percent of question-administrations, compared to 76 percent in web mode. A less dramatic difference in the same direction was observed for the sequence of first viewing of options 2 and 3.

Next, we explore how many times response options were viewed for the first time in sequential and reverse order and compare these between modes. For some question-administrations, when all options were viewed, they were first fixated on in sequential order. However, this varied drastically between modes: there were only 20 percent of such question-administrations in PAPI mode compared to 44 percent in web mode (Table 2). Only PAPI mode had question-administrations (8 percent) for which all response options were read in reverse order, from bottom to top. It is clear that the tendency to read response options sequentially is higher in web mode than in PAPI mode: around 88 percent of question-administrations in web had one or fewer reverse viewings.

Table 2. Percent of question-administrations with sequential or reversed viewings of adjacent response options

	PAPI	Web	Total
All viewed in sequential order	20%	44%	34%
Mixed viewing	72%	56%	63%
All viewed in reverse order	8%	0%	3%

5. Discussion

This article demonstrates the use of a real-world eye tracker in a survey methods context, allowing the study of visual attention in a natural setting for survey modes such as PAPI, in addition to the previously explored web mode. We started with two general aims, which we will now discuss separately.

5.1 Feasibility, advantages and limitations of real-world eye-tracking across modes

With minor restrictions, in web and PAPI modes, the eye-tracker could track eye-movements wherever a respondent looked, with no limitations on the respondent's head movements or placement of the show cards. For many question-administrations, it was possible to differentiate exactly which response option the respondents were looking at. The main restriction was that participants had to wear the tracking helmet with a small camera and system on it. Importantly, the system has been used in real world situations (e.g. while walking around campus and ordering coffee) and has the potential to be used outside of the Psychology Lab as part of a real survey interview. In addition, the latest systems are considerably less restrictive and fit onto a pair of glasses.

For studying eye-movements in a web mode, the real-world eye-tracker has the important limitation that it requires additional manual coding of the time span of eye-fixations. It is nevertheless a useful tool when the researcher is interested in gazes away from the screen (such as at the mouse or the key board), for example, in studying web responding among novice computer users. Moreover, while the real-world eye-tracker is a good tool for studying visual attention in each mode separately, it has the potential for providing comparable information on visual attention for studies comparing modes.

In PAPI mode, there were some limitations. Specifically, participants preferred to have show cards in front of them on their lap and this required pointing the scene camera down at a sharp angle, making it impossible to capture both looks at the interviewer and at the show cards at the same time. In studies with an interest in gazes at an interviewer, the eye-tracker can be set up to capture views at the interviewer instead.

We also tested real-world eye-tracking in SAQ mode. Because of changes in posture, we found that it was important to calibrate the eye-tracker when participants were already relaxed and had begun filling in the questionnaire. Given that we achieved only two interviews of sufficient quality in this mode, we excluded it from our analysis. The results that we observed based on these two interviews suggest that generally fixations in SAQ mode followed the same pattern as in web mode.

5.2 Insights into the survey response process

As well as demonstrating the feasibility of our approach, this study provides an empirical example of how eye-tracking information can be used for studying the survey response process across modes. Specifically, we were able to compare gazes at the question or response options, as well as the order of reading the response options, across modes. We observed some unique visual behaviour of respondents, not previously discussed in the survey literature. For example, one might expect that listening to a question would take longer than reading the same question. Our analysis of the time spent gazing at the question highlighted three points: first, questions in web mode are not read only once, but are frequently reread after glancing at response options; second, the time spent on question wording is not drastically different when the interviewer reads a question out loud to when a respondent reads a question (including re-readings) in web mode; and third, the difference in total time between modes is only partly due to the difference in time spent on question wording.

Additionally, there were differences in how response options were read. Response options were presented visually in both modes, but we found substantial differences in reading styles of response options between PAPI mode and web mode. For example, it was more common for the lower response options to be read in PAPI mode than in web mode. So, even if a response scale is presented only visually and not read out loud by the interviewer, there is some evidence of differences in how respondents view this scale in the presence of an interviewer in comparison to web mode.

The reading order of response options varied across modes: sequential reading of response options occurred more often in web mode than in PAPI mode. Thus these results suggest that even if response options are presented only visually in PAPI mode, respondents often do not read them sequentially from top to bottom. This is different from web mode for which sequential reading of response options is more common.

Our empirical results also showed evidence of simultaneous consideration of response options and question wording in both modes: in PAPI mode, it was not uncommon for respondents to read response options while an interviewer was reading the question out loud; and, in web mode, many questions were re-read between glances at the response options. This suggests that in a PAPI mode at least some of the time respondents perceive the question wording simultaneously with the response options. It is unclear whether this may be optimal for the perception of the question and the response scale, given that simultaneous visual and aural perception may be natural to humans. One may think that this potentially may lead to differences in the perception of the question and response options across modes, given that some response options are read before the question is completely heard in PAPI mode. That the web mode produces s a number of re-readings of the question wording in between looks at the response options suggests similar simultaneous – or at least interleaved - processing of the question wording and the response scale.

The extent to which these gaze differences may influence perception of the question and scale options is not yet clear. But it is clear that when designing questions, we cannot assume they will be read sequentially, starting with the question wording, and following with each response option in the order in which they are presented in the questionnaire or on the show card. Neither can we assume that the perception order of question and response options is similar across modes.

Given the fast movement of survey data collection to mixed modes, we need to better understand the potential influence of modes on responses. In a search for mode-resistant questions or for questions that are comparable across modes, a clear understanding needs to be developed of how each feature of a question may affect perception and attention, and through these, potentially affect responses. Real-world eye-tracking permits such an investigation.

References

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

Bilgen, Ipek, and Belli Robert F. (2010) Comparison of Verbal Behaviors between Calendar and Standardized Conventional Questionnaires, *Journal of Official Statistics*. 26(3): 481–505

Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health, 27*(3), 281-291.

Burke, Moira, Anthony Hornof, Erik Nilsen, and Nicholas Gorman. 2005. High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. *ACM Trans. Comput.-Hum. Interact*. 12, 4, 423-445.

Dillman, Don A., Smyth Jolene D., Christian Leah M (2009). *Internet, Mail, and Mixed-mode Surveys: The Tailored Design Method*. Hoboken, NJ: Wiley

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.

Freeth, Megan, Tom Foulsham, and Alan Kingstone. (2013). What affects social attention? Social presence, eye contact and autistic traits. PLoS ONE, 8(1).

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., Fowler, F. J., Couper, Mick P., Lepkowski, James M., Singer, Eleanor, & Tourangeau, Roger (2004). *Survey methodology*. Hoboken, NJ: John Wiley and Sons.

Hayhoe, Mary, and Dana Ballard (2005) Eye movements in natural behaviour. *Trends in Cognitive Sciences*. 9(4): 188-194

Hicks, Wendy D., Brad Edwards, Karen Tourangeau, Brett McBride, Lauren D. Harris-Kojetin, Abigail J. Moss (2010).Using Cari Tools To Understand Measurement Error. *Public Opinion Quarterly,* 74 (5): 985-1003.

Jans, Matt (2009). *Verbal Paradata and Survey Error: Respondent Speech, Voice, and Question-answering Behavior can Predict Income Item Nonresponse*. Unpublished dissertation, Program in Survey Methodology, University of Michigan

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 (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*, 5, pp.213-236.

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

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.

Land,M.F.,Mennie,N., and Rusted,J.(1999).The roles of vision and eye movements in the control of activities of daily living. *Perception*. 28(11): 1311-1328.

de Leeuw, E. D. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21*(2), 233-255.

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

Ongena, Yfke (2010). *Interviewer and Respondent Interaction in Survey Interviews: Empirical evidence from behaviour coding studies and question wording experiments*. LAP Lambert Academic Publishing

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.

Schwarz, N., Strack, F., Hippler, H., & Bishop, G. (1991). The impact of administration mode on response effects in survey measurement. *Applied Cognitive Psychology, 5*, 193-212.

Rayner K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology,* 62 (8), 1457-1506

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.

Redline, D. Cleo, and Don A. Dillman, (2002). The Influence of Alternative Visual Designs on Respondents' Performance with Branching Instructions in Self-Administered Questionnaires. In R. M. Groves, D. A. Dillman, J. L. Eltinge & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 179-196). New York: John Wiley and Sons, Inc.

Reichle, E., D., Reineberg, A.E. and Schooler, J. K. (2010). Eye Movements During Mindless Reading. *Psychological Science*, 21: 1300-1310

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.

Schwarz, Norbert. *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation.* Mahwah: New Jersey: Lawrence Erlbaum Associates, Publishers.

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

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *QuarterlyJournal of Experimental Psychology:* 59 (11), 1931 - 1949.

Willis, Gordon B. (2005). *Cognitive Interviewing: A Tool for Improving Questionnaire Design.* Thousand Oaks, CA: Sage Publications, Inc.