

WHY ARE NEGATIVE QUESTIONS DIFFICULT TO ANSWER? ON THE PROCESSING OF LINGUISTIC CONTRASTS IN SURVEYS

NAOMI KAMOEN*
BREGJE HOLLEMAN
PIM MAK
TED SANDERS
HUUB VAN DEN BERGH

Abstract Previous studies show that respondents are generally more likely to disagree with negative survey questions (e.g., *This is a bad book. Yes/No*) than to agree with positive ones (e.g., *This a good book. Yes/No*). In the current research, we related this effect to the cognitive processes underlying question answering. Using eye-tracking, we show that during the initial reading of the question, negative evaluative terms (e.g., *bad*) require more processing time than their positive counterparts (e.g., *good*). In addition to these small differences in the initial stages of question answering, large processing differences occur later in the question answering process: Negative questions are reread longer and more often than their positive counterparts. This is particularly true when respondents answer *no* rather than *yes* to negative questions. Hence, wording effects for contrastive questions probably occur because response categories such as *Yes* and *No* do not carry an absolute meaning, but are given meaning relative to the evaluative term in the question (e.g., *good/bad*). As answering *no* to negative questions requires more processing effort in particular, a likely explanation for the occurrence of the wording effect is that *no* answers to a negative question convey a

NAOMI KAMOEN is an assistant professor in the Department of Communication and Information Sciences at Tilburg University, Tilburg, the Netherlands. BREGJE HOLLEMAN and PIM MAK are assistant professors in the Department of Languages, Literature, and Communication at Utrecht University, Utrecht, the Netherlands. TED SANDERS and HUUB VAN DEN BERGH are full professors in the Department of Languages, Literature, and Communication at Utrecht University, Utrecht, the Netherlands. *Address correspondence to Naomi Kamoen, Department of Communication and Information Sciences, PO Box 90153, 5000 LE Tilburg, the Netherlands, +31 13 466 4249; e-mail: N.Kamoen@uvt.nl.

doi:10.1093/poq/nfx010

Advance Access publication June 6, 2017

© The Author 2017. Published by Oxford University Press on behalf of the American Association for Public Opinion Research.

All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

mitigated meaning. The activation of this additional pragmatic meaning causes additional processing effort and also causes respondents to pick a *no* answer to negative questions relatively easily.

An important advice in survey research is to balance positive and negative wordings (e.g., Churchill 1979; Sudman and Bradburn 1982). Theoretical constructs are therefore often measured with a mix of positive items (e.g., *I like the prime minister. Yes/No*) and negative items (e.g., *I dislike the minister for foreign affairs. Yes/No*). One reason for adopting this strategy is to “alert inattentive respondents that item content varies” (Swain, Weathers, and Niedrich 2008, 116). Another reason is to mitigate the effect of respondents choosing the same answer for all items (Weisberg 2005); when positive and negative items are mixed, such “straightliners” will not lead to an extreme score across items.

Whereas mixing contrastive wordings is done to avoid biases on the survey level, a vast body of experimental evidence shows that the choice for a positive or negative wording is relevant for individual questions: Respondents are more likely to answer *no* to *negative* questions than to answer *yes* to *positive* ones. In other words, respondents express a more favorable attitude toward the attitude object when the question is formulated using words with a negative appraisal (for an overview of previous research, see Kamoen, Holleman, and Van den Bergh 2013). This effect can be generalized across experimental settings (Holleman 1999a) and across different word pairs (Kamoen, Holleman, and van den Bergh 2013), even though the effect does not occur for each individual question in each individual study (Bishop et al. 1988). In addition, the effect of question wording has proven to be stable across different types of response scales, such as two-point-scale yes/no questions and five-point-scale agree/disagree questions (Holleman 2000).

The fact that question wording affects survey answers indicates that current survey advice to mix positive and negative wordings also has unintended consequences. To learn more about which question wording is to be preferred, insight is required into the validity of contrastive questions. Recently, research has been undertaken to gain a better understanding of the validity of positive and negative questions by relating them to the cognitive processes underlying question answering (Holleman 1999b; Chessa and Holleman 2007; Kamoen et al. 2011).

In aforementioned studies, the Tourangeau model for question answering processes is used as a starting point (Tourangeau, Rips, and Rasinski 2000). In a simplified version of this model, two stages are distinguished: a comprehension-retrieval stage and a mapping stage (Chessa and Holleman 2007). The question answering process starts with the comprehension-retrieval stage, during which a respondent follows three steps in order to form an opinion about the attitude object in the question. First, the respondent interprets the

question, which means that a representation of the question is made and that the respondent determines what opinion or attitude is being asked for. Second, the respondent retrieves relevant attitudinal information from long-term memory. Some respondents will retrieve a summary evaluation of their beliefs directly (*smoking is despicable*). Others will have to take an additional third step of weighting and scaling individual, and sometimes contrasting, beliefs to reach a judgment (*smoking is bad for my health* versus *smoking relaxes me*). Either way, in the second, larger stage of the model, the mapping stage, the respondent fits his or her judgment to the answering options in the question. Respondents may adapt their answers during this process, for example, for reasons of social desirability.

Kamoen et al. (2011) conducted an eye-tracking study to relate wording effects for contrastive questions to these two larger cognitive processes in the Tourangeau model. In this study, the time and occurrence of reading the question and the answers were measured and compared between positively and negatively worded questions. Their results showed no differences in the time needed to read positive and negative questions for the first time, nor during rereading. The only processing difference was related to the occurrence of rereading: The question and answer options are reread more frequently for negative questions than for positive ones. Based on these results, the authors concluded that wording effects arise late in the question answering process and only the mapping process is affected by the choice of wording. This suggests that respondents activate the same judgment about the attitude object for positive and negative questions during the comprehension-retrieval stage. Therefore, positive and negative questions can be considered equally valid: They measure the same underlying attitude. In the mapping stage, this same judgment is translated differently to the response options depending on the choice of wording. Hence, wording effects probably arise because the meaning of the answering options differs for positive and negative questions: Even though response options such as *yes* and *no* and evaluative terms such as *allow* and *forbid* are straight opposites, the meaning of *yes/no* as an answer to a positive question is not identical to the answer *no/yes* to a negative question. This conclusion matches the results of a reaction time study (Chessa and Holleman 2007) and an analysis of the statistical congenericity of positive and negative questions (Holleman 2006).

While previous research indicates that wording effects for contrastive questions arise in the mapping stage, a more stringent test should be conducted to examine the impact of question polarity on the comprehension-retrieval stage once again, because in previous studies only one, rather coarse-grained, measure was used for this stage: The reading time for the entire question. Small effects on more fine-grained regions may therefore have gone unnoticed.

There are at least two reasons why processing effects may be expected to arise in the initial stage of question answering after all. First, in natural

language use, words with a negative appraisal are often longer in terms of word length and less frequent in terms of usage (Horn 1989). As both longer words and less frequent words are associated with increased processing times (Rayner and Duffy 1986; Pollatsek et al. 2008), these superficial characteristics may be the cause of longer processing times for negative questions. Based on previous research, we therefore expect negative evaluative terms to receive more processing time than their positive counterparts. These effects are expected to be local, that is, bound to the manipulated evaluative term.

A second hypothesis regarding processing effects in the comprehension-retrieval stage can be derived from linguistic research. Outside the domain of surveys, several reaction time studies have shown that terms with a negative appraisal cause more processing time than their positive counterparts (e.g., Hoosain 1973; Sherman 1973; Clark 1976). Clark (1971: 503) explains these effects as follows: "(...) certain inherently negative lexical items are represented in immediate memory, not as unitary wholes, but rather as complexes containing at least two components, a positive core and an embedding negative." Hence, in order to comprehend a term such as "unhappy," this term has to be converted into the format "not happy." Relative to the word length and word frequency account, this cognitive load account predicts more widespread processing differences between positive versus negative questions: An extra mental operation is involved in interpreting negative terms. This extra operation may (1) be more difficult to perform when the respondent already has other information in his short-term memory and (2) it also may cause other question parts after the evaluative term to be affected by the question wording. This latter point is important from a survey methodological perspective because in such instances, the conclusion that positive and negative questions are equally valid should be reconsidered.

To distinguish between the two different causes for a potential processing effect in the early stages of question answering, we set up an eye-tracking study in which we varied not only the question polarity, but also the word order in the question. If there is an effect caused by superficial characteristics such as word length and word frequency, we expect the evaluative term to receive longer fixation times when the question is worded negatively, and the size of this effect should not depend on the position of the evaluative term in the question. However, if negative terms cause an increased cognitive load, we expect processing differences for positive and negative wordings to increase when these terms are positioned at the end of the sentence, when a respondent's cognitive load is already high. In addition, in such cases we may also find effects on other parts of the question, such as on the attitude object.

Other than investigating the comprehension-retrieval stage for positive versus negative questions in a more fine-grained manner, a second aim of the current research is to explore what happens during the mapping stage. Previous research shows that mapping an answer to the response options takes more processing effort for negative questions than for positive ones (Kamoen et al. 2011).

We will explore in the current research whether this additional effort can be related to responding to negative questions in general, or whether it is conditional to the answer a respondent provides (either *yes* or *no*).

Methods

MATERIALS AND EXPERIMENTAL DESIGN

To disentangle our hypotheses related to the comprehension-retrieval stage, we constructed four versions of a survey on environmental issues. These survey versions included 38 questions that were manipulated according to a two (question polarity: positive or negative) by two (ordering of the question: evaluative term at the beginning of the sentence or at the end) design, and 75 filler items that had an identical wording across the four survey versions. The full list of experimental materials is available via Dataverse (<https://dataverse.nl/dataset.xhtml?persistentId=hdl:10411/20857>).

As for the manipulation of position, we refer to example questions (1) and (2). In Dutch, the language in which the experiment was written, survey questions like (1) often occur. Such questions can be divided into four parts that always have the same sequence: a proposed subject (*it is*), an evaluative term (*allowable*), an indicator of opinion (*in my opinion*), and an attitude object (*if stores sell cosmetics that have been tested on animals*). Questions like (1), in which the evaluative term precedes the attitude object (in short, ET-AO questions), can also be reversed (see [Van Berkum et al. 2009](#) for a similar design). This results in questions comparable to (2), in which the attitude object (AO) precedes the evaluative term (ET), in short, AO-ET questions. The ordering of the critical regions in such questions is attitude object (*if stores sell cosmetics that have been tested on animals*), construction with a postponed subject in it (*that is*), evaluative term (*allowable*), and indicator of opinion (*in my opinion*).

1. Het is/aanvaardbaar (onaanvaardbaar)/naar mijn mening/als winkels cosmetica verkopen die op proefdieren is getest.
It is/allowable (unallowable)/in my opinion/if stores sell cosmetics that have been tested on animals.
2. Als winkels cosmetica verkopen die op proefdieren is getest/is dat/aanvaardbaar (onaanvaardbaar)/naar mijn mening.
If stores sell cosmetics that have been tested on animals/that is/allowable (unallowable)/in my opinion.

Besides the manipulation of the question ordering, we also systematically varied the linguistic polarity: Questions were either positively worded like (1) and (2), or phrased negatively (negative wording between brackets in the example questions). [Appendix 1](#) provides an overview of the positive and negative evaluative

terms used for the manipulation of polarity. As can be read from the appendix, positive poles were always words that have a positive connotation or valence (see [Moors et al. 2013](#); [Warriner, Kuperman, and Brysbaert 2013](#)). Negative poles, on the other hand, have a negative connotation, and these were always implicit rather than explicit linguistic contrasts (e.g., *forbid/allow* instead of *not allow/allow*). We controlled for the length of these words, as this factor may be an alternative explanation for fixation time differences between linguistic contrasts ([Rayner and Duffy 1986](#)). Word frequency, however, appeared to be hard to control for, and therefore the word frequency of the positive terms is higher on average.

Using these experimental materials, we can distinguish between word frequency effects and effects of cognitive load. The cognitive load account predicts larger effects of question polarity when the respondent already has prior information in his working memory, whereas the word frequency account predicts the polarity effect to be stable across different sentence positions. Hence, the word frequency account predicts a main effect of question polarity, whereas the cognitive load account predicts an interaction between question polarity and the position of the evaluative term in the question: The effect of question polarity is expected to be larger when the evaluative term is placed at the end of the question. In addition, we will explore what happens during the mapping stage by comparing processing patterns between respondents answering *yes* or *no* to positive and negative questions, irrespective of the position of the evaluative term in the question.

PARTICIPANTS

For this study, we draw on a convenience sample of Humanities students who could voluntarily enroll for participation in an eye-tracking study ($N = 122$). All respondents but one were between 16 and 30 years, and 90 percent of respondents were female.¹

PROCEDURE

Respondents were invited to the laboratory for individual sessions. They were told that they were about to answer attitude questions on environmental issues. A Tobii 1750 remote eye-tracker was used to record respondents' eye movements every 20 milliseconds (50 Hz). The hardware of the Tobii eye-tracker looks exactly the same as an ordinary computer monitor. To link the position of the respondents' eyes to the position on the screen they look at, a calibration procedure is needed. When calibration was completed successfully, the experimenter started the survey. Respondents completed the survey in approximately

1. We advertised for our experiment in several classes in Communication Studies, which might explain the high percentage of females. Due to the random assignment of participants to conditions, this did not affect the internal validity of our study. We cannot estimate the impact on the external validity of our results because gender differences are task dependent ([Rupp and Wallen 2007](#)).

20 minutes, and they were paid 5 Euros for their participation. Data collection was completed between September 10, 2010 and November 2, 2010.

THE SURVEY

The survey started with an introduction to the topic of the survey: environmental issues. The respondents were instructed to answer the questions about this topic truthfully, and select the answer that best matched their opinion. They were told that an asterisk (*) would appear on the screen before each question. Respondents were asked to look at the asterisk and to click on it. By clicking on the asterisk, the first word of the question would appear exactly in that spot. This was done to ascertain that respondents started reading the question at the first word.

The actual survey started right after the introductory text. The questions were presented one by one to the respondents, each on a separate screen. We made sure that a question always fitted one line on the screen, to avoid searching fixations when respondents proceeded to the next line.

The questions in the survey were clustered based on their content (see [table 1](#)). Four clusters contained mainly manipulated questions, which were always associated with a two-point response scale. Besides manipulated questions, these clusters contained about five filler questions about the same topic, which had a different kind of response scale (a four-point agree-disagree scale, a four-point often-never scale, and a four-point often-never scale including a fifth “don’t know” option).

Next to the four clusters with mainly manipulated questions, the survey contained three clusters consisting of only fillers. The dominant type of response

Table 1. Characteristics of the Clusters of Questions

Topic	Dominant scale type	N manipulated questions	N filler questions
Clusters with manipulated questions			
Biological nutrition	2-point scale	10	4
Energy awareness	2-point scale	8	11
Landscape design	2-point scale	11	4
Separation of waste	2-point scale	9	5
Clusters with only filler questions			
Animal testing	4-point scale		14
Genetic manipulation	4-point scale		15
Car use	4-point scale		14

NOTE.—In the actual survey, each cluster with predominantly manipulated questions was followed by a cluster with only filler questions. In addition to the 67 filler questions in table 1, there were also eight filler questions related to demographic variables.

scale in these clusters was always a four-point agree-disagree scale, mixed with some two-point-scale *yes/no* questions, some four-point-scale often-never questions, and some four-point-scale often-never questions including a fifth “don’t know” option.

Using this setup, the type of response scale was mixed throughout the survey to ensure that the respondent was not aware of the response options until the answers were fixated. For the same reason, we made sure that the question wording did not reveal the type of response scale used. In addition, although in reading texts there is no indication that readers acquire useful semantic information from the line in the text below the line fixated (Rayner 1998, 380), we decreased the chance that readers were aware of the response options by placing them far below the question (2.5 centimeters), and we made sure that the different response scales covered an identical range on the screen so that the different scale types were hard to distinguish as long as the question was fixated. Because of these measures, all fixations on the question are likely to reflect comprehension-retrieval processes until the respondent fixates the answers; it is impossible to start mapping an opinion onto the response options when the response options are unknown.

EYE-TRACKING MEASURES

To analyze the eye-tracking data,² each webpage was divided into five regions: the subject construction, the evaluative term, the indicator of opinion, the attitude object, and the answers. For the first four regions, which concern sub-parts of the question, the total reading time was computed until the respondent viewed the response options for the first time. As has been explained earlier, these measures are likely to reflect comprehension-retrieval processes.

Next, the total reading time for the answers was registered up to the point where the respondent either responded or switched back to the question for rereading. Because not all respondents are actually involved in rereading, we estimated the initial reading time for the answers as well as the initial reading times for the question-bound regions separately for so-called single-reading trials, in which the respondent gives an answer when he sees the answering options for the first time, and rereading trials, in which the question and the answers are reread before an answer is given.

In terms of cognitive processes, the mapping process can start from the moment the respondent sees the response options onward. This does not mean that all comprehension-retrieval activity then finishes abruptly. However, because this survey contains relatively short questions, and because the respondents are relatively skilled readers, the main cognitive activity involved is likely to be mapping an answer to the response options (compare Kamoen et al. 2011).

2. Fixations were determined by an algorithm that restricts fixations to data points within 30 pixels. The viewing time in a region was computed as the time from the beginning of the first fixation until the end of the last successive fixation in a region. A minimal fixation length of 50 ms was allowed in the analysis.

In addition to the initial reading time of the answers, we included the remaining total fixation time for all five regions as additional measures for the mapping process. Figure 1 provides an example of how this translates to concrete experimental material.

As we observed that respondents sometimes skip a question region, we registered dichotomously for each trial whether the region was viewed during the initial reading as well as during rereading. Hence, in addition to the processing time measures that are registered if and only if the respondent has looked at a specific region at all, we also registered dichotomously whether each of the 5 (region) x 2 (initial reading or rereading) eye-tracking measures displayed a non-zero fixation time. The distinction between reading times and these “reading occurrences” is important, as Kamoen et al. (2011) showed that positive and negative items are reread for the same duration, that is, once rereading occurs, whereas the process of rereading happens more often for negative trials. Hence, by analyzing both times and occurrences, a distinction can be made between regions that are fixated longer and regions that are fixated more often. As previous research showed that large differences between positive and negative questions are observed for the rereading occurrences, we will use these measures to explore the rereading occurrences for respondents answering *yes* versus *no* to positive and negative questions.

STATISTICAL ANALYSES

As in our study each respondent reacted to 38 manipulated items, the answers, the processing times, and the reading occurrences vary both between items, between respondents, and due to the interaction between respondent and item. These data were therefore analyzed using mixed models, the estimations and equations for which appear in online appendices 1–3.

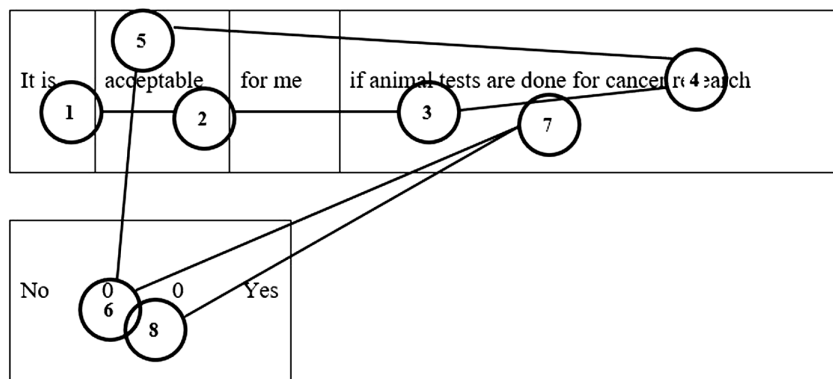


Figure 1. Fictitious Example of Eye-Movement Pattern.

Results

DATA AND MANIPULATION CHECKS

Respondents in the survey versions were found to be comparable with respect to their age ($\chi^2 = 6.33$; $df = 9$; $p = 0.70$), gender ($\chi^2 = 2.71$; $df = 3$; $p = 0.44$), their answers to the filler question, and the processing times for the filler questions (in all cases: $\chi^2 < 3.28$; $df = 1$; $p > 0.07$). Therefore, randomization problems or sampling error are unlikely to account for differences on any of the dependent measures.

The quality of the eye-tracking data was also checked before analyzing the data. The eye-tracking data of 12 respondents (8 percent) were of poor quality and therefore not taken into account. An additional two respondent and item combinations (< 1 percent) were discarded because they displayed unusually low fixation times. In these cases, respondents probably accidentally clicked on an answer before even reading the question. When these trials were removed, assumptions of normality were met for the processing time data after applying a log transformation (compare Yan and Tourangeau 2008).

WORDING EFFECTS FOR CONTRASTIVE QUESTIONS

Table 2 presents an overview of the mean answers given to positive and negative questions. Analyses show a main effect of question polarity: Respondents express their opinions more positively when the question is worded negatively ($\chi^2 = 19.74$; $df = 1$; $p < 0.001$; Cohen's $d = 0.40^3$). Hence, when generalizing over a substantial set of questions with a large variety in types of linguistic contrasts, the existence of wording effects for contrastive questions is shown again. In terms of percentages, respondents are 5.5 percent more likely to answer *no* to negative questions than *yes* to positive ones.

Processing Times

GENERAL PATTERN OF PROCESSING TIMES: SINGLE-READING TRIALS VERSUS REREADING TRIALS⁴

As in some trials respondents immediately give an answer upon seeing the response options (single-reading trials), whereas in other trials respondents reread parts of the question before providing an answer (rereading trials), we

3. Cohen's d is a measure for the effect size relative to the standard deviation. In multilevel models, there are several variance components, and hence the effect size can be calculated relative to each of those, or to a combination of each of those. We calculated the effect size relative to the sum of the between-item and between-person variance.

4. We first performed multivariate tests for the main effect of question polarity, the main effect of position of the evaluative term, and the interaction polarity*position. Individual tests for the separate dependent measures in the model, for example, the subject construction, the evaluative term, and so forth, were performed only if the multivariate test showed an overall main effect. This was done to reduce the risk of a Type I error.

Table 2. Mean Answers to the Positive and Negative Questions with the Evaluative Term at the Beginning and the Attitude Object at the End (ET-AO), and to Positive and Negative Questions with the Attitude Object at the Beginning and the Evaluative Term at the End (AO-ET)

	Proportion	Positive		Proportion	Negative	
		Logit	(SE)		Logit	(SE)
Means						
ET-AO	0.54**	0.18	(0.18)	0.59	0.37	(0.18)
AO-ET	0.55**	0.22	(0.18)	0.61	0.45	(0.18)
Variances						
S ² questions		1.06	(0.25)		1.06	(0.25)
S ² persons		0.07	(0.03)		0.07	(0.03)

NOTE.—For the sake of convenience, the mean answers are given in proportions and in the logits used for the analysis. A higher proportion represents a more positive evaluation of the attitude object. The variances and standard errors are only provided in logits.

** $p < 0.001$ for the difference between positive and negative questions.

first checked if there are general differences in reading times for these different response strategies. As can be read from [table 3](#), no differences related to the initial reading time of the question occur (including these parameters did not improve the model fit: $\chi^2 = 23.74$; $df = 16$; $p = 0.11$). Hence, the fixation times for the subject construction, the evaluative term, the indicator of opinion, and the attitude object do not depend on whether rereading is eventually involved.

By contrast, for the initial reading of the response options, a difference between the single-reading trials and the rereading trials can be observed. Results show that the reading times for the response options are almost twice as long (e.g., 757 versus 413 ms for positive ET-AO questions) when there is no rereading involved (in all cases: $z = 15.75$; $p < .001$). This finding is not surprising, as for single-reading trials the entire mapping process is captured in this measure.

Polarity effects: More important is the effect of question polarity. For both single-reading trials and rereading trials, an effect of question polarity is found for the evaluative term during the initial reading: negative terms are fixated longer than positive ones ($\chi^2 = 47.83$; $df = 1$; $p < 0.001$; Cohen's $d = 1.40$). Crucially, this effect does not depend on the position of the evaluative term in the question.⁵ The other measures related to the initial reading of the question

5. If the manipulation of position is strong enough to disentangle the word frequency and cognitive load effects, we would expect polarity effects for the initial reading of the evaluative term to increase when the attitude object includes more characters. The length of the attitude object in our experimental materials ranged between 37 and 75 characters. In a separate multilevel model, we established that polarity effects for the initial reading of the evaluative term indeed increase when the attitude object contains more characters, but nevertheless both for longer and shorter attitude objects only a main effect of question polarity occurs.

Table 3. Average Processing Times for the Positive and Negative ET-AO and AO-ET Questions

	Subject construction		Evaluative term (ET)		Indicator of opinion		Attitude object (AO)		Answers	
	Ms	Log (ln) (SE)	Ms	Log (ln) (SE)	Ms	Log (ln) (SE)	Ms	Log (ln) (SE)	Ms	Log (ln) (SE)
<i>Means</i>										
	<i>Fixation times during the initial reading for the rereading trials</i>									
Positive ET-AO	254*	5.54 (0.04)	384**	5.95 (0.04)	361*	5.89 (0.04)	2,091*	7.65 (0.03)	413	6.02 (0.04)
Negative ET-AO	262	5.57 (0.04)	471	6.15 (0.04)	348	5.85 (0.04)	2,169	7.68 (0.03)	386	5.94 (0.04)
Positive AO-ET	220	5.40 (0.04)	368	5.91 (0.04)	213	5.36 (0.04)	2,248	7.72 (0.03)	413	6.03 (0.04)
Negative AO-ET	233	5.45 (0.04)	412	6.02 (0.04)	225	5.42 (0.04)	2,245	7.72 (0.03)	413	6.02 (0.04)
	<i>Fixation times during the initial reading for the single-reading trials</i>									
Positive ET-AO	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	757**	6.63 (0.04)
Negative ET-AO	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	781	6.66 (0.04)
Positive AO-ET	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	796	6.68 (0.04)
Negative AO-ET	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	871	6.77 (0.04)
	<i>Fixation times during rereading (only for the rereading trials)</i>									
Positive ET-AO	164*	5.10 (0.07)	236**	5.47 (0.05)	218	5.39 (0.06)	537*	6.29 (0.07)	434*	6.07 (0.05)
Negative ET-AO	195	5.27 (0.06)	289	5.67 (0.04)	219	5.39 (0.05)	624	6.44 (0.06)	410	6.02 (0.04)
Positive AO-ET	178	5.18 (0.07)	283	5.65 (0.06)	232	5.45 (0.07)	468	6.15 (0.05)	498	6.21 (0.05)
Negative AO-ET	192	5.26 (0.07)	298	5.70 (0.05)	190	5.25 (0.06)	457	6.13 (0.05)	464	6.14 (0.05)

Continued

Table 3. Continued

	Subject construction		Evaluative term (ET)		Indicator of opinion		Attitude object (AO)		Answers	
	Ms	Log (ln) (SE)	Ms	Log (ln) (SE)	Ms	Log (ln) (SE)	Ms	Log (ln) (SE)	Ms	Log (ln) (SE)
Variances										
S ² questions	0.01	(0.00)	0.01	(0.00)	0.01	(0.00)	0.01	(0.00)	0.01	(0.00)
S ² persons	0.04	(0.04)	0.04	(0.04)	0.04	(0.04)	0.04	(0.04)	0.04	(0.04)
S ² residual initial reading	0.45	(0.01)	0.51	(0.01)	0.50	(0.01)	0.26	(0.01)	0.40	(0.01)
S ² residual rereading	0.47	(0.03)	0.61	(0.02)	0.48	(0.03)	1.24	(0.04)	0.73	(0.02)

NOTE.—The mean processing times are given in milliseconds, and in the natural logs (ln) used for the analysis; the variances and SEs are only provided in ln. The number of trials on which the parameter estimates are based can be found in online appendix 4.

* $p < 0.05$ for the main effect of polarity; $\times p < 0.05$ for the main effect of position (the effect of the position of the evaluative term can be summarized as follows: Sentence chunks that are positioned at the beginning of the sentence are read longer initially than sentence chunks that are positioned at the end of the sentence. In rereading, this pattern reverses); $\mp p < 0.05$ for the difference between the rereading and the single-reading trials.

show no differences between positive and negative wordings (always $\chi^2 < 2.13$; $df = 1$; $p > 0.14$).

For the initial reading of the response options, there also is a difference between positive and negative questions: Negative questions receive a longer response time than positive ones. However, this effect occurs only for single-reading trials in which the respondent gives an answer upon first viewing the response options ($\chi^2 = 5.38$; $df = 1$; $p = 0.02$; Cohen's $d = 0.54$) and not for rereading trials ($\chi^2 = 2.19$; $df = 1$; $p = .14$).

When the question is reread, respondents fixate the subject construction and the evaluative term longer for negative questions ($\chi^2 = 4.30$; $df = 1$; $p = 0.04$; Cohen's $d = 1.14$; $\chi^2 = 8.55$; $df = 1$; $p < 0.01$; Cohen's $d = 1.14$, respectively); there are no differences in the time needed to reread the other parts of the question ($\chi^2 < 3.12$; $df = 1$; $p > 0.08$). Also, the answers are reread equally long for contrastive questions ($\chi^2 = 2.68$; $df = 1$; $p = 0.10$).

ROBUSTNESS CHECK: THE ROLE OF WORD FREQUENCY AND WORD LENGTH

Following the line of reasoning presented in the Methods section, it is plausible to attribute the processing effects for the evaluative term to differences in word frequency. To put this explanation to a stricter test, we ran several additional analyses.

First, we checked the correlation between the log transformed word frequencies in [appendix 1](#) and the initial reading time for the evaluative term. If word frequency is the cause for the processing differences for positive versus negative questions, we would expect to observe a negative correlation across all test items. Indeed, we found the correlation to be negative ($r = -0.15$; $p < .001$).

Second, we performed analyses on the first-pass first-fixation duration (see online appendix 5 for the parameter estimates). The first-pass first-fixation duration reflects the time of the first fixation in a region in which there are no preceding fixations assigned to a higher object. If the polarity effects for the evaluative term are caused by question characteristics, such as word frequency, a main effect of question polarity must be visible in this superficial eye-tracking measure too. Indeed, results show a main effect of question polarity for the first-pass first-fixation duration ($\chi^2 = 4.73$; $df = 1$; $p = 0.03$; Cohen's $d = 0.27$).

Finally, we performed additional analyses on a subset of items ($N = 12$) related to the word pair *forbid/allow* (*toegelaten/verboden* in Dutch). This is the only word pair in our set of items for which the word frequency for the positive and negative term is comparable, whereas the word length is longer for the positive wordings. If characteristics such as word length and word frequency caused the polarity effect for the evaluative term in our main analyses, we would therefore expect the effect of question polarity to reverse for this specific set of items. Indeed, analyses show that when only examining the 12

occurrences of *forbid/allow*, the evaluative term receives a longer processing time for positive rather than for negative items ($\chi^2 = 41.20$; $df = 1$; $p < 0.001$; Cohen's $d = 2.04$; see online appendix 5).

Overall, based on these additional analyses, we are confident that the processing differences between positive and negative wording for the initial reading of the evaluative term are caused by superficial question characteristics rather than cognitive load.

READING OCCURRENCE

In the previous sections, we discussed the fixation times. Equally important, however, is an analysis of how often each region is viewed: Reading and rereading *times* are averaged across trials only in which a certain region is actually viewed, whereas the reading and rereading *occurrence* summarize patterns for all possible trials. In absence of relevant differences between single-reading trials and rereading trials, we present a pooled analysis of single-reading trials and rereading trials.

Differences in processing occurrences for contrastive questions: During the initial reading of the question-answer pair, no differences at all can be observed between positive and negative questions (always $\chi^2 < 3.11$; $df = 1$; $p > 0.08$). This means that all regions are read and skipped roughly equally often for contrastive questions.

By contrast, several effects occur during rereading. All regions are reread more often for negative questions, but in some cases these are main effects and in other cases there is an interaction with the ordering of the question segments (in all cases $\chi^2 > 4.43$; $df = 1$; $p < 0.04$; Cohen's $d > 0.55$). As the exact nature of these effects is not relevant for our purposes here, we refer to [table 4](#) for further elaboration.

Exploration Rereading relative to the answer given: In order to explore the nature of the mapping differences further, we performed an additional analysis on the rereading occurrence data, differentiating between respondents answering *yes* versus *no* to positive and negative questions. [Table 5](#) shows the results of this analysis. Please note that, for the sake of convenience, the position of the evaluative term is not taken into account and we have focused on a coarse-grained question and answer region (compare [Kamoen et al. \[2011\]](#)). This is because our earlier analysis showed that in fact all regions are reread more often for negative questions.

Results show that for positive questions, answering *yes* or *no* does not affect the probability that switching between looking at the response options and the question occurs ($\chi^2 = 0.72$, $df = 1$, $p = 0.40$): Rereading the question happens in 44 percent of the trials, and the response options are reread in 30 percent

Table 4. Probability That Reading a Region Occurs for Positive and Negative Questions with the Evaluative Term at the Beginning (ET-AO) or at the End (AO-ET) of the Question

	Proposed subject		Evaluative term (ET)		Indicator of opinion		Attitude object (AO)		Answers	
	Proportion	Logit (SE)	Proportion	Logit (SE)	Proportion	Logit (SE)	Proportion	Logit (SE)	Proportion	Logit (SE)
<i>Means</i>										
<i>Proportion of trials in which the region is fixated during the initial reading</i>										
Positive ET-AO	0.85*	1.75 (0.12)	0.95	2.87 (0.19)	0.68*	0.77 (0.09)	0.99*	4.24 (0.42)	1.00	5.35 (0.56)
Negative ET-AO	0.86	1.81 (0.12)	0.97	3.32 (0.21)	0.71	0.90 (0.09)	0.98	3.88 (0.39)	1.00	6.02 (0.71)
Positive AO-ET	0.46	-0.18 (0.10)	0.95	3.00 (0.19)	0.67	0.70 (0.09)	1.00	5.25 (0.54)	1.00	5.75 (0.64)
Negative AO-ET	0.46	-0.17 (0.10)	0.96	3.08 (0.20)	0.65	0.61 (0.09)	0.99	4.75 (0.47)	1.00	5.55 (0.60)
<i>Proportion of trials in which the region is fixated during rereading</i>										
Positive ET-AO	0.12+	-2.01 (0.12)	0.34**	-0.65 (0.08)	0.17**	-1.62 (0.10)	0.29**	-0.92 (0.10)	0.32+	-0.74 (0.09)
Negative ET-AO	0.18	1.49 (0.11)	0.47	-0.13 (0.08)	0.21	-1.31 (0.09)	0.35	0.61 (0.09)	0.45	-0.20 (0.09)
Positive AO-ET	0.10	-2.19 (0.13)	0.19	-1.46 (0.09)	0.11	-2.14 (0.12)	0.41	-0.35 (0.09)	0.31	-0.82 (0.09)
Negative AO-ET	0.11	-2.07 (0.12)	0.26	-1.04 (0.09)	0.13	-1.89 (0.11)	0.44	-0.25 (0.09)	0.32	-0.58 (0.09)
<i>Variances</i>										
S ² questions	0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)		0.00 (0.00)	
S ² persons initial reading	0.55 (0.09)		1.60 (0.30)		0.30 (0.06)		11.33 (1.83)		11.23 (2.53)	
S ² persons rereading	0.52 (0.10)		0.20 (0.04)		0.26 (0.06)		0.41 (0.07)		0.30 (0.06)	

NOTE.—The mean occurrences of reading are given in proportions and in the logits used for the analysis; the variances and SEs are only provided in logits.

* $p < 0.05$ for the main effect of polarity; ** $p < 0.05$ for the main effect of position (for the occurrence data, the effect of the position of the evaluative term can be summarized as follows: sentence chunks that are positioned at the beginning of the sentence are read and reread more often than sentence chunks that are positioned at the end of the sentence); + $p < 0.05$ for the interaction polarity * position.

Table 5. Probability That Rereading a Question or an Answer Occurs for Respondents Answering Yes and No to Positive and Negative Questions

	Proportion	Question region		Proportion	Answer region	
		Logit	(SE)		Logit	(SE)
Means						
Positive wording						
Yes answer	0.43	-0.30	(0.09)	0.30	-0.86	(0.09)
No answer	0.45	-0.19	(0.10)	0.30	-0.84	(0.10)
Negative wording						
Yes answer	0.49	-0.05 ^a	(0.10)	0.37	-0.52	(0.10)
No answer	0.55	0.21	(0.09)	0.40	-0.42	(0.09)
Variances						
S ² questions		0.06	(0.02)		0.06	(0.02)
S ² persons		0.36	(0.06)		0.31	(0.06)

NOTE.—For the sake of convenience, the rereading frequencies are given in proportions and in the logits used for the analysis. The variances and standard errors are only provided in logits.

^a $p < 0.05$ for the difference between *no* + a negative term and *yes* + a negative term (combined effect for question and answer region).

of the cases. By contrast, for negative questions, answering *no*, on average, requires more switching than answering *yes* ($\chi^2 = 6.53$, $df = 1$, $p = 0.01$; Cohen's $d = 0.28$): The question is reread 55 versus 49 percent of the time, and the answers 40 versus 37 percent. We will elaborate on these results in the Conclusion and Discussion section.

Conclusion and Discussion

A vast body of research shows that respondents are more inclined to disagree with negative questions than to agree with equivalent positive ones. We related this effect on the answers to the cognitive processes underlying question answering. Respondents ($N = 122$) answered attitude questions while their eye movements were recorded. A 2×2 factorial design was implemented in which the question polarity (positive or negative) and the word order of the question (evaluative term at the beginning or at the end) were varied.

Analyses demonstrate the existence of wording effects for contrastive questions once again: Respondents are more likely to disagree with negative questions than to agree with positive ones. Therefore, a further understanding of how respondents answer positive and negative questions is called for.

Eye-tracking data show that there are two distinct ways of answering a question. For some trials, respondents read the question, switch to the answers, and

immediately give an answer (single-reading trials). For other trials, the question and the answers are read initially and some parts of the question and/or the response options are reread before the respondent gives an answer (rereading trials).

Comparisons of fixation times show that for single-reading trials the evaluative term receives more processing time in negative questions than in positive ones. This effect arises irrespective of the position of the evaluative term in the question. In addition, the response options are read longer for negative questions than for positive ones.

For rereading trials, we also observed that during the initial reading of the question, only the evaluative term receives more processing time for negative questions than for positive ones. Hence, this effect is comparable to the effect found for single-reading trials. In contrast to single-reading trials, we did not find longer reading times for the initial reading of the response options. However, we did locate effects related to the rereading time and to the occurrence of rereading. First, the verb and the evaluative term receive a longer fixation time for negative questions. Second, large differences in processing occurrence are shown: During rereading, all regions of negative questions are fixated more often than those in positive ones. In addition, we found that the occurrence of rereading is also bound to the specific answer a respondent gives: For negative questions, rereading occurs even more frequently when respondents answer *no* rather than *yes* to the same negatively worded question. For positive questions, on the other hand, answering *yes* versus *no* does not affect the rereading frequency.

Relating these results to the cognitive processes underlying question answering, a first conclusion is that something happens in the comprehension-retrieval stage: The initial reading of the evaluative term takes more time when the question is worded negatively. In advance, we formulated two possible explanations for such an effect: cognitive load (e.g., Clark 1976) and word frequency (e.g., Pollatsek et al. 2008). The difference between these accounts is that word frequency effects are expected to occur irrespective of the position of the evaluative term in the question, whereas cognitive load predicts larger differences in processing time when the evaluative term is read with other information already in working memory, that is, when the evaluative term is positioned at the end. As we only observed a main effect of question polarity in the current study, word frequency seems the most plausible explanation for the effects observed. This interpretation is particularly likely as additional analyses showed that an effect of question polarity for the evaluative term is visible already in the first-pass first-fixation duration, a measure indicative of superficial language processes. Moreover, the effect of question polarity was reversed for a subset of items ($N = 12$) in which the positive and negative terms had a comparable word frequency, but for which the negative terms had a higher word length. The reversal of the effect of question polarity also

makes other alternative explanations, such as negative attention capture (Scott, O'Donnell, and Sereno 2012), unlikely.

A second conclusion is that our results clearly show differences between positive and negative questions that can be related to the mapping stage. First, for single-reading trials, reading the answering options takes more time for negative questions than for positive ones. Because of the design of our study, we can be quite sure that respondents could not have started the mapping process before arriving at the response options. Therefore, for single-reading trials, this is evidence that the mapping process is affected by the choice of wording.

Second, for rereading trials, we find negative questions to be reread longer and more often. We consider these effects to be reflections of mapping activity mainly: If rereading were to be a reflection of comprehension-retrieval processes, we would expect rereading to occur equally often for *yeah*- and *naysayers* and we found rereading to be dependent on the answer respondents provide. More specifically, it is answering *no* to negative questions that takes extra processing effort in particular.

Taken together, these findings suggest that, although respondents find negative questions slightly more difficult to comprehend than positive ones, wording effects for contrastive questions probably arise when respondents have to translate their own opinion to the precoded response options. This process requires more switching and takes more time for negative questions than for positive ones, and in particular for respondents who want to answer *no* rather than *yes* to negative questions. This suggests that answering options, such as *yes* and *no*, are not simple categories, but are interpreted relative to the evaluative term in the question. If we speculate about *how* the meaning of the response options differs, there are two options to consider. For one, the tendency to report more favorable attitudes for negative questions could be due to a more mitigated meaning of *no*-answers in general, or to a more mitigated meaning of *no*-answers to negative questions in particular. Based on our exploration, we favor the latter option, as we found that answering *no* to negative questions in particular requires more mapping effort.

In order to develop this more specific mapping hypothesis further, we suggest that qualitative studies should be conducted in which respondents are asked to judge the intended strength of *yes*- and *no*-answers to positive and negative questions. We expect that such studies will demonstrate an asymmetrical meaning pattern, that is, *yes* to a negative question may have the same meaning as *no* to a positive question, but we expect *no* answers to negative questions to convey a more mitigated meaning than *yes* answers to positive ones.

In addition, we suggest experimental studies to be conducted on the underlying cause for these meaning patterns. Linguistic theories provide several possible causes as to why *no* to a negative question would convey a mitigated

meaning. This could, for example, be due to the specific pragmatic meaning of double negatives (Blutner 2004), or the pattern could stem from politeness principles, as respondents may simply avoid an impolite *yes*-answer to a negatively phrased question (Brown and Levinson 1987; Colston 1999). In order to test if an asymmetry is more likely to be caused by either one of these mechanisms, an experimental study could be conducted in which not only the polarity of the evaluative term in the question is varied, but also the “polarity” or “desirability” of the attitude object (*The government should allow/forbid individuals to invest in gray/green energy*). If an asymmetrical interpretation stems from politeness, we expect the response asymmetry to reverse depending on the desirability of the attitude object. If the asymmetry is caused by the general mitigated meaning of double negatives, no such reversal should occur.

Overall, these and similar future studies are relevant not just for survey research, but also for the development of the linguistic theories: “after all, although the human species did not evolve in an environment that included attitude surveys, filling out such surveys is a real-world task that uses language for a purpose” (Van Berkum et al. 2009, 1098).

Appendix 1. Word Frequency of the Word Pairs Used

Table A1 shows the manipulated evaluative terms used in the experiment and their word lengths and word frequencies. The average word length for the positive terms is 9.03 characters (SD = 2.27), and for the negative terms 9.57 (SD = 2.64). This difference is not significant ($t(72.37) = 1.12, p = .27$).

The word frequencies in this table are taken from the SoNaR reference corpus of contemporary Dutch (Oostdijk et al. 2013). This corpus includes over 500 million words taken from various sources, such as newspapers and news media. On average, positive terms are more frequent in the SoNaR corpus than their negative equivalents ($t(38.80) = 2.93, p < .01$). In our view, it is impossible to control statistically for these differences in word frequency, because some negative terms are derived from their positive equivalents (important/unimportant) and in such cases family size effects may play a role.

Table A1. Word Frequencies of the Evaluative Terms in the Experiment

# in experiment	Dutch Positive (#characters)	English translation	Word freq.	Dutch Negative (#characters)	English translation	Word freq.
8	Belangrijk (10)	Important	78,460	Onbelangrijk (12)	Unimportant	2,560
5	Acceptabel (10)	Acceptable	1,276	Onacceptabel (12)	Unacceptable	761
2	Veel (4)	Many	724,563	Weinig (6)	Few/Little	153,190
3	Begrijpelijk (12)	Comprehensible	4,631	Onbegrijpelijk (14)	Incomprehensible	4,194
12	Toegelaten (10)	Allowed	18,657	Verboden (8)	Forbidden	18,931
2	Wenselijk (9)	Desirable	3,776	Onwenselijk (11)	Undesirable	288
5	Goed (4)	Good	621,117	Slecht (6)	Bad	72,699
1	Nuttig (6)	Useful	8,755	Nutteloos (9)	Useless	2,188

Note.—For *toegelaten/verboden (allowed/forbidden)*, the reported frequencies are the “lemma” frequencies, which means that inflexions of this verb are also taken into account.

Supplementary Data

Supplementary data are freely available at *Public Opinion Quarterly* online.

References

- Bishop, George, Hans-J. Hippler, Norbert Schwarz, and Fritz Strack. 1988. "A Comparison of Response Effects in Self-Administered and Telephone Surveys." In *Telephone Survey Methodology*, edited by Robert M. Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls, and Joseph Waksberg, 273–82. New York: Wiley.
- Blutner, Reinhard. 2004. "Pragmatics and the Lexicon." In *Handbook of Pragmatics*, edited by Laurence H. Horn and Gregory Ward, 488–519. Oxford: Blackwell.
- Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Chessa, Antonio G., and Bregje C. Holleman. 2007. "Answering Attitudinal Questions: Modelling the Response Process Underlying Contrastive Questions." *Applied Cognitive Psychology* 21:203–25.
- Churchill, Gilbert A. Jr. 1979. "A Paradigm for Developing Better Measures of Marketing Constructs." *Journal of Marketing Research* 16:64–73.
- Clark, Herbert H. 1971. "The Chronometric Study of Meaning Components." In *Current Problems in Psycholinguistics*, edited by Jacques Mehler, 490–505. Paris: Centre National de la Recherche Scientifique.
- . 1976. *Semantics and Comprehension*. The Hague: Mouton.
- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates.
- Colston, Herbert L. 1999. "'Not Good' Is 'Bad,' but 'Not Bad' Is Not 'Good': An Analysis of Three Accounts of Negation Asymmetry." *Discourse Processes* 28:237–56.
- Holleman, Bregje C. 1999a. "Response Effects in Survey Research: Using Meta-Analysis to Explain the Forbid/Allow Asymmetry." *Journal of Quantitative Linguistics* 6:29–40.
- . 1999b. "The Nature of the Forbid/Allow Asymmetry: Two Correlational Studies." *Sociological Methods & Research* 28:209–44.
- . 2000. *The Forbid/Allow Asymmetry: On the Cognitive Mechanisms Underlying Response Effects in Surveys*. PhD diss. Amsterdam/Atlanta: Rodopi Publishers.
- . 2006. "The Meanings of 'Yes' and 'No': An Explanation for the Forbid/Allow Asymmetry." *Quality & Quantity* 40:10–38.
- Hoosain, Rumjahn. 1973. "The Processing of Negation." *Journal of Verbal Learning and Verbal Behaviour* 12:618–26.
- Horn, Laurence R. 1989. *A Natural History of Negation*. Chicago: Chicago University Press.
- Kamoen, Naomi, Bregje C. Holleman, Pim Mak, Ted Sanders, and Huub van den Bergh. 2011. "Agree or Disagree? Cognitive Processes in Answering Contrastive Survey Questions." *Discourse Processes* 48:355–85.
- Kamoen, Naomi, Bregje C. Holleman, and Huub van den Bergh. 2013. "Positive, Negative, and Bipolar Questions: The Effect of Question Polarity on Ratings of Text Readability." *Survey Research Methods* 7:181–89.
- Moors, Aagnes, Jan De Houwer, Dirk Hermans, Sabine Wanmaker, Kevin Van Schie, Anne-Laura Van Harmelen, Maarten De Schryver, Jeffrey De Winne, and Marc Brysbaert. 2013. "Norms of Valence, Arousal, Dominance, and Age of Acquisition for 4300 Dutch Words." *Behavior Research Methods* 45:169–77.
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste, and Ineke Schuurman. 2013. "The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch." In *Essential Speech and Language Technology for Dutch: Results by the STEVIN-Programme*, edited by Peter Spyns and Jan Odijk, 219–47. Springer: Berlin.

- Pollatsek, Alexander, Barbara J. Juhasz, Erik D. Reichle, Debra Machacek, and Keith Rayner. 2008. "Immediate and Delayed Effects of Word Frequency and Word Length on Eye Movements in Reading: A Reversed Delayed Effect of Word Length." *Journal of Experimental Psychology: Human Perception and Performance* 34:726–50.
- Rayner, Keith. 1998. "Eye Movements in Reading and Information Processing: 20 Years of Research." *Psychological Bulletin* 124:372–422.
- Rayner, Keith, and Susan A. Duffy. 1986. "Lexical Complexity and Fixation Times in Reading: Effects of Word Frequency, Verb Complexity, and Lexical Ambiguity." *Memory & Cognition* 14:191–201.
- Rupp, Heather A., and Kim Wallen. 2007. "Sex Differences in Viewing Sexual Stimuli: An Eye-Tracking Study in Men and Women." *Hormones and Behaviour* 51:524–33.
- Scott, Graham G., Patrick J. O'Donnell, and Sara C. Sereno. 2012. "Emotion Words Affect Eye Fixations During Reading." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 38:783–92.
- Sherman, Mark A. 1973. "Bound to Be Easier? The Negative Prefix and Sentence Comprehension." *Journal of Verbal Learning and Verbal Behavior* 12:76–84.
- Sudman, Seymour, and Norman M. Bradburn. 1982. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco: Jossey-Bass Publishers.
- Swain, Scott D., Danny Weathers, and Ronald W. Niedrich. 2008. "Assessing Three Sources of Misresponse to Reversed Likert Items." *Journal of Marketing Research* 45:116–31.
- Tourangeau, Roger, Lance J. Rips, and Kenneth A. Rasinski. 2000. *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Van Berkum, Jos J. A., Bregje C. Holleman, Mante Nieuwland, Marte Otten, and Jaap Murre. 2009. "Right or Wrong? The Brain's Fast Response to Morally Objectionable Statements." *Psychological Science* 20:1092–99.
- Warriner, Amy B., Victor Kuperman, and Marc Brysbaert. 2013. "Norms of Valence, Arousal, and Dominance for 13,915 English Lemmas." *Behavior Research Methods* 45:1191–207.
- Weisberg, Herbert F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.
- Yan, Ting, and Roger Tourangeau. 2008. "Fast Times and Easy Questions: The Effects of Age, Experience and Question Complexity on Web Survey Response Times." *Applied Cognitive Psychology* 22:51–68.