Field Methods
OnlineFirst

https://doi.org/10.1177/1525822X221115506

SAGE journals

Original Manuscript

The Devil Is in the Details: A Randomized Experiment Assessing the Effect of Providing Examples in a Survey Question across Countries

Eva Aizpurua , Gianmaria Bottoni, and Rory Fitzgerald

Abstract

Despite the widespread use of examples in survey questions, very few studies have examined their impact on survey responses, and the evidence is mainly based on data collected in the United States using questionnaires in English. This study builds on previous research by examining the effects of providing examples using data from a cross-national probability-based web panel implemented in Estonia (n = 730), Great Britain (n = 685), and Slovenia (n = 529) during Round 8 of the European Social Survey (2017/18). Respondents were randomly assigned a survey question measuring confidence in social media using Facebook and Twitter as examples, or another condition in which no examples were offered. The results show that confidence in social media was significantly lower in the example condition, although the effect size was small. Confidence in social media varied across countries, and the effect of providing examples was heterogeneous across countries and education levels. The implications of these findings are discussed.

European Social Survey Headquarters, City, University of London, London, UK

Corresponding author(s):
Eva Aizpurua, European Social Survey Headquarters, City, University of London, Northampton Square, London EC1V 0HB, UK.
Email: eva.aizpu@gmail.com

Introduction

Optimally answering survey questions often demands substantial cognitive effort from respondents. They are required to (1) interpret the intended meaning of the question, instructions, and response options; (2) retrieve relevant information from memory; (3) integrate the information into a judgment; and (4) map the judgment onto the response options provided to them (Tourangeau 2017; Tourangeau et al. 2000). Multiple strategies are implemented to facilitate this process, from improving the design of the questionnaire (e.g., shortening reference periods to reduce recall biases) to planning the administration (e.g., randomizing the order in which response options are shown to minimize response-order effects) (Dillman et al. 2014). One of the strategies used to assist respondents in the response process is the use of clarifications features, such as providing examples, in survey questions.

Providing examples mostly affects the processes involved in comprehension and retrieval. To comprehend a question, respondents must understand their meaning and scope in a way that ensures the item is measuring the intended concept. Questions that are conceptually or linguistically complex are

more likely to produce greater comprehension difficulties (Holbrook et al. 2006; Lenzner et al. 2011; Olson et al. 2019). In surveys involving multiple populations, comprehension difficulties might differ across groups, introducing different levels of measurement error that threaten the comparability of the findings (Aizpurua 2020; Harkness et al. 2010; Smith 2018). Some behavior-coding studies conducted in the United States have shown that interviews conducted in languages other than English result in more requests for clarification (Harkness et al. 2010; Kapousouz et al. 2020). Also in the United States, it has been found that, when interviews are conducted in English, minority populations tend to express more comprehension difficulties than non-Hispanic Whites (Johnson et al. 2006). These findings warrant caution in the use of examples when surveying multiple populations, as they could increase comparability error if processed differently across cultural or linguistic groups.

Examples, however, can help clarify the scope or the specificity of a question, minimizing comprehension difficulties. This might be particularly helpful in self-administered surveys where interviewers are not available to provide clarification, minimizing the risk of respondents misinterpreting the intent meaning of the question, which increases measurement error. An early study from the U.S. Census Bureau comparing two questions measuring ethnicity revealed that offering examples of nationalities (e.g., Colombian, Salvadoran, Spaniard) resulted in significantly more respondents providing specific nationalities as a response than the version using general descriptors (i.e., Hispanic/Latino) (Martin 2002). Other studies examining the impact of question characteristics, however, have found that including definitions in the question stem has little impact on response behaviors (Olson et al. 2019). In a telephone survey, Olson and colleagues (2019) found that providing examples was unrelated with non-substantive responses or requests for clarification from respondents, although they increased the risk of un-codable answers.

Examples can also influence the retrieval stage by stimulating recall, serving as reminders of instances that might, otherwise, be overlooked. In a study by Tourangeau and colleagues, it was found that respondents who received examples of food categories (i.e., dairy, poultry, vegetables, and grain) reported consuming more servings than those who did not receive examples (Tourangeau et al. 2014). Another study conducted with university applicants in Germany provided evidence that adding a wide range of examples to four questions resulted in higher reports of physical impairment symptoms (e.g., headaches), sources of information about college (e.g., websites), and perceived challenges associated with college (e.g., grades) than the same questions with no examples (Metzler et al. 2015).

The risk, however, is that examples introduce bias by enhancing recall for the example items while reducing recall for non-examples. Evidence of this phenomenon, referred to as the “focusing hypothesis,” was obtained in a study analyzing the impact of providing examples on a question about multitasking. While providing examples did not increase the number of secondary activities reported by respondents, it switched the focus to the activities used as examples. Respondents who received example activities were more likely to report them, while those who did not receive examples more frequently listed activities outside of the examples. This meant there was no difference in the average number of activities mentioned, but significant differences in the specific activities reported by the two groups (Aizpurua et al. 2021).

In this regard, excessive focus on examples represents a form of satisficing, as it implies spending minimum effort to provide a satisfactory response. Similar to other forms of weak satisficing, such as response-order effects or acquiescence, respondents would execute the four stages of processing (i.e., comprehension, retrieval, mapping, and retrieving), but engaging in them (particularly the retrieval stage) in a superficial manner. According to the theory of satisficing, factors such as cognitive ability and motivation would encourage engagement in optimal behavior while task difficulty would promote

satisficing (Krosnick 1991).

Present Study

Despite examples being included in survey questions, few studies have analyzed their impact on responses. This body of research suggests that respondents are generally more likely to select the examples which are provided to them both in open- and close-ended questions (Aizpurua et al. 2021; Tourangeau et al. 2014, 2016). One of the weaknesses of previous research is that they are mostly derived from studies conducted in the United States, with questionnaires administered in English. The extent to which the findings from those studies replicate in places outside of the United States and languages other than English is unknown. Although some of these prior studies have been conducted using web surveys (Tourangeau et al. 2014, 2016), their findings are based on non-random samples, limiting the generalizability of the results.

We contribute to this literature by analyzing the results of a randomized experiment examining the impact of providing examples on confidence in social media. This experiment was included in the first wave of a cross-national, probability-based panel implemented in three European countries (Villar et al. 2018). Based on previous research, the following hypotheses were proposed:

H1 *Levels of confidence in social media will differ depending on whether the question includes examples.*

H2 *The number and specific social media platforms considered by re-spondents will differ depending on whether the question includes examples.*
We will find support for the "focusing hypothesis" if examples do not result in respondents recalling more social media platforms, but rather shifting their focus to the instances mentioned to them. Conversely, finding that respondents in the example condition select a greater number of social platforms would signal that using examples assists by triggering the retrieval process more generally and that respondents use the examples as non-exhaustive cues.

H3 *Providing examples will have heterogeneous effects across educa-tional levels and countries.*
Based on the theory of satisficing (Krosnick 1991; Roberts et al. 2019), we would expect examples to have stronger effects among respondents with lower education levels (proxy for cognitive ability; see Roberts et al. 2019).
Considering the existence of large differences in Internet penetration and social media use among Estonia, Slovenia, and the United Kingdom at the time of data collection, we anticipated differences in the effect of the examples across these countries. The United Kingdom had the highest proportion of Internet and social media users (95% and 71% in the last 12 months, respectively), followed by Estonia (with figures similar to the average in the EU28 at 89% and 60%, respectively), and Slovenia (80% and 45%) (Eurostat 2017a, 2017b).

Methods

Data

The European Social Survey (ESS) is a cross-national, academically driven survey conducted every two years across large parts of Europe (Jowell et al. 2007). In Round 8, a cross-national web panel (CRONOS) was implemented, recruiting respondents off the back of the ESS in three countries: Estonia,

Great Britain, and Slovenia. To minimize coverage errors, respondents with no Internet access were offered tablets with high-speed connection during the duration of the panel (Villar et al. 2018). Table 1 provides additional information about the characteristics of the ESS and CRONOS.

Table 1. Key Characteristics of ESS R8 and CRONOS.

	ESS R8	CRONOS
Participating countries	AT, BE, CH, CZ, DE, EE, ES, FI, FR, GB, HU, IE, IL, IS, IT, LT, NL, NO, PL, PT, RU, SE, SI	EE, GB, SI
Population	Individuals 15 and over living in private households	Individuals 18 and over who participated in the ESS Round 8
Sampling	Probability-based: Country-specific	Probability-based: Piggyback from ESS R8
Survey mode	Face-to-face	Web (multi-device)
Data collection	2016–2017	2017–2018
Survey length	~60 minutes (English)	~20 minutes (English)
Data access	Free of charge for non-commercial use: https://www.europeansocialsurvey.org	Free of charge for non-commercial use: https://www.europeansocialsurvey.org

The cross-national web panel was the first attempt to develop a cross-national, probability-based panel following an input-harmonized approach from the recruitment to data processing stages. In total, six waves of data were collected between February 2017 and February 2018. Surveys took approximately 20 minutes to complete and were administered in English (Great Britain), Slovenian (Slovenia), Estonian and Russian (Estonia).

In this article, we analyze data from individuals who participated in the first wave of CRONOS, making use of their corresponding information from the main ESS. All eligible respondents were provided with 5€/£5 unconditional incentives each wave.[1] Response rates[2] ranged from 15% in Great Britain to 25% in Estonia (Slovenia = 23%), while participation rates[3] were all above 50% (56% in Great Britain, 67% in Slovenia, and 78% in Estonia) (Villar et al. 2018).

Experimental Design and Measures

Wave 1 of CRONOS comprised 99 items adapted from the European Values Study. Several randomized experiments were part of wave 1, including the one analyzed in this article.[4] This between-subjects experiment was embedded in a question assessing levels of confidence in social media. Respondents were randomly assigned to a condition in which "Facebook and Twitter" were used as examples, or a condition in which no examples were provided. The question ("How much confidence do you have in social media [like Facebook and Twitter"]) was measured using a four-point, unipolar scale ranging from "a great deal" to "none at all."

As expected, there were no differences in the proportion of respondents allocated to each condition by

country (Examples condition: 50.5% EE, 49.9% GB, 49.0% SI, $X^2 = 0.109,\ p = .896$). Immediately after the question, respondents were queried about the platforms that they took into consideration while answering the former question. In total, 15 platforms were listed, and respondents had to indicate whether they had considered each one (recording yes or no for each). These items were located toward the middle of the questionnaire, at the end of a battery of questions measuring trust in multiple institutions and organizations (e.g., the press, major companies, environmental organizations[5]). The survey was programmed following a paging design, in which each item was presented on a single screen.

To increase the precision of the estimates, several covariates were incorporated in the multivariate models, including age (measured in years and then recoded into age groups), sex (male, female), level of education (output harmonized into seven categories, which were grouped distinguishing among lower and upper secondary education and tertiary education), and Internet usage (never, occasionally/a few times a week, most days, and every day). The wording of all questions is provided in the Appendix. These models are fully consistent with those estimated with no covariates, which are available in the Appendix (Table A2)

Analytical Strategy

Our analysis was organized in three steps. We first examined whether providing examples influenced levels of confidence in social media by comparing the responses between the groups using chi-square tests. The magnitude of the differences was assessed using *Cramer's V.* Then, we investigated whether the platforms considered to assess confidence in social media varied by experimental condition by using chi-square tests and computing effect sizes (*Cramer's V*). To test whether providing examples improved recall, a count variable was created by summing the number of platforms selected by respondents (range = 0–15). Differences in the average number of platforms between the groups were explored using an independent-samples *t*-test.

To further analyze the impact of the examples on social media confidence, ordinal regression models were estimated. Confidence in social media was regressed on the experimental condition while controlling for the country where the survey was fielded, the age, sex, and education level of the respondent, and their frequency of Internet use. There was no indication of multicollinearity in the model (VIF <1.50). Finally, to assess whether the effect of the examples was moderated by the respondents' education level or country, two additional models were estimated including interaction terms between the experimental condition and these two variables. All analyses were computed using weighted data, adjusting the post-stratified ESS R8 design weight for nonresponse at wave 1 of CRONOS.

Results

Description of the Sample

Of those invited to participate in the first wave of CRONOS, 1,944 took part [(730 in Estonia, 685 in Great Britain, and 529 in Slovenia)]. Table 2 summarizes the characteristics of the sample, overall and disaggregated by group. Consistent with the random allocation of respondents to the experiment, the two groups were balanced, and there were no significant differences between them in any of the variables under study, except for age measured as number of years ($F = 5.09$, $p = .024$), but not when categorized in groups ($X^2 = 1.316,\ p = .250$).

Table 2. Sample Composition (Weighted Data).

	No

	Overall		examples		Examples	
Variable	%	*N*	%	*N*	%	*N*
Country						
Estonia	37.6	730	37.1	370	38.8	360
Slovenia	27.2	529	27.7	270	26.7	252
Great Britain	35.2	685	35.2	326	35.3	359
Sex						
Male	49.2	845	48.3	415	50.1	430
Female	50.8	1,099	51.7	558	49.9	541
Age M (SD)	48.1	17.9	46.9	17.6	49.3	18.1
18–29	18.7	325	20.9	178	16.5	147
30–39	16.8	359	16.1	172	17.5	187
40–49	16.9	376	17.6	199	16.1	177
50–59	17.8	357	17.2	172	18.4	185
60–69	16.8	331	17.1	170	16.6	161
70+	12.6	192	10.8	81	14.4	111
Item nonresponse	0.4	4	0.3	1	0.6	3
Education						
Lower secondary and below	21.8	274	21.8	139	21.9	135
Upper secondary	50.1	997	50.7	494	49.6	503
Tertiary education	26.9	662	27.0	337	26.8	325
Item nonresponse	1.1	11	0.5	3	1.7	8
Internet usage						
Never	10.7	133	9.8	64	11.7	69
Occasionally/a few times a week	10.6	180	8.7	77	12.5	103
Most days	7.9	162	7.9	82	7.9	80
Every day	70.8	1,469	73.6	750	68.0	719

Note: M = mean; SD = standard deviation. Unweighted number of observations and weighted proportions are displayed. Internet usage was measured before being recruited to CRONOS.

Respondents' ages ranged from 18 to 91 (M = 48.09, SD = 17.86), and roughly half of the sample were females (50.8%) (see Table 2). In terms of education levels, approximately half of respondents (50.7%) had completed upper secondary and post-secondary education (ISCED[6] levels 3, 4, and 5), with the remainder having tertiary education (27.2%, ISCED levels 6 and 7) and lower secondary education and below (22.1%, ISCED levels 1 and 2). Internet use was common, with seven in 10 respondents using it every day (70.8%).

Social Media Confidence by Experimental Condition

We first examined whether confidence placed in social media differed between experimental conditions.

As shown in Table 3, our findings reject the null hypothesis that confidence in social media is comparable between the groups ($X^2 = 3.583$, $p = .007$). When examples were offered, the proportion of respondents who reported not having confidence was 5 percentage points higher than in the condition with no examples (28.7% versus 23.7%). Similarly, the group receiving the examples indicated to a lesser extent having quite a lot or a great deal of confidence when compared to the no example condition (14.3% vs 21.0%). The effect size, however, was small (*Cramer's V* = 0.102).

Table 3. Confidence in Social Media and Social Media Platforms Considered by Experimental Condition (Weighted Data).

	No examples % (*n*)	Examples % (*n*)	X^2	*p*-value
Confidence				
A great deal	0.3 (3)	0.6 (6)	3.583	0.007
Quite a lot	20.7 (178)	13.7 (121)		
Not very much	51.6 (528)	53.2 (535)		
None at all	23.7 (233)	28.7 (276)		
Item nonresponse	3.7 (31)	3.9 (33)		
Platforms mentioned				
Facebook	77.2 (710)	79.3 (743)	0.756	0.385
Twitter	53.0 (433)	50.2 (438)	0.910	0.340
Snapchat	37.3 (260)	30.5 (228)	5.442	0.020
Instagram	47.0 (360)	41.4 (343)	3.653	0.056
LinkedIn	31.6 (248)	28.6 (239)	1.211	0.271
Youtube	70.3 (561)	66.3 (562)	2.328	0.127
Google+	63.6 (502)	59.0 (467)	2.626	0.105
Reddit	17.8 (116)	14.6 (108)	1.923	0.166
Tumblr	20.3 (128)	15.9 (118)	3.260	0.071
Pinterest	26.9 (191)	24.3 (187)	0.972	0.324
Wikipedia	52.9 (414)	49.0 (399)	1.768	0.184
Flickr	16.7 (115)	14.9 (118)	0.736	0.391
Vine	17.1 (103)	13.0 (91)	3.285	0.070
Periscope	13.5 (87)	11.4 (82)	1.037	0.309
Other	26.0 (157)	19.3 (121)	5.819	0.016
	M (SE)	M (SE)	t-test	p-value
Number of platforms	4.71 (0.15)	4.50 (0.13)	1.180	0.278

Note: M = mean; SE = standard error. Design-based chi-square and adjusted Wald tests (*F*) are reported here. Unweighted number of observations and weighted proportions are displayed.

To examine potential differences in the platforms that respondents considered when reporting their

confidence in social media, we compared both the percentage of respondents who selected each platform and the total number of platforms by experimental condition. As shown in Table 3, the overall number of platforms considered by respondents was comparable between the groups. On average, respondents reported between four and five different platforms ($M_{noexamples} = 4.71$ and $M_{examples} = 4.50$).

Very few differences were identified in the platforms that respondents considered when responding to the question about confidence in social media. Of the 15 platforms under consideration, only two showed significant differences between the two conditions: Snapchat and "other." In both cases, the proportion of respondents indicating having considered them was lower in the group receiving the examples (30.5% vs 37.3% for Snapchat and 19.3% vs 26.0% for "other"). Respondents who were not given examples were as likely as those receiving Twitter and Facebook as examples to indicate having considered these platforms in their responses. In fact, Facebook was the social media platform that most people selected regardless of the condition, with nearly eight in 10 participants indicating having considered it when reporting their confidence in social media.

Social Media Confidence: The Role of Examples, Country, and Education

To further explore the influence of the examples on social media confidence, we estimated a series of ordinal regression models. In them, confidence in social media was recoded, combining the two highest categories (lowest frequencies), and reversing the direction of the scale so that higher values represented increased confidence. The first model regressed confidence in social media on the experiment and the countries to see if levels of confidence varied across countries and between the two experimental conditions. Sex, age, frequency of Internet use, and education level were included as covariates. The results from this model are consistent with the bivariate findings, showing that providing examples is associated with lower confidence in social media (OR = 0.711, $p = .002$). The country where the study was conducted was also significant, with respondents in Slovenia having increased odds of greater confidence than those in Great Britain (OR = 1.850, $p < .001$). In this model, confidence in social media was higher for females than males (OR = 1.282, $p = .023$) and lower for those with tertiary education when compared to those with upper secondary education (OR = 0.755, $p = .013$). Levels of social media confidence were comparable across age groups and frequency of Internet use (see Table 4).

Table 4. Ordinal Regression Models Predicting Confidence in Social Media by Experimental Condition (Weighted Data).

	Model 1 (main effects)			Model 2 (interaction education)			Model 3 (interaction country)		
Variable	OR	SE	95% CI	OR	SE	95% CI	OR	SE	95% CI
Examples provided	0.711**	0.078	0.573, 0.882	0.609***	0.084	0.464, 0.799	0.949	0.197	0.631, 1.426
Country (ref. great Britain)									
Estonia	1.290	0.172	0.992, 1.676	1.298	0.173	0.999, 1.687	1.463*	0.276	1.010, 2.118
Slovenia	1.850***	0.272	1.387, 2.468	1.855***	0.272	1.391, 2.473	2.598***	0.551	1.714, 3.937
Female	1.282*	0.140	1.034, 1.588	1.285*	0.140	1.037, 1.592	1.280*	0.140	1.033, 1.586
Age (ref. 18–29)									
30–39	0.824	0.156	0.568, 1.194	0.833	0.159	0.573, 1.210	0.810	0.154	0.559, 1.175
40–49	0.806	0.148	0.562, 1.155	0.811	0.150	0.564, 1.167	0.798	0.147	0.557, 1.145
50–59	0.836	0.156	0.580, 1.204	0.843	0.158	0.585, 1.217	0.816	0.152	0.566, 1.177
6,069	0.901	0.184	0.603, 1.345	0.910	0.187	0.608, 1.361	0.868	0.180	0.578, 1.304
70+	0.700	0.176	0.428, 1.146	0.697	0.175	0.426, 1.141	0.682	0.173	0.415, 1.120
Internet use (ref. every day)									
Never	1.355	0.356	0.809, 2.269	1.355	0.354	0.812, 2.263	1.352	0.355	0.807, 2.263
Occasionally	0.768	0.165	0.504, 1.169	0.763	0.164	0.500, 1.164	0.774	0.164	0.510, 1.174
Most days	1.229	0.266	0.804, 1.878	1.227	0.264	0.804, 1.872	1.249	0.274	0.812, 1.921
Education (ref. upper secondary)									
Lower secondary	0.899	0.165	0.628, 1.288	0.842	0.222	0.502, 1.412	0.903	0.164	0.632, 1.291
Tertiary	0.755*	0.085	0.606, 0.942	0.603**	0.097	0.440, 0.827	0.755*	0.085	0.605, 0.942
Examples*education									

Examples*education								
Examples provided*Lower			1.145	0.412	0.565, 2.320			
Examples provided*Tertiary			1.564*	0.339	1.022, 2.393			
Examples*country								
Examples provided*Estonia						0.776	0.200	0.468, 1.287
Examples provided*Slovenia						0.509*	0.150	0.285, 0.907
F	3.865***		3.600***			3.700***		
N	1,865		1,865			1,865		

Note: OR = odds ratio; SE = standard error; CI = confidence interval.
*p < .05; **p < .01; ***p < .001.

To assess whether the effect of the experimental manipulation was comparable across education levels, an interaction term was included in Model 2, while we adjusted for all other variables (i.e., country, sex, age, and Internet use). The results were significant, suggesting that the effect of the examples differed depending on respondents' education levels (see Table 4). As can be seen in Figure 1, the experiment produced no differences among respondents with tertiary education, while those with secondary education (both lower and upper) were more sensitive to the experimental manipulation.

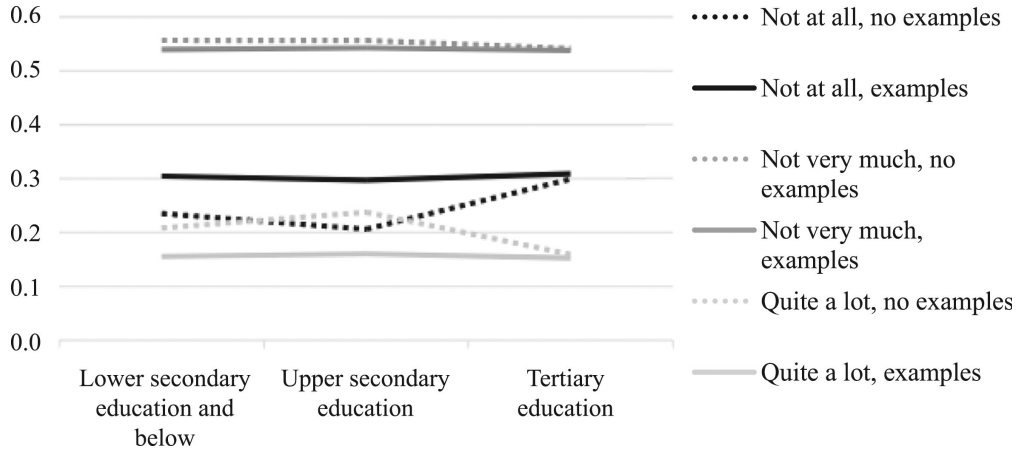

Figure 1. Predicted probabilities—Confidence in social media by experimental condition and education level.

To assess whether the effect of providing examples was comparable across countries, an interaction term between these two variables was added in Model 3. The results suggest that the effect was different in Slovenia (OR = 0.509, p = .022). As can be seen in Figure 2, overall differences were mostly driven by this country. Whereas in Great Britain and Estonia predicted probabilities were very similar across conditions, in Slovenia, respondents who received no examples were more likely to report the highest level of confidence in social media (light gray dotted line), and less likely to report the lowest level of confidence (black dotted line).

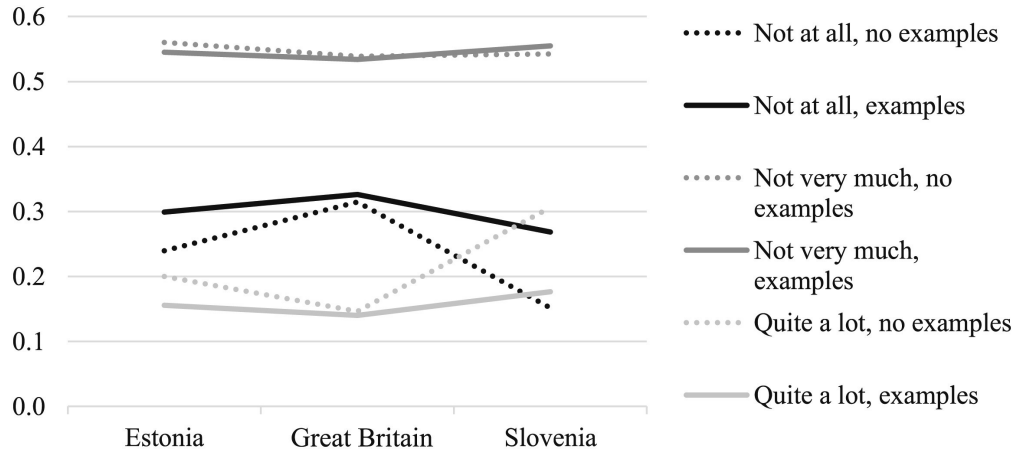

Figure 2. Predicted probabilities: Confidence in social media by experimental condition and country.

Discussion

There has been little empirical research into the impact that offering examples in survey questions has on the answers provided by respondents. We have contributed to addressing this gap in the literature by examining the results of a randomized experiment implemented in a probability-based, cross-national web survey fielded in three European countries. Our results show that levels of confidence in social media vary depending on whether the question stem included Facebook and Twitter as examples. When these examples were offered, confidence was reduced, although the effect size was small. This is consistent with the hypothesis that examples influence the comprehension stage of the response process, making the examples more salient.

Beyond this global effect, our findings suggest that the effect of examples on survey responses might vary across education levels. While those with tertiary education showed comparable levels of confidence in social media regardless of the condition they were assigned to, respondents with lower levels of education were more affected by the inclusion of examples. This is consistent with the satisficing hypothesis positing that lower respondent ability increases the risk of measurement error (Krosnick 1991). If respondents are optimizing and interpreting the examples as a non-exhaustive list of instances, the use of different sets of examples would result in small differences. On the contrary, if respondents complete the tasks involved in the response process superficially, they might excessively focus on the examples, failing to infer from the examples or to retrieve information pertaining to other instances.

When analyzing the impact of the examples across countries, significant differences were found, with Slovenia—the country with the lowest Internet penetration and social media use at the time—driving the overall differences. This finding suggests that the effect of providing examples might vary across cultural or linguistic groups. In such a case, the use of examples in the context of cross-national research might increase comparison error (Smith 2018). Finding examples that are meaningful across cultures and display the same intensity and meaning can be challenging. Our results suggest that even when the examples might be relevant in all participating countries, their effects might be heterogeneous, introducing different levels of measurement error. If examples are included in cross-cultural surveys, carefully pre-testing the questions with different groups of the population and in as many countries as possible is strongly recommended.

Consistent with previous research (Aizpurua et al. 2021), we did not find evidence that the use of examples improves the recall process. If this were the case, respondents in the example condition would have reported a greater number of platforms than those in the no example condition. On average, respondents in both groups indicated having considered between four and five different platforms, with Facebook, YouTube, Google+, and Twitter being the most reported among both groups. The fact that the examples provided to respondents were among the most well-known social media platforms might explain these findings. A study conducted by Tourangeau and colleagues in the United States showed that atypical examples had greater impact on survey responses than typical examples, perhaps because respondents are likely to consider the typical examples regardless of whether they are presented to them (Tourangeau et al. 2014). Our results are consistent with this hypothesis, with individuals in the no example condition being as likely to report Facebook and Twitter as those in the example group. This finding provides no support to the "focusing hypothesis," according to which respondents would have been more likely to report the examples that were provided to them.

Our question examining the specific platforms that respondents had considered when reporting their

confidence in social media displayed 15 social media outlets and asked panelists whether they have considered each one when reporting their levels of confidence. Previous research suggests that response options are endorsed at higher rates when using yes–no questions, such as the ones we used, when compared to alternative formats such as "mark-all-that-apply" (Neuert 2020). Because of this, and taking into account the relatively high number of platforms reported by respondents (*Mdn* = 4), further studies testing the focusing hypothesis with alternative question formats (e.g., mark-all-that-apply, open-ended) are warranted. Similarly, confidence in social media represents a salient topic for many people and understanding the impact of examples in topics with different levels of interest, and various types of questions (e.g., attitudinal vs behavioral) deserves further consideration.

The results of this study suggest that examples influence the way respondents process survey questions and respond to them, although their effects are small. There are, however, a number of limitations of the current work. First, only one set of examples was used, and they were typical, representing two of the most commonly used social media platforms at the time of data collection. Although our study includes three countries, and the survey was administered in four languages, further cross-cultural research, with a wider and more diverse set of countries is needed.[7] Third, we use educational level as a proxy for respondent ability, mindful that this represents an imperfect indicator of this construct. Future research would benefit from including indicators of task difficulty and respondent motivation, given their potential additive or multiplicative effect (Roberts et al. 2019). Finally, because panelists were recruited off the back of the ESS, the response rate was lower than that of the ESS, increasing the risk of nonresponse bias. Recent studies, however, provide evidence that the CRONOS sample is not extremely divergent from the target population or to the data from the main ESS (Bottoni and Fitzgerald 2021; Maslovskaya and Peter Lugtig 2022).

Conclusion

Examples are widely used in survey questions to clarify their intended scope, indicate the type of expected responses, and/or remind respondents of instances that might otherwise go unnoticed. The current study suggests that examples influence the cognitive process involved in answering questions, with responses being different depending on whether examples were provided. Although our study does not provide an answer to the question of which version of the item produces higher data quality, it suggests that confidence in social media is reduced when examples are offered, and that this effect is stronger in Slovenia than in the Great Britian. Because measurement quality varies greatly across countries (Bosch and Revilla 2021), it is possible that providing examples, even when they have differential effects, might result in measurement errors being more comparable. However, using examples can also amplify existing differences, increasing comparability error. This requires caution when it comes to including examples in the context of cross-national surveys. If examples are included, carefully pre-testing the questions to ensure that they are interpreted as intended across countries and groups is necessary to be confident that differences or similarities found are not an artifact of measurement error. Studies using multitrait-multimethod experiments could also be used to estimate the measurement quality of questions with and without examples, providing much needed guidance for survey researchers and practitioners.

Acknowledgments

The authors thank those who made CRONOS possible, including those who designed and implemented this experiment (in particular Ana Villar, Angelica Maineri and Ruud Luijkx) and the respondents who kindly gave their time to take part in the survey. A special thanks to the ESS National Coordinators in Estonia, Great Britain, and

Slovenia, and those working on the CRONOS Central Team (City, University of London, Norwegian Center for Research Data, University of Ljubljana, and Universitat Pompeu Fabra). We are also grateful for the feedback received at AAPOR on an earlier version of this work, and the input provided by the reviewers and editors of *Field Methods*.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Although the article was written in its entirety while all authors were working at the European Social Survey Headquarters, the first author joined Facebook (one of the two platforms used in the examples) after submitting the manuscript to Field Methods.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Horizon 2020 Framework Programme.

ORCID iD

Eva Aizpurua https://orcid.org/0000-0001-7045-5535

Notes

1 In Great Britain, a split-ballot experiment was embedded in CRONOS, where respondents were randomly allocated to receive a £5 unconditional incentive with each wave or a £30 unconditional incentive for the six waves upfront with wave 1.
2 Number of interviews (partial and completes) as a proportion of the original ESS gross sample for each country.
3 Number of participants (partial and completes) as a proportion of sample units invited to participate in CRONOS.
4 Further information about CRONOS experiments can be found at: https://www.europeansocialsurvey.org/docs/cronos/CRONOS_experiments_description.pdf
5 For the full questionnaire, see https://www.europeansocialsurvey.org/docs/cronos/CRONOS_source_questionnaire_wave1.pdf
6 International Standard Classification of Education.
7 Although the survey was administered in two languages in Estonia, the small sample size, particularly for Russian responses (n = 129, with 64 in the control group and 65 in the treatment group), prevented us from conducting additional analyses examining within-country differences. We replicated the bivariate findings presented in Table 2 and found consistent results for both languages. The number of social media platforms was comparable between the groups, with panelists reporting around five outlets in both conditions and languages. The chi-squared tests examining differences in social media confidence between the experimental groups did not reach statistical significance, but the patterns were similar to those in the aggregate data, with panelists in the example condition reporting slightly lower levels of confidence in social media.

References

Aizpurua E. 2020. Pretesting methods in cross-cultural research. In *The essential role of language in survey research*, eds. Sha M., Gabel T., 129–50. Research Triangle Park: NC: RTI International.

Aizpurua E., Park K. H., Heiden E. O., Losch M. E.. 2021. I say, they say: Effects of providing examples in a question about multitasking. *International Journal of Social Research Methodology* 24:781–87. Crossref

Bosch O. J., Revilla M.. 2021. The quality of survey questions in Spain: A cross-national comparison. *Revista Española de Investigación Sociológica* 175:3–26. Crossref

Bottoni G., Fitzgerald R.. 2021. Establishing a baseline: Bringing innovation to the evaluation of cross-national probability-based online panels. *Survey Research Methods* 15:115–33.

Dillman D., Smyth J., Christian L.. 2014. *Internet, phone, mail, and mixed mode surveys: The tailored design*

method. 4th ed. Hoboken, NJ: John Wiley & Sons.

Eurostat. 2017a. Internet use in the last 12 months. Percentage of individuals. https://ec.europa.eu/eurostat /databrowser/view/isoc_ci_ifp_iu/default/table?lang=en (accessed May 27, 2022).

Eurostat. 2017b. Internet use: Participating in social networks (creating user profiles, posting messages or other contributions to Facebook, twitter, etc.). Annual. Percentage of individuals. https://ec.europa.eu/eurostat /databrowser/view/isoc_ci_ac_i/default/table?lang=en (accessed.May 27, 2022).

Harkness J. A., Edwards B., Hansen S. E., Miller D. R., Villar A.. 2010. Designing questionnaires for multipopulation research. In *Survey methods in multicultural, multinational, and multiregional contexts*, eds. Harkness J. A., Braun M., Edwards B., Johnson T. P., Lyberg L. E., Mohler P. P., Pennell B.-E., Smith T. W., 1–57. Hoboken, NJ: Wiley.

Holbrook A., Cho Y. I., Johnson T.. 2006. The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly* 70:565–95. Crossref

Johnson T. P., Cho Y. I. K., Holbrook A. L., O'Rourke D., Warnecke R. B., Chavez N.. 2006. Cultural variability in the effects of question design features on respondent comprehension of health surveys. *Annals of Epidemiology* 16:661–68. Crossref

Jowell R., Roberts R., Fitzgerald C., Eva G.. 2007. *Measuring attitudes cross-nationally: Lessons from the European social survey*. Thousand Oaks, CA: Sage.

Kapousouz E., Johnson T. P., Holbrook A. L.. 2020. Seeking clarifications for problematic questions: Effects of interview language and respondent acculturation. In *The essential role of language in survey research*, eds. Sha M., Gabel T., 23–46. Research Triangle Park, NC: NCL RTI International.

Krosnick J. A. 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5:213–36. Crossref

Lenzner T., Kaczmirek L., Galesic M.. 2011. Seeing through the eyes of the respondent: An eye-tracking study on survey question comprehension. *International Journal of Public Opinion Research* 23:361–73. Crossref

Maslovskaya O., Lugtig P.. 2022. Representativeness in six waves of CROss-National Online Survey (CRONOS) panel. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*:1–21. Crossref

Martin E. 2002. The effects of questionnaire design on reporting of detailed Hispanic origin in Census 2000 mail questionnaires. *Public Opinion Quarterly* 66:582–93. Crossref

Metzler A., Kunz T., Fuchs M.. 2015. The use and positioning of clarification features in web surveys. *Psihologija* 48:379–408.

Neuert C. E. 2020. Do forced-choice (FC) questions trigger deeper cognition than check-all-that-apply (CATA) questions? *Journal of Survey Statistics and Methodology* 8:617–35. Crossref.

Olson K., Smyth J. D., Ganshert A.. 2019. The effects of respondent and question characteristics on respondent answering behaviors in telephone interviews. *Journal of Survey Statistics and Methodology* 7:275–308. Crossref

Roberts C., Gilbert E., Allum N., Eisner L.. 2019. Research synthesis: Satisficing in surveys. A systematic review of the literature. *Public Opinion Quarterly* 83:598–26. Crossref

Smith T. W. 2018. Improving multinational, multiregional, and multicultural (3MC) comparability using the total survey error (TSE) paradigm. Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural Contexts (3MC) edited by Timothy P. Johnson, Beth-Ellen Pennell, Ineke A.L. Stoop and Brita Dorer (see https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118884997.ch2).

Tourangeau R. 2017. The survey response process from a cognitive viewpoint. *Quality Assurance in Education* 26:169–81.

Tourangeau R., Conrad F. G., Couper M. P., Ye C.. 2014. The effects of providing examples in survey questions. *Public Opinion Quarterly* 78:100–25.

Tourangeau R., Rips L. J., Rasinski K., eds. 2000. *The psychology of survey response*. New York: Cambridge

University Press.

Tourangeau R., Sun H., Conrad F. G., Couper M. P.. 2016. Examples in open-ended survey questions. *International Journal of Public Opinion Research* 29:690–702. Crossref

Villar A., Sommer E., Finnøy D., Gaia A., Berzelak N., Bottoni G.. 2018. *CROss-national online survey (CRONOS) panel: Data and documentation user guide*. https://www.europeansocialsurvey.org/docs/cronos/CRONOS_user_guide_e01_1.pdf (accessed May 28, 2022).

Appendix

Table A1. Question Wording.

Question stem	Response options
Outcome variables (CRONOS)	
Confidence in social media (Example condition)	A great deal Quite a lot Not very much
How much confidence do you have in social media like Facebook or Twitter?	None at all Don't know Prefer not to answer
Confidence in social media (No example condition)	A great deal Quite a lot Not very much
How much confidence do you have in social media?	None at all Don't know Prefer not to answer
In the previous question, you were asked about social media. Please indicate whether you considered each of the following social media platforms	
Please select yes or no for each example.	
Facebook	
Twitter	
Snapchat	
Instagram	Yes
Linkedin	No
Youtube	
Google+	
Reddit	
Tumblr	
Pinterest	
Wikipedia	
Flickr	
Vine	
Periscope	
Other	
Covariates (from the main ESS)	
Sex	Male

Interviewer-coded	Female
Education (output harmonized)	ES-ISCED I, less than lower secondary ES-ISCED II, lower secondary ES-ISCED IIIb, lower tier upper secondary ES-ISCED IIIa, upper tier upper secondary ES-ISCED IV, advanced vocational, sub-degree ES-ISCED V1, lower tertiary education, BA level ES-ISCED V2, higher tertiary education, ≥ MA level Other
Frequency of Internet use People can use the internet on different devices such as computers, tablets and smartphones. How often do you use the internet on these or any other devices, whether for work or personal use?	Never Only occasionally A few times a week Most days Every day Refusal Don't know

Note: the source questionnaires for all CRONOS waves and all ESS rounds are available at https://www.europeansocialsurvey.org/

Table A2. Missing Data by Variable.

	N	%
Confidence in social media	64	3.3
Experimental condition	0	0.0
Country	0	0.0
Sex	0	0.0
Age group	4	0.2
Internet use	0	0.0
Education	11	0.6

Table A3. Ordinal Regression Models Predicting Confidence in Social Media with no Covariates (Weighted Data).

	Model 1 (main effects)			Model 2 (country interaction)			Model 3 (education interaction)		
Variable	OR	SE	95% CI	OR	SE.	95% CI	OR	SE	95% CI
Examples provided	0.709**	0.077	0.573, 0.879	0.924	0.188	0.619, 1.378	0.607***	0.083	0.465, 0.794
Country (ref. Great Britain)									
Estonia				1.483*	0.280	1.024, 2.147			
Slovenia				2.598***	0.557	1.707, 3.955			
Examples*Country									
Examples provided*Estonia				0.790	0.200	0.467, 1.310			
Examples provided*Slovenia				0.531*	0.154	0.301, 0.938			
Education (ref. upper secondary)									
Lower secondary education							0.848	0.220	0.509, 1.410

Tertiary education	0.560***	0.089	0.410, 0.765
Examples*education			
Examples provided*lower	1.103	0.396	0.545, 2.232
Examples provided*tertiary	1.559*	0.340	1.017, 2.390

Note: OR = odds ratio; SE = standard error; CI = confidence interval.
*p < .05; **p < .01; ***p < .001.