

This appendix relates to

Engel, Uwe & Britta Köster. 2015. Response Effects and Cognitive Involvement in Answering Survey Questions. In Uwe Engel, Ben Jann, Peter Lynn, Annette Scherpenzeel & Patrick Sturgis (Eds.), *Improving Survey Methods. Lessons from Recent Research*, 35–50. New York/London: Routledge.

URL for accessing the PPSM Panel Report (last page) updated September 19, 2024

Table A4.1 Frequency distributions of respondent- and interviewer-ratings in the recruitment interviews† of the PPSM access panel

	1	2	3	4	5	6	7	N
Sensitivity*	49.6%	13.4%	10.4%	10.6%	8.6%	3.8%	3.6%	5,697
Satisficing**	50.7%	30.8%	11.9%	4.4%	1.4%	0.6%	0.3%	5,702
Reluctance***	33.8%	27.4%	18.9%	10.9%	5.3%	2.3%	1.5%	5,737

†Conducted in 2009.

*The perceived degree of sensitivity is rated by: “I would like to come back again shortly to the health questions. How personally did you feel the questions about alcohol consumption to be? On a scale from 1 to 7, the 1 means this time, that you felt the question to be not at all personal, and the 7, that you felt the question to be much too personal. With the values in between you can grade again your opinion.” Campanelli (2008: 187)¹ discusses the use of such a direct question to assess the sensitivity of a survey question with reference to Sudman & Bradburn (1982). According to that discussion, an indirect question would have been better than a direct question, since that would be less threatening to respondents. In view of our own experience with the above-cited direct question, we encountered no such difficulties and see actually no reason to turn to a more indirect way of asking about the sensitivity of survey items.

**The degree of satisficing behavior is rated by: “Attempt to assess the response behavior of the respondent: Did the respondent take care in answering the question properly and conscientiously or did the respondent “rush” through the questions, without thinking of his/her answers properly? Rate the response behavior on a scale from 1 to 7, where the 1 means that the respondent did answer the questions very conscientiously and the 7, that the respondent did answer the questions not at all conscientiously.”

*** The degree of reluctance is rated by the degree which was necessary to win a target person over to the interview: “On a scale from 1 to 7, how do you assess the general readiness of the interviewee to get interviewed”. It followed a 7pt scale ranging from 1=“immediately ready” to 7=“ready only after investing a great deal of convincing effort”. Note that an equivalent scale was used in cases where despite all efforts no interview came about. It was worded as: “On a scale from 1 to 7, how do you assess the general chance to win the respondent over to an interview?” The 7pt scale ranges from 1=“I consider this possible (upon other terms)” to 7=“I consider this absolutely impossible”.

¹ Campanelli, P. (2008) Testing Survey Questions. Pp. 176-200 in De Leeuw, E.D.; Hox, J.J.; Dillman, D.A. (Eds.): *International Handbook of Survey Methodology*. New York: Lawrence Erlbaum Associates

Table A4.2 Stability of satisficing response behavior and reluctance over a sequence of five measurement occasions and occasion-specific satisficing – reluctance correlations

(a) Kendall's tau b

Kendall's τ_b		t ₁		t ₂		t ₃		t ₄		t ₅
		S ₁	R ₁	S ₂	R ₂	S ₃	R ₃	S ₄	R ₄	S ₅
t ₁	S ₁									
	R ₁	0.22*								
t ₂	S ₂	0.08*								
	R ₂		0.05*	0.40*						
t ₃	S ₃			0.04*						
	R ₃				0.02	0.23*				
t ₄	S ₄					0.07*				
	R ₄						0.02	0.24*		
t ₅	S ₅							0.10*		
	R ₅								0.03	0.35*

S=Satisficing behavior, R=Reluctance; t₁ = recruitment interviews (conducted in 2009), t₂ = panel welcoming interviews (on average seven weeks after recruitment, conducted in 2009), t₃, t₄, t₅ = 1st, 2nd and 3rd wave of two panel surveys, one in single-mode (PPSM panels surveys PS₂₁, PS₂₂, PS₂₃) and one in (telephone and web) mixed-mode (PPSM panel surveys PS₁₂, PS₁₄, PS₁₅) conducted in 2010 and 2011 (spring and autumn), with panel distances of about six resp. eight months.

(b) Number of cases

Kendall's τ_b		t ₁		t ₂		t ₃		t ₄		t ₅
		S ₁	R ₁	S ₂	R ₂	S ₃	R ₃	S ₄	R ₄	S ₅
t ₁	S ₁									
	R ₁	4,092								
t ₂	S ₂	3,543								
	R ₂		3,594	3,577						
t ₃	S ₃			3,575						
	R ₃				3,593	3,612				
t ₄	S ₄					2,465				
	R ₄						2,466	2,466		
t ₅	S ₅							1,961		
	R ₅								1,961	2,002

The number of cases is taken from the corresponding computation of Spearman's rho respectively. Note that the computation of Kendall tau b and Spearman's rho yields approximately identical results respectively.

Table A4.3: The effect of perceived sensitivity and satisficing behavior on the strength of a response-order effect on answers to a question about the frequency of alcohol consumption

Satisficing ↓	Perceived sensitivity of a question about alcohol consumption							
	1 low		2 middle		3 high		all	
	N	z*	N	z*	N	z*	N	z*
1 low	1,491	10.54	665	7.09	702	3.24	2,858	12.86
2 middle	791	10.18	458	3.72	484	4.35	1,733	11.24
3 high	513	7.55	220	3.83	303	5.10	1,036	9.86
All	2,795	16.31	1,343	8.64	1,489	7.08	5,627	19.69

* $z=b_1/s.e.$ The intercept varies significantly across the interviewers in most of the constellations. This applies to the overall condition, to five of six marginal conditions and to four of the nine combinations of sensitivity and satisficing. A total of 184 interviewers were involved. As to the single marginal conditions, this number varies between 143 and 174, while the range is between 87 and 152 in the cell conditions. The respondent rating of sensitivity is recoded as (1 = 1 'low')(2, 3 = 2 'middle')(4, 5, 6, 7 = 3 'high'). The interviewer rating of satisficing is recoded as (1 = 1 'high')(2 = 2 'middle')(3, 4, 5, 6, 7 = 3 'low'). Scale for rating alcohol consumption ranges from 1 (never) to 6 (daily). Data source: PPSM recruitment interviews (random samples 1 and 2). These interviews lasted 20 minutes on average and were conducted in 2009.

Comparison of two subject matters in terms of perceived sensitivity

As putative sensitive questions are not necessarily to be felt as sensitive ones, we pursued the approach to let the respondents rate the perceived sensitivity of survey questions with respect to two subject areas: deviant behavior and health behavior. Based on a split-half design, the two meta-questions were worded as follows (The analyses of the present chapter are confined to split A):

Random split A: "I would like to come back again shortly to the health questions. How personally did you feel the questions about alcohol consumption to be?"

Random split B: "I would like to come back again shortly to the questions on behavior like tax fraud and prostitution. How personally did you feel these questions to be?"

It then followed: "On a scale from 1 to 7, the 1 means this time, that you felt the question to be not at all personal, and the 7, that you felt the question to be much too personal. With the values in between you can grade again your opinion."

Controlling for interviewer effects by a random-intercept regression, the mean perceived sensitivity of alcohol consumption turns out to be 2.67 on the above scale from 1 to 7, while the expected mean difference to the deviance questions amounts to -0.45, yielding a mean value of 2.22 for these latter questions ($z=-9.77$; $N=5,747$ respondents, 184 interviewers; random recruitment samples 1 and 2). On average the survey questions on justification of different offences (e.g. accepting social benefits if not entitled to, tax fraud, having an affair as a married person, being bribed etc.) appear thus less sensitive than the question on personal alcohol consumption. This is certainly due to the fact that the latter question is concerning personal behavior while the former question concerns justification in general not aimed specifically on ones' own behavior.

Table A4.4: The effect of satisficing behavior on the strength of a question-wording effect on answers to statements about distributive justice

Images of distributive justice	
JUST 1	It is just if someone who keeps working hard earns more than someone who don't.
JUST 2	It is just if everybody gets even without own efforts what he needs for life.
JUST 3	It can only be just if all people are financially equally well off.
JUST 4	It is just if occupational success alone determines who has more or less money for life at one's disposal.
JUST 5	Not success is decisive but that one always aims to do one's best on the job.
JUST 6	It is just if parents who are able to afford it enable their children a better education at a private school.
JUST 7	It is just if parents pass their wealth down to their children, even if that means that children of wealthy parents have better prospects in life.

Images of distributive justice	Satisficing							
	high		middle		low		all	
	\bar{x}_j	Δ	\bar{x}_j	Δ	\bar{x}_j	Δ	\bar{x}_j	Δ
Justice 1	5.6	-1.4	6.4	-3.0*	6.2	-0.8*	6.2	-1.6*
Justice 2	3.2	0.6	3.9	-0.5	3.3	0.7*	3.5	0.3
Justice 3	2.6	1.4*	2.9	0.3	3.0	0.5	3.0	0.6*
Justice 4	3.9	0.7	4.0	0.3	4.0	0.2	4.0	0.3
Justice 5†	6.0	-1.4*	5.5	0	5.2	-0.1	5.4	-0.3
Justice 6	4.1	-0.3	3.8	0	4.1	0.2	4.0	0.1
Justice 7	5.8	-0.6	5.0	-0.1	4.8	0.1	5.0	0
N	49 to 51		109 to 111		208 to 211		368 to 373	

† This item is identical in both random splits. *Sig. ($p \leq 0.05$). The table displays mean values using the "just" formulation. Δ indicates the extent to which the corresponding mean value falls below (negative sign) or exceeds (positive sign) the mean reported to the left of it. The original 7pt scales range from 1="I agree not at all" to 7="I agree totally". This agreement is recoded to disagreement for the "unjust"-items to obtain corresponding frequency distributions in both split groups. The general introduction to the seven items was phrased as follows: "Since the court decision of the Federal Constitutional Court on the standard rates of HARTZ IV [welfare aid] in February this year many are wondering how a just distribution of prosperity should look like actually? I would like to read seven statements to you and ask you to tell me to each statement how much you agree with it on a scale from 1 to 7. 1 means that you agree not at all with the statement and 7 that you agree totally. With the figures in-between you can grade your opinion. Here the first statement: ... " Data source: PPSM Spring 2011 panel survey PS₃₂.

Table A4.5: The effect of satisficing behaviour on the strength of a question-wording effect on answers to a statement about the violation of rules

	"Some argue that in Germany scarcely anybody takes no liberties with the keeping of rules"					
Satisficing →	low		higher		All	
Split group →	A	B	A	B	A	B
Percent (agree)	56.7	52.1	49.3	49.3	54.6	51.3
N*	171	165	67	71	238	236

*Percentage base. In split group A it was asked: "Do you agree with this statement or do you not agree with this statement?" (0 = do not agree, 1 = agree). Split group B: "Do you agree with this statement or do you refuse this statement?" (0 = refuse, 1 = agree). Data source: PPSM spring 2011 panel survey PS₄₁.

A modifying impact of satisficing behavior is also observed in an experiment that varied the category labels of a dichotomous response scale (see Table A5, above). While an "agree" to the statement "Some argue that in Germany scarcely anybody takes no liberties with the keeping of rules" was used in both random splits, disagreement was worded "not agree" in one and "refuse" in the other group. As is shown above, the two wordings produced only a small difference of about three percentage points: 54.6 percent agreement if this category is contrasted with "not agree" vs. 51.3 percent agreement if contrasted with "refuse". While a low degree of satisficing resp. a high cognitive involvement intensifies this difference slightly by a factor of 1.4, the difference vanishes completely if the satisficing degree of behavior is higher than "low". Thus the semantic difference produces its effect only if the question is very conscientiously answered. This appears contrary to the effect patterns observed in case of the distributive-justice items (Table A4, above) respectively the items on environmental behavior (Table A6, below).

Table A4.6: The modifying impact of satisficing on response behavior using four response scale formats

Cramer's V	Survey mode		Satisficing		
	All	Online	Telephone	low	higher
One should be willing to accept higher financial costs for the sake of protection of the environment	0.31	0.31	0.30	0.29	0.36
One should always choose environment-friendly means of transportation, even if the tours take more time than	0.28	0.33	0.27	0.26	0.31
Climate protection is a luxury which in society not everyone can permit oneself	0.31	0.26	0.31	0.30	0.34
One can achieve an effective climate protection first of all by means of application of modern technique	0.33	0.22	0.32	0.33	0.32
Germany is a country which will master even the climate change	0.30	0.29	0.28	0.27	0.32
N	1,364 to 1,391	369 to 384	995 to 1,007	642 to 646	353 to 361

This experiment compares the responses to a 7pt scale (only endpoints labeled), a 5pt scale (only endpoints labeled), a 5pt scale (all categories labeled) and a 5pt scale (branching format). Satisficing = "low" means that the survey questions were answered "very conscientiously", while satisficing = "higher" means that the questions were answered less conscientiously than very conscientiously. The entries in this table indicate that the observed frequency distributions deviate less from the distributions, we would have to expect in case of statistical independence of responses across these scale formats, if the survey questions on environmental behavior are answered very conscientiously. This means that we come a little bit closer to the ideal case in which it doesn't make a difference which of the scale formats is used. Note that the distinction according to satisficing refers to the telephone mode only. Data source: PPSM spring 2011 panel survey PS₁₄.

Table A4.7 LCA: Conditional response probabilities used to reveal response styles

	7pt scales				5 pt scales		
	Class 1	Class 2	Class 3		Class 1	Class 2	Class 3
The questions were in part difficult to answer.							
1	0.484	0.296	0.640				
2	0.166	0.248	0.091	1	0.652	0.483	0.740
3	0.112	0.120	0.035	2	0.123	0.252	0.080
4	0.078	0.106	0.048	3	0.101	0.136	0.060
5	0.067	0.144	0.065	4	0.065	0.095	0.059
6	0.040	0.071	0.048	5	0.059	0.035	0.061
7	0.051	0.015	0.074				
The questions were clear and precise.							
1	0.057	0.005	0.026				
2	0.057	0.027	0.005	1	0.105	0.028	0.025
3	0.063	0.026	0.009	2	0.048	0.047	0.017
4	0.052	0.029	0.009	3	0.090	0.066	0.016
5	0.088	0.104	0.007	4	0.094	0.195	0.048
6	0.161	0.337	0.065	5	0.662	0.664	0.895
7	0.523	0.472	0.878				
The questions have made me ponder.							
1	0.443	0.048	0.256				
2	0.250	0.135	0.069	1	0.632	0.158	0.216
3	0.149	0.138	0.054	2	0.158	0.227	0.080
4	0.089	0.199	0.094	3	0.123	0.313	0.162
5	0.043	0.293	0.163	4	0.035	0.242	0.177
6	0.005	0.163	0.136	5	0.052	0.061	0.365
7	0.021	0.023	0.226				
The questions were interesting.							
1	0.143	0.001	0.019				
2	0.197	0.000	0.003	1	0.296	0.003	0.001
3	0.338	0.039	0.017	2	0.256	0.076	0.000
4	0.208	0.202	0.051	3	0.335	0.410	0.076
5	0.073	0.451	0.126	4	0.026	0.435	0.213
6	0.018	0.274	0.239	5	0.088	0.077	0.711
7	0.023	0.034	0.545				
I enjoyed the survey.							
1	0.149	0.004	0.009				
2	0.212	0.016	0.003	1	0.276	0.002	0.015
3	0.253	0.067	0.012	2	0.197	0.130	0.009
4	0.225	0.216	0.046	3	0.337	0.447	0.076
5	0.087	0.389	0.112	4	0.082	0.367	0.213
6	0.051	0.263	0.221	5	0.107	0.055	0.688
7	0.023	0.046	0.596				
The Interview seemed a bit lengthy to me.							
1	0.148	0.063	0.631				
2	0.142	0.263	0.098	1	0.260	0.104	0.560
3	0.143	0.149	0.024	2	0.109	0.252	0.129
4	0.123	0.214	0.027	3	0.151	0.306	0.112
5	0.119	0.182	0.089	4	0.103	0.250	0.095
6	0.119	0.103	0.055	5	0.377	0.088	0.105
7	0.205	0.026	0.076				
The time passed faster than I thought.							
1	0.252	0.023	0.069				
2	0.187	0.093	0.021	1	0.371	0.048	0.089
3	0.163	0.119	0.019	2	0.126	0.167	0.028
4	0.118	0.198	0.031	3	0.110	0.292	0.080
5	0.078	0.259	0.054	4	0.072	0.333	0.117
6	0.114	0.242	0.128	5	0.322	0.160	0.687
7	0.088	0.065	0.677				
Entropy = 0.79; N = 2,833				Entropy = 0.74; N = 2,897			

A German-speaking version of this table is given in Engel et al. 2012: 308f. All scales treated as ordinal scales (Logit-Link). Data source: PPSM recruitment interviews (samples 1 and 2)

Table A4.8: Satisficing behavior (in percent), by response style

Response style	Class Size*	Scale	Satisficing				N
			low	middle	high		
In percent							
1 Endpoint (weak)†	19.0	5pt.	42.2	31.4	26.3	100	547
2 Midpoint	42.2		51.3	32.0	16.6	100	1,214
3 Endpoint (none)	38.8		57.1	28.1	14.8	100	1,117
All	100		51.8	30.4	17.8	100	2,878
1 Endpoint (weak)	25.0	7pt.	41.5	33.0	25.5	100	703
2 Midpoint	40.6		50.4	33.3	16.4	100	1,144
3 Endpoint (none)	34.5		54.3	27.4	18.3	100	971
All	100		49.5	31.2	19.3	100	2,818

*Column percents. Classification of respondents is based on their most likely latent class membership.

†Cognitive anchor (tendency to let one's response differ from this anchor).

Table A4.9 Happiness and satisfaction with current life: the impact of naming the midpoint of an 11pt scale on response behavior when the response format is read to the respondent in a telephone interview

↓ Survey Mode*	Estimated mean value if response format is:			Happiness	
	Midpoint not named	Midpoint named		b	Sig.
FL [FL (+CP)]	7.6	7.5	Mode		0.01
CP [CP]	7.7	7.4	Response format mode		0.25
O [O]	7.2	7.4	Propensity (panel survey)	0.67	0.05
FL [FL (+CP) or O]	7.9	7.6	Propensity (recruitment)	-0.10	0.00
CP [CP or O]	7.5	8.0	Satisficing	-0.14	0.00
O [FL (+ CP) or O]	7.2	6.6	N=3,951; Adj. R ² =0.013		
				Satisfaction	
FL [FL (+CP)]	8.0	8.0	Mode		0.00
CP [CP]	8.0	7.8	Response format mode		0.29
O [O]	7.3	7.1	Propensity (panel survey)	1.34	0.00
FL [FL (+CP) or O]	8.3	7.9	Propensity (recruitment)	-0.07	0.01
CP [CP or O]	8.4	8.1	Satisficing	-0.12	0.00
O [FL (+ CP) or O]	7.4	6.7	N=3,953; Adj. R ² =0.027		

*Survey mode employed [survey mode(s) offered in the preceding recruitment interview]: FL = Fixed line network; CP = Cell phone; O Online. The tables displays mean happiness/satisfaction values for the respective (mode × response-format group) in case the respective group means of response propensity and satisficing are inserted in the prediction equation. Propensity scores are computed with respect to auxiliary variables using logistic regression equations. Please refer to Engel (2013)² for a description of the sets of auxiliary variables used for that purpose. Satisficing is measured by interviewer-ratings as obtained from the recruitment interviews and as already described above. Data sources: Pooled PPSM 2011 panel surveys S₂₁ and S₁₂.

² Engel, U. (2013) *Access Panel and Mixed-Mode Internet Survey. PPSM Panel Report.*
<https://www.viewsandinsights.com/videoportal>