

Brian Kleiner, Oliver Lipps, and Eliane Ferrez

Satisficing and language proficiency

Lausanne, October 2012

FORS Working Papers

2012-1

FORS Working Paper series

The FORS Working Paper series presents findings related to survey research, focusing on methodological aspects of survey research or substantive research. Manuscripts submitted are papers that represent work-in-progress. This series is intended to provide an early and relatively fast means of publication prior to further development of the work. A revised version might be requested from the author directly.

Further information on the FORS Working Paper Series can be found on www.fors.unil.ch.

Copyright and Reserved Rights

The copyright of the papers will remain with the author(s). Formal errors and opinions expressed in the paper are the responsibility of the authors. Authors accept that the FORS reserves the right to publish and distribute their article as an online publication.

FORS may use the researcher's name and biographical information in connection with the advertising and promotion of the work. For any comment, suggestion or question on these guidelines, please do not hesitate to contact us (paperseries@fors.unil.ch).

Editorial Board

Peter Farago

Brian Kleiner

Oliver Lipps

Georg Lutz

Isabelle Renschler

Carmen Borrat-Besson

Valérie-Anne Ryser

Marlène Sapin

Robin Tillmann

Michèle Ernst Staehli

Boris Wernli

Responsible editor: Marieke Voorpostel

How to cite this document:

Kleiner, B., Lipps, O., & Ferrez, E. (2012). Satisficing and language proficiency. *FORS Working Paper Series*, paper 2012-1. Lausanne: FORS.

Acknowledgements

We would like to thank Marieke Voorpostel and Boris Wernli for their very helpful feedback, as well as Stefano Regazzoni and Ludovic Jaccard for their careful work in question coding.

ISSN 1663-523x (online)

FORS

c/o University of Lausanne, Géopolis

1015 Lausanne

Switzerland

E-mail: paperseries@fors.unil.ch

© 2012 Brian Kleiner, Oliver Lipps, Eliane Ferrez



Summary

Especially in multicultural diverse populations, unit nonresponse can be attributed in part to language barriers, if no proxies are available or willing to answer on behalf of people, and if there is no translated version of the survey into the native language of the potential respondent. In such cases, sampled individuals may simply opt out of the survey.

However, little is known about the data collected from sampled individuals who are not native speakers of the survey language and who decide to participate in surveys. Specifically, there is little in the literature addressing the data quality of survey respondents who may not be fully proficient in the language of survey administration. Our research examines the question of whether the data of nonnative speakers may be compromised to some extent due to problems of comprehension of survey questions, especially questions that are more complicated both in terms of content and form.

Operationalising data quality within the terms of Krosnick's satisficing theory, we compared native and nonnative speaking groups in two large-scale Swiss national surveys with respect to a set of dependent measures, including item nonresponse, extreme responding on scales, recency effects, and straightlining. The characteristics of the selected surveys and the questions within them allowed us to distinguish possible effects of language comprehension from other effects on data quality, such as age and level of education, and effects of motivation.

Our results indicate that foreigners from neighbouring countries that speak one of the three Swiss national languages (i.e., from Germany, Austria, France, and Italy) tended to produce data on the same level of quality as the Swiss nationals. On the other hand, there was significantly poorer data quality for the different dependent measures for foreign populations in Switzerland that are *not native speakers* of the available survey languages (German, French, Italian). However, in addition to language proficiency, the poorer quality may be attributable in part to respondent motivation.

Satisficing and language proficiency

Brian Kleiner, Oliver Lipps, Eliane Ferrez¹

1. Respondent language proficiency and data quality in surveys

This paper examines the question of whether the data of non-native speakers of available survey languages may be compromised to some extent due to problems of comprehension of survey questions, especially questions that are more complicated in terms of content and/or form. Little is known so far about the data collected from sampled individuals who are not native speakers of a survey language who opt to participate in surveys. Specifically, there is little in the literature addressing the data quality of survey respondents who may not be fully proficient in the language(s) of survey administration. Given increasing migration and linguistic diversification in many countries, it is important to examine further whether the data provided by native and non-native speakers of available survey languages are of the same quality. More generally, there is the question of the quality of data collected from foreign groups who might not be as motivated as native ones to respond to surveys. These questions have implications for research that aims to assess the living conditions and/or opinions of minority populations in relation to majorities.

We examine data from two large-scale Swiss surveys within the framework of satisficing, comparing the responses of Swiss respondents with two foreign groups: 1) respondents from neighbouring countries who speak natively one of the three national languages (German, French, Italian), and 2) respondents from other countries who do not speak natively one of the three national languages. Examining several measures of satisficing, such comparisons allow us disentangle to some extent the effects of language proficiency and motivation among the two foreign respondent populations.

¹ FORS – Brian.Kleiner@fors.unil.ch

1.1. Related research

There is ample evidence in the literature that elderly people, very young people, and people with lower education levels tend to have more problems with question comprehension and/or answering (e.g., Yan and Tourangeau 2008; Fricker et al. 2005; Narayan and Krosnick 1996), as measured by a variety of dependent variables, such as item nonresponse, extreme responding, and response times. In addition, some work has shown that comprehension of survey items is correlated with cultural or racial/ethnic differences (Holbrook et al. 2006; Harkness et al. 2003; Warnecke et al. 1997). Still other research shows that there are cultural styles of responding to survey questions that may create bias in data across groups (Johnson et al. 2006a,b; Johnson et al. 1997), while others have shown that the language of survey responding itself may create a sort of cultural lens for interpreting and responding to questions (Peytcheva 2008; Johnson et al. 2006a).

Moreover, there has been little research on the effects of language proficiency on question comprehension for people who are not native speakers of survey languages. This may be because it is generally assumed by survey researchers that such people – once they opt to participate in surveys – are sufficiently competent in the available language of survey administration and therefore understand survey questions in the same way as native speakers. One exception is the work of Gray et al. (2011), which showed by way of cognitive interviews that non-native speakers of English in the U.K. had serious problems in understanding certain questions.

On the other hand, there is a large body of literature that focuses on question complexity and how this may influence comprehension and responding to questions. In general, research suggests that questions that are complex in content and in form carry a higher cognitive load for respondents, which may result in misinterpretation of questions, satisficing, or breaking off in a survey. With respect to question content, some work indicates that questions that are subjective (Yan and Tourangeau 2008; Bassili 1996); that are abstract or that have vague concepts (Johnson et al. 2006b; Holbrook et al. 2006); that have infrequent words (Lenzner et al. 2009); that have memory demands (Tourangeau et al. 2000); that are hypothetical (Lenzner et al. 2009); or that require calculation or numeric values (Johnson et al. 2006b; Lenzner et al. 2009;) are generally more difficult.

With respect to question form, research suggests that questions that are longer (Johnson et al. 2006b; Holbrook et al. 2006); that contain more clauses (Yan and Tourangeau 2008); that

are syntactically more complex (Lenzner et al. 2009); or that include negative particles (Saris et al. 2010) are generally harder for cognitive processing and responding. In addition, number and types of response categories may also affect the cognitive burden of a question (Yan and Tourangeau 2008; Tourangeau et al. 2000), and the placement of a question in a questionnaire may be a factor in question comprehension (Yan and Tourangeau 2008).

Studies have demonstrated that lower educated people and older people may have more difficulties in handling the increased processing demands of more complex questions (Yan and Tourangeau 2008; Krosnick et al. 1996). Along these lines, it may also be the case that non-native speakers have more or less difficulty depending on the relative complexity of survey questions. Again, there is little in the literature that addresses this.

1.2. Theoretical framework and assessing data quality

To assess whether the survey data of non-native speakers might be of poorer quality due to language proficiency, especially for more complex or difficult types of questions, we employed Tourangeau, Rips, and Rasinski's cognitive model for survey responding (2000) and the framework of satisficing (Krosnick 1991). The work of Tourangeau, Rips, and Rasinski identifies four major components of survey responding.² First, respondents must understand and interpret a question's meaning and pragmatic intent. Second, a respondent must retrieve the relevant information from memory. Third, he/she must form a judgement or estimate based on the understanding of the question and the information retrieved from memory. Finally, he/she must provide an acceptable answer that meets the needs of the survey item by mapping judgements or estimations to available response options.

Each of these processes in answering survey questions will carry a certain cognitive load, and questions will vary in terms of which of these components are the most taxing for respondents. Thus, for example, long, complicated, or abstract questions may create additional burden for comprehension, while questions that require digging into long-term memory will burden retrieval processes. Further, questions that demand calculation, evaluation, or hypothetical thinking will place a burden on the third stage. Comprehension and reporting are especially affected by question properties, whereas retrieval and judgement are less based on the question and more on characteristics of respondents (Yan and Tourangeau 2008).

² Earlier influential work modelling the survey response process was done by Cannell, Miller, and Oksenberg (1981), and developed further by Tourangeau and Rasinski (1988).

The idealised path modelled by Tourangeau, Rips, and Rasinski represents the best case scenario, where respondents make a full cognitive effort to carry out the task required of them. However, such full-faith effort on the part of respondents is not a given or even the norm. Krosnick's framework of satisficing³ takes Tourangeau, Rips, and Rasinski's model a step further by addressing specifically respondents' effort-conserving behaviours in the face of heightened cognitive demands.⁴ Satisficing is a responding strategy whereby respondents minimize cognitive effort by providing minimally acceptable answers in relation to the difficulty of questions, their abilities, and their level of motivation.

Krosnick distinguishes weak and strong satisficing. Weak satisficing involves shortcutting on retrieval and judgement processes, while strong satisficing is when these two steps are avoided altogether. Examples of both types include: selecting the last mentioned response option for a question in an orally administered survey ("recency effects"); providing an expected answer (so-called "acquiescence"); providing the same answer across a series of questions with the same options (referred to sometimes as "straightlining"); and declaring that one does not know the answer to a question or does not want to respond ("item non-response").

According to Krosnick's model, there are three possible factors that determine the extent to which satisficing takes place: task difficulty, respondent ability, and respondent motivation. Task difficulty depends on features of the question and its response options, such as question complexity and memory demands, while the survey context can also affect task difficulty, like the rate at which interviewers ask questions, and the internal state of the respondent (e.g., if he/she is tired, rushed).

Second, respondent ability will influence the extent of satisficing. Krosnick defines this primarily in terms of the cognitive skills of respondents. Third, different levels of motivation can affect satisficing, depending on how important a respondent judges his/her participation in the survey (e.g., topic interest, need for cognition, conscientiousness, or likeability of the interviewer).

Following the frameworks of Krosnick and of Tourangeau, Rips, and Rasinski, one would generally expect more satisficing for questions with higher task difficulty that place a greater cognitive burden on able and willing respondents with respect to comprehension, retrieval,

³ The term "satisficing" dates from the 1950s and is a blend of the words "satisfy" and "suffice".

⁴ Earlier work by Krosnick and Alwin (1987; 1988; 1989) introduced Herbert Simon's (1947) concepts of 'satisficing' and 'optimizing' to the field of survey methodology.

judgement, and reporting. Further, there should be more satisficing behaviour among respondents who are less able to cope with the processing demands of more difficult questions.

While we agree that cognitive skills are important in this regard, there is another aspect of ability that seems to have been neglected in Krosnick's framework – language proficiency. An individual who has weaker linguistic competence in a survey language might be predicted to behave in a similar way to one who has poorer cognitive skills, that is, by showing higher levels of satisficing. The processing demands especially connected to question comprehension could easily derail the responding process for non-native speakers, and could lead them to seek a "minimally acceptable answer" for questions that are difficult to understand. This includes notably item non-response, but other forms of satisficing could be involved as well.

If a significant number of non-native speaking respondents have trouble in processing and responding to questions (especially more complex ones), then this should be reflected in respondent data for various selected dependent measures of satisficing. Using data from several large-scale Swiss representative telephone surveys, we hypothesized that non-native speakers of the three national languages (French, German, and Italian) would exhibit higher levels of satisficing, as measured by item-nonresponse, extreme responding, recency effects, and straightlining. We also hypothesized that the effects would be most significant for more complex types of questions that place greater cognitive burden on respondents.

With respect to motivation, one might expect foreign respondents to generally feel less concerned by national surveys on attitudes and living conditions and thus to put less effort into responding. Our third hypothesis was that, beyond questions of language proficiency, foreign respondents would be less motivated in responding and would therefore exhibit higher levels of satisficing than Swiss respondents.

2. Study approach: data and analyses

2.1. Overall approach

Our approach involved analyses of several large-scale representative surveys in Switzerland that permitted comparisons on selected satisficing measures between Swiss respondents,

respondents from neighbouring countries who speak French, German, or Italian natively, and all other foreigners.⁵ Such comparisons would presumably allow us to disentangle effects of language/comprehension from motivational effects of being a foreigner. In addition, several features of these surveys permit assessment of the relative strengths of effects due to language ability and motivation.

To assess the effects of question characteristics and difficulty, we coded survey questions along many of the different dimensions that have been documented in the literature (e.g., on question length, number of clauses, need for memory and calculation, and so on – see below). Our analyses modelled the effects of language in relation to question characteristics using dependent measures such as item nonresponse, extreme responding, recency effects, and straightlining, while controlling for other variables like age and education.

2.2. The data

In order to test our hypotheses, we employed the 2008 Swiss Labour Force Survey (SLFS), and the 2004 Swiss Household Panel (SHP). These specific surveys were chosen for a variety of reasons. First, their relatively large sample sizes allow for sufficient statistical power in comparing the Swiss group with the two foreigner groups. Further, both surveys include the necessary variables for adequately distinguishing (1) Swiss native speakers of the three national languages German, French, and Italian from (2) non-Swiss native speakers of these languages (i.e., those from the neighbouring countries of Germany, France, Austria, Liechtenstein, and Italy), and (3) non-Swiss non-native speakers of these languages.

Second, the surveys include a wide range of question types that allow for examination of aspects of question complexity in relation to respondent ability (including language proficiency) and motivation. Also, the diversity of question types in the two surveys allows for analysis of multiple measures of satisficing, such as item nonresponse, recency effects, and extreme responding. Third, in order to control better for different possible effects on satisficing, both surveys were conducted using the same mode – by telephone.

⁵ Foreigners make up more than 20 percent of the population of Switzerland. Besides those from neighbouring countries, the largest foreign populations are from ex-Yugoslavian countries, Portugal, Spain, and Turkey. Our division of foreigners into two groups was modelled on previous work that focused on unit nonresponse and representation of national minorities in Swiss general population surveys (Laganà et al. 2011). In that work, this interest was in the effects of cultural distance and language on survey participation.

There are several differences between the two surveys that create some interesting potential for analysis as well. Most notably, the SLFS is a highly factual survey, while the SHP includes many subjective and evaluative kinds of questions. Related to this, because the SHP includes many opinion questions concerning life in Switzerland (e.g., questions about politics), some of these may be more likely to be considered as irrelevant for foreigners, which could influence satisficing by reducing motivation. Finally, both of these surveys have special features that allow us to disentangle effects of language proficiency from effects of motivation. In particular, the SHP has a variable where interviewers rate overall respondent understanding of questions, while in the 2008 SLFS most foreign respondents had the option of doing the survey in their own native language⁶, and thus we can compare those foreigners who participated in their own language with those who did so in a non-native language with regard to the dependent measures of satisficing. Table 1 lays out some key elements of the two surveys. It should be noted that the data used for the SHP was from a refreshment sample of 3,389 cases, and so all of these respondents were participating in the survey for the first time.

Table 1: Features of the SHP (2004) and SLFS (2008)

	Swiss Household Panel (2004): refreshment sample	Swiss Labour Force Survey (2008)
Composition	3,389 cases: 2,956 Swiss; 221 from neighbouring countries; 212 from other countries	47,390 cases: 29,927 Swiss; 7,740 from neighbouring countries; 9,723 from other countries
Scope	Swiss residential adult population	Swiss residential adult population
Themes	Health, well being and attitudes, politics, social networks, economics, education, and labour	Structure of the labour force and employment behaviour patterns
Types of questions	Subjective, factual, demographic	Factual, demographic
Survey type	Longitudinal	Cross-sectional
Available survey languages	German, French, Italian	German, French, Italian, Serbo-Croatian, Albanian, English, Portuguese, Turkish
Survey mode	Telephone (CATI)	Telephone (CATI)

2.3. Question coding

Following findings from the literature that point to aspects of question complexity, we selected a set of question codings, to the extent that it was feasible given the questions available in the surveys. Questions were divided into three types: The demographic ones referred to stable social characteristics of respondents or related individuals, such as age,

⁶ English, Serbo-Croatian and Albanian translations were added to the SLFS in 2003, while Turkish and Portuguese were added in 2005.

sex, level of education, income, and family size. The factual/behavioural questions referred to facts, conditions, or behaviours of or around respondents (e.g., "How frequent are your contacts with your partner?"). The evaluative (attitude/belief/expectations) questions involved some appeal to the judgement of respondents (e.g., "Would you like to change the number of hours that you work each week?").

The questions were also coded according to their response format. A first type was those that require a yes or no answer. The multiple response questions were those with two or more response options, each labelled (not including yes/no questions). The scale questions included a sequence of ordered numbers (at least four) in their response format, with the extreme points labelled and the points in between not all labelled. It should be noted that this kind of question, frequent in the SHP, was not used in the SLFS. Open questions had no specific options available to choose from. Related to this, the number of response categories (spoken or explicitly made available to respondents by the interviewers) was also coded for each question. The length and the structure of questions were coded, by counting the number of sentences, clauses⁷, and words in it (including those within response options that are spoken out loud by interviewers). This was done for German, French, and Italian separately.

Questions were coded for whether they required counting or some degree of calculation on the part of respondents (e.g., "For how many hours or days per week does a close relative normally look after the children?"), as well as for hypothetical thinking about a possible state of affairs (e.g., "Let's assume that you can't find a part-time job. In that case would you be prepared to accept a full-time job?"). Questions were also coded for whether they required searching in long-term memory for something that occurred in the past (e.g., "How many times in the last 10 years were you unemployed?"). We distinguished between intensive searching, moderate searching, and little or no searching in long-term memory. For the SLFS questions we took into account the presence of negation in the formulation, which should add some complexity to its comprehension. No forms of negation were found for any of the SHP questions.

In addition, questions were coded for characteristics that might mitigate any resulting differences between the country groups for the dependent measures of satisficing, such as relevance. Foreigners might indeed feel that certain questions do not really apply to them and so may not be motivated to answer (e.g., "Overall, how satisfied are you with the way in

⁷ We defined them as follows: A sentence included at least one subject and one verb. A clause included at least and *at maximum* one subject and one conjugated verb.

which democracy works in our country, if 0 means "not at all satisfied" and 10 "completely satisfied"?). The sensitivity or social desirability of the question was also coded. A question might be embarrassing, highly personal or of a sensitive nature for some respondents, who might think that answering will put them in a vulnerable position (e.g. "Have you ever stopped working for longer than 6 months to study, because you were unemployed, or for family, health or other reasons?"). Similarly, a question may have a socially desirable response option where there is a clearly socially preferable answer (e.g. "Did you complete your eighth or ninth year of compulsory schooling?"). Given the difficulties of clearly distinguishing sensitivity and social desirability, we combined these into a single code.

The placement of questions in the survey (i.e., the first third, the second third, or last third of questions) was coded, since whether a question is earlier or later in a survey interview could influence satisficing behaviours, given the effects of fatigue or declining motivation. Finally, we coded questions as to whether they were eligible for examining recency effects, as well as extreme and mid-5 responding. Questions that could involve recency effects were defined as those with three or more labelled response options. Extreme responding and mid-5 responding were examined for 11-point scale questions, only possible for the SHP⁸, where only the extreme points were labelled. All questions in the SHP with 11-point scales were thus coded as eligible for these analyses.

Variables that were constructed or not orally administered during the survey were excluded from analyses. In the end, we coded 395 variables for the SHP and 404 variables for the SLFS. The questions were coded separately by two people. Their codings were then compared, discussed, and resolved if necessary. For the most part, the coding of the questions was a mechanical exercise (e.g. number of sentences, words), but even for the more subjective question characteristics (e.g., memory, question type), there was a high degree of agreement, and all disputed cases were resolved.

2.4. Analyses

The dependent measures of satisficing included item nonresponse, recency effects, extreme responding, mid-5 responding, and straightlining (each defined below). For each measure, the level of satisficing was defined as the proportion of variables affected by one of the satisficing measures for each individual, that is, the number of satisficing answers over the number of all answers for the variables coded. Since the dependent variable is a proportion

⁸ The 2008 SLFS did not contain any questions with scales.

of satisficing responses, we employed general linear models with the assumption of a binomially distributed dependent variable and a logit link function (Papke and Wooldridge 1996). The robust option was used to estimate unbiased standard errors.

While both surveys allowed for assessment of item nonresponse as an indicator of satisficing, the other measures adopted differed across the two surveys according to the distribution of question types. For example, in the SHP there were too few variables that could allow for a reliable examination of recency effects, the risk being that the contents of the response options for a small number of questions could be a stronger determinant of selecting a last-spoken option than satisficing. Also, given that the SHP included many sequences of Likert-type scale questions (unlike the SLFS), three of the dependent variables were only applicable for the SHP.

Table 2: Dependent measures of satisficing for analyses

	Swiss Household Panel	Swiss Labour Force Survey
Item nonresponse	x	x
Recency effects		x
Extreme responding	x	
Mid-5 responding	x	
Straightlining	x	

Item nonresponse was defined as the proportion of "don't know" and "refuse to answer" responses for individuals across all asked questions. The measure of recency effects was based on the proportion of responses that selected the last mentioned item from a list of three or more items for all relevant questions. Extreme responding was defined as the number of responses that were at the extreme beginning or end points of 11-point scales divided by the total possible number of such cases for relevant questions. Mid-5 responding was the proportion of responses at the centre point of 11-point scales. Finally, straightlining was defined as the proportion of cases where responses were identical (i.e., same selected response category) across a series of questions with the same stem and response format.

The "country" variable was constructed as follows. Respondents claiming Swiss nationality or who had lived in Switzerland since at least the age of 2 were categorised as "Swiss". Otherwise, those who claimed nationality from any of the countries of Germany, Austria, Liechtenstein, France, or Italy were treated as "Neighbours", and all remaining respondents were deemed to be "Other foreigners".⁹ This constructed variable serves as a proxy for the native language of respondents, with the first two groups presumably fully proficient in one of

⁹ We excluded cases in the two surveys where a proxy was used to help complete the survey.

the three national languages in which the surveys were administered (German, French, and Italian).¹⁰ Respondents in the "other foreigner" group may or may not have been fully proficient in one of the three languages, depending on different factors (e.g., length of time living in Switzerland, country of origin). For the SLFS, we further divided the "other" group into those who conducted the survey in one of the three national languages (n=5,991) versus those who conducted the survey in some other language (n=3,732).

Other independent variables included level of education, age, and sex. For the SHP, we made use of the variable corresponding to respondent understanding, where interviewers had to code for each respondent whether they thought the understanding of questions was "good", "fair", or "poor". As will be shown later, this variable allowed us to distinguish to some extent effects of language proficiency and motivation using regression analyses.

For the analyses themselves, in addition to descriptive statistics (means and standard errors distinguished by country group), we built regression models using the above mentioned dependent and independent variables, including variables related to question complexity and relevance.

3. Results

Table 3 presents the coefficients and overall significance findings for the five different dependent measures of satisficing, by the country variable, with the Swiss respondents as reference group. To determine significance, each dependent measure was included in a regression model controlling for age, level of education, and sex. For the SLFS, the "other" category is divided into those who responded in one of the three national languages and those who responded in some other language (presumably their own native language).

With the exception of mid-5 responding, the results indicate a significant level of satisficing in the predicted direction across the different measures for the foreign groups.¹¹ For both

¹⁰ These assumptions about language hold true generally but are not without exception. For example, a small number of respondents in the "other foreigner" group may have been from a non-neighbouring country where French, German, or Italian was their native language. In addition, there was a small percentage of cases of Swiss or "neighbour" respondents who answered the survey in a non-native language (e.g., a Swiss German who responded in French, or an Italian who responded in German).

¹¹ It should be noted that the regression model results across the different satisficing measures were entirely consistent and in a predictable direction for age and education groups. The most educated respondents were less likely to satisfice than less education respondents, while older respondents were more likely than younger ones to do so.

surveys, the other foreigner group that completed the survey in German, French, or Italian had significantly higher levels of item nonresponse than the Swiss group, again controlling for age, level of education, and sex. For the SLFS, the other foreigner group that completed the survey in their own native language were no different than the Swiss or Neighbours with respect to item nonresponse. Within the SHP, Neighbours also had higher levels of item nonresponse than the Swiss, but to a lesser degree than the Other foreigners. Also within the SHP, Other foreigners were more likely than the Swiss and neighbour groups to respond to the extremes of scales and to straightline.

Although the results from the two surveys indicate that the Other foreigners who completed the survey in one of the three national languages were more likely to satisfice than the other country groups, it would be premature to conclude from this that it was due to language proficiency. It appears that motivation may have also been playing a role among the foreign groups, since the Neighbours, who responded to the survey in a native language, were also different from the Swiss for item nonresponse in the SHP. Similarly, in the SLFS, all three foreigner groups showed significantly higher levels of recency effects than the Swiss group. The conclusion from this is that foreign groups may have satisficed more in part because they felt less concerned by the survey questions. To disentangle the effects of language proficiency and motivation among the foreign groups, we pass now to results from more in-depth analyses of item non-response, with consideration of question characteristics.

Table 3: Binomial regression (logit link) results in coefficients for measures of satisficing, by "country"

	Swiss Household Panel	Swiss Survey	Labour Force
Item non-response			
Neighbours	.467**		-.006
Others 1 (de_fr_it)	.969**		.110**
Others 2 (sc_al_en_tu_po)	(n/a)		.003
Recency effects	(n/a)		
Neighbours			.086**
Others 1 (de_fr_it)			.090**
Others 2 (sc_al_en_tu_po)			.079**
Extreme responding			(n/a)
Neighbours	-.025		
Others 1 (de_fr_it)	.188**		
Others 2 (sc_al_en_tu_po)	(n/a)		
Mid-5 responding			(n/a)
Neighbours	-.054		
Others 1 (de_fr_it)	.039		
Others 2 (sc_al_en_tu_po)	(n/a)		
Straightlining			(n/a)
Neighbours	.135		
Others 1 (de_fr_it)	.496**		
Others 2 (sc_al_en_tu_po)	(n/a)		
N	3,389		47,390

NOTE: "Swiss" as reference group, controlling for age, level of education, and sex.

** p<.01, * p<.05

Item nonresponse

The measure of item nonresponse is well suited for examining effects of language proficiency and motivation among foreign groups, since answering "don't know" or refusing to answer is an ideal way to minimise effort and save face for questions that are difficult to understand or simply not relevant for respondents.¹² Indeed, item nonresponse can have both cognitive and motivational causes and meanings (Loosveldt et al. 2002), and so this measure, in relation to other specific variables, can help to disentangle these different factors. Also, not responding to questions is a direct reflection of comprehension problems (de Leeuw et al. 2003), and more so than the other measures of satisficing. To take an example, while responding to the extremes of scales may indicate a reduction of effort in responding, it is not more likely to result from poor question understanding than non-extreme responding. As will be shown in the following progression of results, the characteristics of questions and several special features of the two surveys allow for distinguishing these different drivers of item nonresponse.

¹² "Don't know" responses were three times more common than "refusals" in the SHP and about twice as common in the SLFS.

Table 4 shows the mean item nonresponse across the country groups for the two surveys. It is not surprising that there was a slightly higher level of item nonresponse in the SHP than in the SLFS, since the SHP contains many subjective opinion type questions (e.g., on politics), compared to the SLFS, which is almost entirely factual.

Table 4: Mean item-nonresponse for the SHP and SLFS

Item nonresponse	SHP			SLFS		
	N	mean % item-nr	Standard error	N	mean % item-nr	Standard error
Swiss	2,956	.0131	.0004	29,927	.0089	.0001
Neighbours	221	.0187	.0016	7,740	.0087	.0002
Others 1 (de_fr_it)	212	.0295	.0030	5,991	.0099	.0003
Others 2 (sc_al_en_tu_po)	n/a	n/a	n/a	3,732	.0093	.0004
All	3,389	.0145	.0004	47,390	.0090	.0001

Table 5 shows that respondent understanding was highly correlated with other foreigner status. Over a third of the Other foreigners were judged to have fair or poor understanding by interviewers. By adding this variable to the regression model, we may determine the extent to which language proficiency accounts for levels of item nonresponse.

Table 5: Percentages of respondents with good, fair, and poor understanding in the SHP, by country group.

	SHP		
	Swiss	Neighbours	Others 1 (de_fr_it)
Good understanding	89%	83%	62%
Fair understanding	10%	16%	28%
Poor understanding	1%	1%	9%

Note: Data are from the SHP 2004 refreshment sample

Table 6 presents binomial regression model results for item nonresponse that includes the country variable as well as the control variables age¹³, level of education, and sex (referred to in the table as "SHP1"). In order to be able to compare coefficients across the nested models, we use average marginal effects (Mood 2010). The educ2 and educ3 categories are the middle and highest educated groups, respectively. Next, adding the respondent understanding variable to the model ("SHP2"), we see that while better respondent understanding corresponds to significantly less item nonresponse, it diminishes but does not eliminate other foreigner status as an explanatory factor.¹⁴ Here then is a second indication that language proficiency is involved in other foreigner satisficing, but that it is not in itself a

¹³ The reference group for age was respondents younger than 60, since we expected a certain homogeneity within this group with respect to satisficing, when controlled for the other socio-demographic variables.

¹⁴ On the other hand, the effects of education, especially for the intermediate level, are reduced considerably.

sufficient factor. In other words, motivation appears to be playing a role in the higher level of satisficing among the foreign groups.

Table 6: Binomial regression (logit link) model results for item nonresponse

Item-nonresponse	SHP0	SHP1	SHP2
Average marginal effects			
Neighbours	.0051**	.0061**	.0053**
Others 1 (de_fr_it)	.0116**	.0127**	.0092**
Educ2		-.0011	-.0001
Educ3		-.0036**	-.0019*
Male		-.0051**	-.0053**
Age 60-69		.0046**	.0041**
Age 70+		.0099**	.0077
Good respondent understanding. ¹⁵			-.0074**
Pseudo-log likelihood	-207.4	-204.1	.202.1

Note: Data are from the SHP 2004 refreshment sample. N=3,382

We hypothesized that difficulties in question understanding would be reflected in more satisficing for more complex types of questions, especially for respondents who conducted the survey in a non-native language. For both the SHP and SLFS we included question characteristics in the regressions, following our coding scheme, and found that particular ones significantly contributed to item nonresponse. Table 7 presents the question characteristics for which there was a consistently significantly higher level of item nonresponse for Other foreigners compared to the Swiss and neighbour groups for one or both surveys. The results show that there were no discernible patterns within and across the two surveys, especially with regard to what are presumably more complex question features.¹⁶

¹⁵ Coded 0=poor understanding, 1=fair understanding, 2=good understanding.

¹⁶ There were very few cases in which the Other foreigners had lower mean item nonresponse than the Swiss or Neighbors for particular questions characteristics.

Table 7: Question characteristics for which Other foreigners had significantly higher item nonresponse than Swiss and neighbour groups

	SHP	SLFS
Questions requiring some retrieval from long-term memory*	x	
Questions requiring heavy retrieval from long-term memory*	x	
Questions requiring calculation*	x	
Factual questions	x	x
Demographic questions	x	
Evaluative questions*		x
Questions with three response categories*	x	
Yes/no questions	x	
Single sentence questions		x
Questions with 20 words or more*		x
Questions with multiple labelled responses*		x
Questions with early placement in survey	x	
Questions with middle placement in survey*	x	x
Questions with late placement in survey*		x

*Question characteristics that would be predicted to be more difficult.

In order to evaluate the importance of these question characteristics in relation to each other and respondent characteristics, we created a model with the SHP using all relevant variables and ran a regression model including demographic and question variables that were shown to be significant in prior regressions, as well as the respondent understanding variable and the variable on question relevance. The dependent variables were binary (item-nonresponse vs. substantive answer given) over all relevant variables coded. The data set for this model was extended such that each individual had as many records as relevant question variables, resulting in 651,691 substantive (i.e., excluding not applicable questions) respondent-question combinations. The data structure was then such that both individual respondents were clustered within question variables and question variables within respondents (a “cross-classified” data structure; see Fielding and Goldstein 2006). The data structure can be visualized as follows:

Figure 1: Question-respondent association in the cross-classified data structure

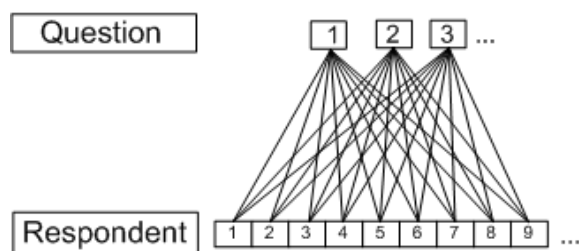


Table 8: Cross-classified binary MCMC estimated regression model (logit link) results for item nonresponse.

	SHP0	SHP1	SHP2	SHP3
Neighbours		.485**	.435**	.433**
Others 1 (de_fr_it)		1.016**	.751**	.749**
Age 60-69		.393**	.357**	.354**
Age 70+		.824**	.658**	.656**
Educ2		-.087*	-.007	-.007
Educ3		-.288*	-.160*	-.160*
Male		-.394**	-.425**	-.425**
Good respondent understanding			-.628**	-.631**
Not relevant questions				1.807*
Questions requiring some retrieval from long-term memory				-.377*
Questions requiring heavy retrieval from long-term memory				1.398*
Questions requiring calculation				.013
Questions with middle placement in survey				.545*
Questions with late placement in survey				.592*
Yes/no questions				-1.107*
Questions with three response categories				-.773*
Demographic questions				-.050
Factual questions				-.449
Constant	-5.199**	-5.083**	-3.990**	-4.003**
Bayesian DIC statistics ¹⁷	81,345	79,472	78,693	78,687
Variance (question level) ¹⁸	2.233**	2.269**	2.284**	1.430*

Note: Data are from the SHP 2004 refreshment sample. *(**)= =coefficient at least 2(10) x standard error.

As shown in table 8, including question characteristics together in a single model (SHP3) reveals that they are very much interrelated, with some becoming non-significant in explaining item nonresponse, when controlling for the others (e.g., calculation, and demographic and factual questions). In a few cases the effect is even reversed, with features that significantly lead to less item nonresponse, as in the case of yes/no questions, questions with three response categories, and questions requiring some retrieval from long-term memory. Only heavy retrieval from long-term memory and middle or late placement of questions in surveys increase item nonresponse. Notably, questions that are less relevant for foreigners significantly increase the likelihood of item nonresponse, although the effect is small.

Taken as a whole, the relatively low significance levels suggest that these question characteristics are less important than respondent characteristics in accounting for item

¹⁷ The Bayesian Deviance Information Criterion (DIC) is an MCMC penalised goodness of fit measure. It is equivalent to the Akaike Information Criterion (AIC) used in maximum likelihood estimation (from Stata help on "runmlwin"). Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best

¹⁸ In logistic models the variance at the lowest level is constrained to the area under the logistic curve ($\pi^2/3 \sim 3.29$); see Snijders and Bosker (1999).

nonresponse. Indeed, adding the question characteristic variables to the model appears to do little to change the contributions of the different respondent characteristics in explaining variance for item nonresponse (see the development of the DIC statistics in table 8).

Finally, it should be noted that we also examined the possible effects of acculturation on item-nonresponse, but found surprisingly that foreigner status interacted with the number of years living in Switzerland was not significantly correlated with this dependent measure of satisficing.

4. Discussion and conclusions

In summary, the foreigners who participated in the surveys in a non-native language indeed had stronger satisficing behaviours compared to respondents who answered using a native language. For example, they were more likely in the SHP to straightline and to answer at the extremes of scales than both the Swiss and respondents from neighbouring countries. In the SLFS, Other foreigners who conducted the survey in a non-native language were more likely than the Swiss, Neighbours, and Other foreigners who conducted the survey in their own native language not to respond to questions. This result is consistent with our first hypothesis that poorer language proficiency should show higher levels of satisficing.

However, the results are not entirely unambiguous. The Neighbours also showed higher levels of satisficing than the Swiss for one of the measures (item nonresponse), an indication that reduced motivation of foreigners may also be playing a role. To disentangle possible effects of language proficiency and motivation, we examined more closely item nonresponse, with consideration of question characteristics. The findings from the SHP suggest that question understanding is indeed closely linked to other foreigner status, but that this still does not account fully for differences in item nonresponse between the foreign groups and the Swiss. Further, the (non)relevance of certain questions for foreign population of respondents explained a small part of their higher tendency for nonresponding.

Question characteristics following our coding scheme do not seem to have been very important in predicting higher levels of item nonresponse among the other foreign group. Rather, our regression modelling indicates that respondent characteristics (such as country, level of education, age, and sex) are far stronger determinants of item nonresponse than question features.

Interpreting these findings within the satisficing framework, we would argue that in the context of the SHP 2004 and SLFS 2008, language proficiency played a significant role in satisficing along the lines of respondent "ability". As corroborated by the interviewer assessments of respondent understanding in the SHP, many in the other foreigner category probably struggled to grasp the meanings of questions and to provide appropriate answers. This appears to have been a problem across the board, and not just for more complex types of questions (i.e., those with a higher level of task difficulty), although it is possible that our coding categories were not sensitive to the kinds of challenges faced for particular question types. Recalling Krosnick's "ability" factor within the satisficing framework, we would argue based on our findings that this concept should be broadened beyond cognitive ability to include language ability as well.

Moreover, motivation also seems to explain some of the satisficing behaviours among foreigners. This is not surprising, since foreigners may generally feel as though the survey is not really for or about them as individuals. They may also feel less concerned by specific survey questions and topics. The effect of reduced motivation thus may be various manifestations of satisficing. While we suggest that motivation may be an important factor, we also are aware that an additional factor may be that certain culture-specific notions may simply be unfamiliar to some foreigner, and this might increase item nonresponse as well.

More study is needed to confirm that satisficing may be stronger among foreign populations, and to carefully distinguish the influences of ability/language proficiency from motivation. It may be the case that motivational factors also account for observed differences in satisficing across education and age groups, in that survey questions are generally geared, unconsciously or not, toward a "normal" respondent, that is, one most similar to survey designers themselves (highly educated, middle-aged, etc.). With respect to question complexity, there may be finer methods to identify specific question characteristics that tend to increase task burden and that reveal empirical effects in different forms of satisficing for different groups.

If the findings of this paper hold, what are the implications for researchers? One conclusion would be that survey designers who are concerned about foreign groups should take steps to reduce the pressures that may lead to increased satisficing. For example, they can provide surveys in other languages, or ensure that survey topics and questions are relevant and motivating for all respondents. Naturally, survey questions should be as clear and simple as possible to promote understanding, not only for older or less educated respondents, but also for those who have limited language proficiency.

5. References

Bassili, J. N. (1996). The how and the why of response latency measurement in telephone surveys. In N. Schwarz, and S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. (319–346). San Francisco: Jossey-Bass.

Cannell, C.F., Miller, P.V. and Oksenberg, L. (1981). Research on interviewing techniques, in S. Leinhardt (ed.), *Sociological Methodology*, San Francisco: Jossey-Bass: 389-437.

Fielding, A. and Goldstein, H. (2006). *Cross-classified and Multiple Membership Structures in Multilevel Models: An Introduction and Review*. Research Report RR791, department for education and skills, University of Birmingham

Fricker, S., Galesic, M., Tourangeau, R., and Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly* 69: 370–392.

Gray, M., D'Ardenne, J., Balarajan, M., and Uhrig, N. (2011). *Cognitive testing of Wave 3 Understanding Society Questions*. Understanding Society Working Paper Series, No. 03.

Harkness, J., Van de Vijver, F. J.R., & Mohler, P. Ph. (2003). *Cross-Cultural Survey Methods*. 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, 4: 565–595.

Johnson, T. P., Holbrook, A. L., and Cho, Y. I. (2006a). The effects of acculturation on survey question comprehension among Latino respondents in the U.S. *ASA Section on Survey Research Methods*: 3186-3192. Downloaded 15-06-12: <http://www.amstat.org/sections/srms/proceedings/y2006/Files/JSM2006-000500.pdf>

Johnson, T. P., Cho, Y. I., Holbrook, A. L., O'Rourke, D. P., Warnecke, R. B., and Chavez, N. (2006b). Cultural variability in the effects of questions designs features on respondent comprehension of health surveys. *Annals of Epidemiology* 16: 661-668.

Johnson, T. P., O'Rourke, D. P., Chavez, N., Sudman, S., Warnecke, R. B., and Lacey, L., (1997). Social cognition and responses to survey questions among culturally diverse populations. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz and D. Trewin (Eds.), *Survey measurement and process quality*: 87-113. New York: Wiley.

Krosnick, J.A., Nuruyan, S., and Smith, W.R. (1996). Satisficing in surveys: initial evidence. *New Directions for Evaluation* 70: 29-44.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology* 5: 213-236

Krosnick, J.A. and Alwin, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly* 51, 2: 201-219.

Laganà, F., Elcheroth, G., Penic, S., Kleiner, B., and Fasel, N. (2011). National minorities and their representation in social surveys: which practices make a difference? *Quality & Quantity*: 1-28.

Leeuw E. D. de, Hox J, and Huisman M. (2003). Prevention and treatment of item nonresponse. *Journal of Official Statistics*. 19: 153–76.

Lenzner, T., Kaczmirek, L., and Lenzner, A. (2009). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology* 24: 1003-1020.

Loosveldt, G., Pickery, J., and Billiet, J. (2002). Item nonresponse as a predictor of unit nonresponse in a panel survey. *Journal of Official Statistics*. 18: 545–557.

Mood, C. (2010). Logistic regression: why we cannot do what we think we can do, and what we can do about it. *European Sociological Review*. 26(1): 67-82.

Narayan, S. and Krosnick, J.A. (1996). Education moderates some response effects in attitude measurement. *Public Opinion Quarterly* 60: 58-88.

Papke, L. E. and Wooldridge, J. (1996) Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics* 11: 619–632.

Peytcheva, E. (2011). Language of administration as a source of measurement error: Implications for surveys of immigrants and cross-cultural survey research. ProQuest, UMI Dissertation Publishing.

Saris, W., Revilla, M., Krosnick, J.A., and Shaeffer, E.M. (2010). Comparing questions with agree/disagree response options to questions with construct-specific response options. *Survey Research Methods* 4: 61-79.

Simon, H. (1947). *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. – 4th ed. in 1997, The Free Press.

Snijders, T. and Bosker, R. (1999). *Multilevel Analysis*. Newbury Park, California: Sage.

Tourangeau, R. and Rasinski, K. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin* 103 (May): 299-314.

Tourangeau, R., Rips, L.J. and Rasinski, K. (2000). *The Psychology of Survey Response*. New York: Cambridge University Press.

Warnecke, R. B., Johnson, T. P., Chavez, N., Sudman, S., O'Rourke, D. P., Lacey, L., and Horm, J. (1997), Improving question wording in surveys of culturally diverse populations." *Annual Epidemiological Psychology* 7: 334–342.

Yan, T., and Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology* 22, 1: 51-68.