

CONTEXT EFFECTS IN SOCIAL SURVEYS: BETWEEN INSTRUMENT AND RESPONDENT

Abstract. The intention of this article is to examine the nature of cognitive representations that respondents create and utilise when making evaluative judgements in surveys. The following two hypothetical statements are the starting point of our analysis: (1) the order of items affects both the cognitive representation of the underlying dimension and the factor structure of the answers; (2) the different levels of cognitive sophistication of respondents affect their recognition of the concepts measured and the nature of their answers. A multidimensional scale of the concept of 'negative nationalism' was analysed. The concept was measured on two dimensions: xenophobia and protectionism. The research results partly confirm the hypothetical statements: the level of a respondent's education (as an indicator of their level of cognitive sophistication) influences their recognition of the concept measured, while other classical context characteristics (such as item order) did not. We can only confirm the possible effect of item order on the nature of their response, which results in the two-dimensionality or unidimensionality of the concept measured.

Keywords: survey context, measurement, scales, cognitive structures, item-order effects

Introduction

Social survey researchers tend to reduce or control as far as possible (systematic) context effects on the replies of their respondents. To this end, researchers must determine whether respondents – as a result of the particular context effect- are able to identify a measured concept 'hidden' within the questions asked. In this article, we will discuss two aspects of context effects: a) so-called 'local context effects' (the effects of the measurement instrument); and b) so-called 'global context effects', which include the

* Dr. Samo Uhan, Assistant Professor, dr. Mitja Hafner Fink, Assistant Professor, Faculty of Social Sciences, University of Ljubljana.

motivational and cognitive basis of attitudes (see Uhan, 1998). We observe the local context as a question-order effect and the global context as the effect of respondents' cognitive sophistication (level of education) (Krosnick, 1992; Johnson et al., 1999).

When measuring a concept (such as nationalism, in-/tolerance, social distance, prejudice) within a social survey, researchers often use statements with which respondents express their agreement or disagreement. These usually take one of two forms: a) a balanced list of positive and negative items that cover the two poles of the same dimension (e.g. tolerant – intolerant) (bipolarity); and b) a list of items covering two or more (sub)dimensions of the same concept, which are not necessarily antagonistic (multidimensionality) (cf. Hafner-Fink and Uhan, 2013). In this article, we focus on the second form.

Context effects are not usually identified directly in respondents' replies, but are rather based on hypothesis. Context effects can be defined as those changes in responses to survey questions that are a consequence of the characteristics of the questionnaire or of the circumstances in which the survey takes place. If these effects were not present, the replies would be different, namely unaffected by context. The key to understanding context effects is the mental representation, or the model of information processing that the respondent references when forming his/her response. The replies of respondents can thus be influenced by the formal characteristics of the instrument used; for instance, the question order, the type of scale, etc. (see Schuman & Presser, 1981; Tourangeau & Rasinski, 1988; Smith, 1988; Sudman, Bradburn & Schwarz, 1996). On the other hand, some authors have considered other possible effects based on the specific circumstances in which the survey interview takes place – such as the personal, cultural and social context of those involved – which have an explicit or implicit influence (see Zaller & Feldman, 1992; Tourangeau, Rips & Rasinski, 2000; Hair, 2005). In this respect, the most frequent question researchers ask is whether the context effects should be treated as a 'temporary disturbance', or whether they represent a serious, systemic fault that may diminish the significance of the survey results or findings.

Although the methodological literature cites many instances of research into context effects, most are experimental studies constructed in order to demonstrate context effects. This does not diminish their validity, but it does raise the question, how often and in what way do context effects appear in non-experimental circumstances?

Researchers often point to the well-known study by Schuman and Presser (1981), who describe the influence of context in researching standpoints in 'normal' non-experimental circumstances. On the basis of their analysis of results in the DAS (Detroit Area Study), they establish that the likelihood of the occurrence of context effects is scarcely any greater than

coincidence. Smith (1988) reaches a similar conclusion following his analysis of the replies in the GSS (General Social Survey). Smith establishes that a random rotation of questions leads to context effects in only four percent of cases. However, these findings which indicate a low or coincidental effect of context on responses can be misleading.

Tourangeau, Rips and Rasinski (2000) demonstrate that context effects are more frequent than one might expect. They contend that researchers often overlook the fact that the appearance of context effects depends on the conceptual links among questions. This means that heterogeneous research often conceals context effects.

In this paper we will verify the appearance of context effects in the 'natural' environment of a public opinion survey. In our research we will check for the traditional effect of survey context on survey responses, i.e. the influence of the order in which questions are asked. At the same time, we will also test the influence of context in the case of related questions or statements, based on the hypothesis that a respondent's cognitive sophistication influences their ability to identify what is being measured (socio-cultural context). And in this situation we expect socio-cultural context (the ability of respondents to identify the concept measured) to override the expected effects of the survey context.

The Problem

In order to operationalise concepts, the standard design of social surveys includes two basic types of questions that can be referred to as objective and subjective. As a rule, respondents have fewer problems framing responses to objective questions than to subjective questions, since rather than the recognition and linking of facts, the latter demands the creative processing of information. This awareness has led to a debate as to whether it is theoretically appropriate to view the underlying variable as a unidimensional continuum.

Bipolar survey items have long been held to be appropriate instruments for the operationalisation of theoretical models. This is because it is relatively simple for them to prompt the respondent to a cognitive representation of the object of research that is close to the relevant concept expected by the researchers. The application of a bipolar scale, however, raises a number of methodological questions. In the face of 'faulty' results, cognitive psychologists have drawn attention to the fact that the actual instrument, namely the bipolar scale of statements, can itself be the object of the cognitive representations that the respondents form of the research and the context in which it is carried out. If cognitive representations play a crucial role in models of information processing, this would also imply the hypothesis that these

representations have a direct influence on the formation of responses to survey questions and, consequently, that they reduce the validity of the results. The experimental findings relating to this methodological problem (presented below) emphasise the importance of taking account of both the socio-cultural context (cultural and social) and the context in which the survey is conducted (the characteristics of the instrument).

Traditionally, social surveys address respondents' attitudes as a latent variable that can be adequately measured on a bipolar response scale. These models assume that the cognitive structures underlying the responses to a bipolar survey scale are both unidimensional and continuous. The term 'unidimensional' refers to the notion that there is a single dimension of variability for the class of stimuli being judged. The term 'continuous' refers to the fact that there are no breaks in the dimension: that is, that there is a seamless gradient from one end of the latent variable to the other. A cognitive structure that has the properties of unidimensionality and continuity should produce judgments that are reciprocally antagonistic: that is, as one moves away from one pole of the response continuum, one will inevitably move towards the opposite pole (Ostrom et al., 1992: 298). If, for example, one were measuring the 'subjective' dimension of '(in)tolerance' then one end of the bipolar scale would consist of a statement that expressed extreme intolerance, while at the other end would be an highly tolerant statement.

Early on, researchers noted that the concept of a unidimensional latent variable was unsuitable for evaluating complex social phenomena. For instance, researchers encountered difficulties when attempting to assess subjective standpoints as to whether the death penalty is morally justifiable. The measuring of such a standpoint proved possible only when the respondents combined evaluations for different aspects of the death penalty (ethical, legal, cultural, historical, etc.). According to Ostrom (Ostrom et al., 1992), respondents face a similar task every time they evaluate the properties of objects that differ from each other but are not mutually exclusive (when measuring a complex, subjective standpoint). As an example of an ineffective scale, Ostrom cites the traditional American presidential question that assesses the conservatism or liberalism of the candidate. The scale includes two poles on what appears to be the same continuum – liberal and conservative- which the survey participants are expected to perceive as mutually exclusive, although much research shows the opposite.

A newer approach to social psychology (above all cognitive theory) reconceptualises the traditional psycho-physiological models for presenting opinion, or, as Ostrom (Ostrom et al., 1992: 298) notes, 'The attitude construct, traditionally viewed as a bipolar continuum, can be reconceptualised in terms of two discrete categories that are separately stored in a semantic network.' The essence of this approach is the desire to identify the discrete

cognitive representations triggered by the survey questions. The respondent has two kinds of cognitive representations available: those that relate to the object of evaluation and those that relate to the instrument of measurement (for instance, the above-mentioned scale of conservatism or liberalism of the presidential candidate). Prior to giving a response, the survey participant's task is to find the highest level of correspondence between the two cognitive representations, the *object* and the *instrument* of measurement. Continuing with our the example of evaluating the presidential candidate on the conservative-liberal scale, it is reasonable to assume that the extreme poles of this scale trigger separate cognitive structures or representations (a prototypical conservative or liberal president). The task of the respondents is to decide which of the prototypical representations (or both together) best describe the candidate.

In spite of the apparently unidimensional bipolar questioning, the likelihood that, in giving their answers, the respondents will make use of separate cognitive categories that can 'support' both poles of the scale (simultaneously conservative and liberal) has prompted theorists to surmise that the latent cognitive structures are dualistic and discrete (Ostrom et al., 1992: 298). In discrete categories, the two poles are linked by content but are nevertheless independent of each other; each pole is not necessarily a negation or inverse property of the category at the other pole. Dualistic categories are linked but not necessarily antagonistic. In the conservative-liberal example, the respondent can express his or her attitude as non-conservative, but not at the same time liberal. In other words, we are no longer talking about one dimension on a bipolar conservative-liberal continuum, but about two separate dimensions - liberal and conservative. Based on the assumption of duality, Ostrom designed an experiment to reject the hypothesis that the latent variables were merely unidimensional and continuous (Ostrom et al., 1992). The starting point for Ostrom's experiment was the 'Donald case', which is known in socio-psychological research and which describes an average day in the life of a fictitious character known as Donald. Donald has both the positive and negative features of an average person. In the original experiment, the researchers tested a number of hypotheses with a pre-prepared evaluation test on the unidimensional continuum of 'friendly-unfriendly (malicious)'. To this end, Ostrom made use of a special procedure that involved a short presentation of Donald, which incorporated a 'disturbing story'. The purpose of this story was to negate the effect of recentness.

In carrying out his experiment, Ostrom used three different sequences of questions: each sequence incorporated the same twelve statements which served to prompt the respondents to express their impressions of the target person. Six statements fitted the latent dimension 'friendly-unfriendly (malicious)', while the other six were construct-irrelevant.

Ostrom used an eleven-step scale of agreement presented on the two pages of the questionnaire, which each carried six statements. The sequences in Ostrom's experiment differed in terms of the location of the relevant and irrelevant statements. In the random ordering sequence, the twelve statements were ordered at random with three relevant and three irrelevant statements on each page. In the 3-6-3 sequence, the first three statements on the first page were relevant to one of the poles of the scale, while the last three statements were relevant to the opposite pole; the middle six statements in this sequence were irrelevant. The 6-6 sequence placed all the relevant statements on one page. According to Ostrom's findings, duality (two-dimensionality) can be expected when the measurement instrument enables the direct identification of the concept due to the ordering of the questions in both block sequences. In the case of the random sequence of items, bi-polarity (unidimensionality) would be expected. The findings of other researchers suggest that when the instrument enables the direct identification of the concept, the opposite effect (a one-dimensional structure) can be expected (Hafner-Fink & Uhan, 2013). It seems that the latter indicates the 'global context' effect (the content of the concept and the cognitive sophistication of respondents), which may override the question order effect (ibid. 850-851).

The Research Model

These findings formed the starting point of our research. Apart from question order effect ('local context'), we were interested in the ability of respondents to identify what was being measured ('global context'). The test was performed in the autumn of 2003, within the framework of the Slovenian Public Opinion Survey (SJM 2003/3 and SJM 2003/4; see Toš et al., 2004a, 2004b). We obtained 1777 questionnaires which answered all the relevant questions (2002 adult residents of Slovenia were surveyed).

The respondents were divided into three sub-groups, each of which responded to a different version of the questionnaire.¹ With the help of a five-step (Likert-like) scale, the participants expressed agreement or disagreement with a group of twelve statements, six of which were designed to express two dimensions of negative nationalism (protectionism and xenophobia), while the remaining six were construct-irrelevant. In each version of the questionnaire, the position of the construct-relevant items were changed in line with Ostrom's sequence, so that, in the first version, the first three statements expressed xenophobia and the last three protectionism, while the six intermediate statements were construct-irrelevant (i.e.

¹ *The respondents completed a survey sheet which was attached to the basic SJM questionnaire.*

using a 3-6-3 sequence). In the second version, all six construct-relevant statements appeared in one block that comprised the concept 'negative nationalism', while the six irrelevant statements appeared separately on the other page of the questionnaire (i.e. a 6-6 sequence). In the third version, the twelve statements appeared in random order, with three relevant and three irrelevant statements on each side of the questionnaire, irrespective of the concept.

Analysis and Results

We supplemented Ostrom's thesis on the influence of the measuring instrument (bipolarity and item order) with the thesis on the influence of the 'content' of the concept measured. We thus carried out a test in which all the statements were unidirectional (which means that we excluded the possible influence of bipolarity) and we researched only the influence of the item order on the 'formation' of the two-dimensionality (or unidimensionality) of the concept of *'negative nationalism'*. A Likert scale was used to measure the level of *negative nationalism*, which was hypothetically divided into two dimensions: xenophobia and protectionism. Here, each dimension was represented by three statements that were combined in the questionnaire with six irrelevant statements². The scale was constructed as follows:

- a. the dimension of *'protectionism'* was represented by the following three statements:
 - Slovenia should limit the import of foreign products to protect its economy. (x1)
 - Foreigners should not be allowed to buy land in Slovenia. (x2)
 - Slovenian television stations should give precedence to Slovenian films and programmes. (x3)
- b. the dimension of *'xenophobia'* was represented by the following three statements:
 - The crime rate is increasing because of the number of immigrants. (x4)

² The six irrelevant statements covered concepts (dimensions) of authoritarianism and traditionalism that are theoretically (conceptually) related to the relevant dimension (concept) of nationalism. Because of this affinity of the relevant and irrelevant concepts, the 'task' of 'identification' of the relevant dimension was not trivial for respondents. The following six irrelevant items were included in the questionnaire:

- Homosexuals should not be allowed to express their sexual orientation in public. (x7)
- Old customs are being destroyed by modern times. (x8)
- I am always prepared to support new things. (x9)
- Human life is determined by destiny. (x10)
- In general, a child will benefit if he/she accepts his/her parents' way of thinking. (x11)
- A community that tolerates large differences in beliefs cannot survive in the long run. (x12)

- Non-Slovenes should not be allowed to hold public posts. (x5)
- Mixing people of different ethnic and cultural backgrounds brings only problems. (x6)

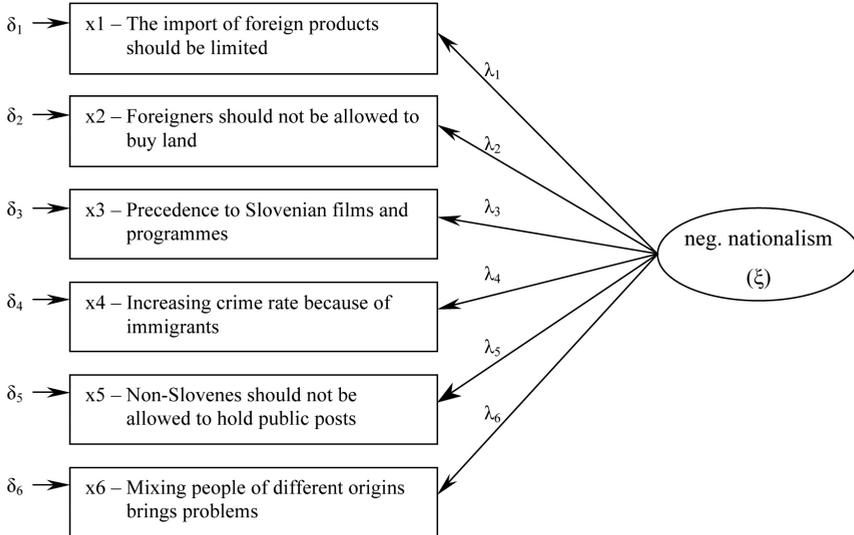
To test both models (unidimensionality and two-dimensionality) for each of the three conditions of ordering (3-6-3; 6-6; and random), we used confirmatory factor analysis (CFA) on the basis of a correlation matrix of the six relevant statements shown. We first carried out an exploratory factor analysis (the principal axis method) which in this case yielded similar results to those arrived at by Ostrom and his colleagues: in both block sequences there is a clear two-dimensional solution (protectionism, xenophobia); while in the case of the random ordering of statements, a unidimensional solution makes more sense.³ Using confirmatory factor analysis for all three conditions of ordering the relevant statements, we tested two measurement model variants: a) a two-factor model (a thesis of the two-dimensionality of the concept of *negative nationalism*) – *protectionism* (ξ_1) and *xenophobia* (ξ_2); and b) a single-factor model, which presupposes the unidimensionality of the concept of *negative nationalism* (ξ). Here too, we tested using both models, and for each model we tested the three conditions of ordering the relevant statements (*Diagram 1*). We compared the results of the testing of all six models to the same degrees of freedom (df). To evaluate the models' fit we applied two statistics: the ratio χ^2/df and the root mean square error of approximation (RMSEA)⁴.

³ In the instances when irrelevant items were also included in the model, the results were structurally the same: a) in both block sequences, the two expected factors of nationalism (*xenophobia* and *protectionism*) and the two factors of irrelevant items were formed; b) in the random sequence, two factors were formed – one factor of relevant items (*nationalism*) and one factor of irrelevant items. There are only a few deviations: a) one irrelevant item (about homosexuals) 'joined' the factor of *xenophobia* in both block sequences; and b) one irrelevant item (about homosexuals) 'exchanged' positions with one relevant item.

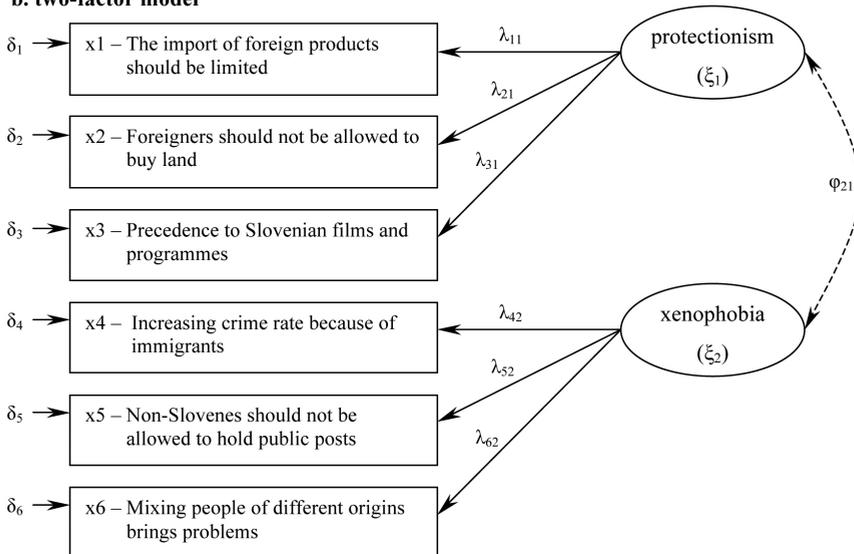
⁴ There are different views in the literature as to the threshold value for each statistic. For the χ^2/df ratio, 'different researchers have recommended using ratios as low as 2 or as high as 5 to indicate a reasonable fit' (Marsh and Hocevar, 1985: 567). There is also disagreement regarding the RMSEA value: some authors are more conservative and put the value of 0.05 as the upper boundary, while others are more liberal and put the value of 0.08 as the upper boundary for a good model fit (e.g. MacIntosh, 1998: 87; Li and Wehr, 2007: 376).

Diagram 1: GENERAL HYPOTHETICAL MODELS FOR THE SCALE OF 'NEGATIVE NATIONALISM'

a. one-factor model



b. two-factor model



The Question-Order Effect

Only when allowing for correlations of measurement errors in all three conditions of ordering were we able to obtain a good fit between the model and the data in the case of both the single-factor and the two-factor models. However, when comparing the results of testing for all six models to the same degrees of freedom, differences in the model's fit are present (*Tables 1a, 1b and 1c*). These differences and the specific results for each model do indicate item-order effects. In the case of the two block sequences, the fit between the single-factor model and the data is significantly worse than the fit between the two-factor model and the data. Thus, the single-factor model's fit with the data in the case of the random ordering of statements is stronger than in both block sequences, while the opposite applies for the two-factor model; random ordering means a worse fit between the model and the data. So, for the block sequences, a two-factor solution makes most sense and for the random distribution, a single-factor solution. Similarly, given the correlations among the factors (φ_{21}) in the case of the two-factor model, we can conclude that a single-factor solution makes more sense for the random ordering, as the correlation is considerably higher ($\varphi_{21} = 0.866$) than in either of the block sequences (0.624 and 0.573). Moreover, the correlations among the measurement errors (δx_{ij}) shows the logic of a two-factor solution in the two block sequences: to obtain a good fit for the single-factor model, we were obliged, in both block sequences, to allow for correlations among the measurement errors of the statements within the same measured dimension (specifically, for *protectionism*); however, there was no correlation among the measurement errors for the statements that we included in the different dimensions (*Tables 1a in 1b*). In the case of random ordering, to obtain a good fit between the model and the data we had to allow for correlations among the measurement errors; although, in this instance, we were not dealing with systematic correlations within particular dimensions, but rather correlations of errors between 'protectionist' and 'xenophobic' statements.

The results thus show possible fits between the data and both the single-factor and two-factor models, but only when allowing for correlations between measurement errors for particular statements. On this basis, we can conclude that, behind the measured concepts relating to negative nationalism (common to both), there is another, more general dimension for which we had to allow certain correlations of errors which we initially assumed to be illogical. Namely, we must consider the possibility that irrelevant statements could be a by-product of associated attitudes, such as authoritarianism and traditionalism. For instance, a nationalistic attitude in the form of protectionism or xenophobia can also be imbued with authoritarianism.

Table 1a: THE RESULTS OF CONFIRMATORY FACTOR ANALYSIS (CFA) – TEST OF THE SCALE OF 'NATIONALISM', SEQUENCE 3-6-3 (SJM 2003/3-4)*

statements:	one-factor model		two-factor model		
	lambda - x	R ²	lambda - x		R ²
	neg. nationalism (ξ)		protect. (ξ ₁)	xenoph. (ξ ₂)	
x1 - no foreign products	0.417	0.177	0.661	--	0.337
x2 - no land for foreigners	0.490	0.240	0.653	--	0.426
x3 - precedence to Slovenian films...	0.401	0.161	0.578	--	0.334
x4 - more crime because of immigrants	0.679	0.461	--	0.755	0.570
x5 - no public posts for Non-Slovenes	0.701	0.492	--	0.655	0.429
x6 - mixing of people - problems	0.799	0.639	--	0.881	0.776
	δx ₂₁ = 0.190; δx ₃₁ = 0.205		φ ₂₁ = 0.624		
	df = 7; χ ² = 35.660 (P = 0.000)		δx ₆₄ = -0.135		
	RMSEA = 0.083; χ ² /df = 5.094		df = 7; χ ² = 9.302 (P = 0.232)		
			RMSEA = 0.023; χ ² /df = 1.329		

* Results are shown in a standardised version.

Table 1b: THE RESULTS OF CONFIRMATORY FACTOR ANALYSIS (CFA) – TEST OF THE SCALE OF 'NATIONALISM', SEQUENCE 6-6 (SJM 2003/3-4)*

statements:	one-factor model		two-factor model		
	lambda - x	R ²	lambda - x		R ²
	neg. nationalism (ξ)		protect. (ξ ₁)	xenoph. (ξ ₂)	
x1 - no foreign products	0.362	0.134	0.603	--	0.363
x2 - no land for foreigners	0.465	0.217	0.754	--	0.568
x3 - precedence to Slovenian films...	0.403	0.162	0.669	--	0.448
x4 - more crime because of immigrants	0.684	0.467	--	0.685	0.469
x5 - no public posts for Non-Slovenes	0.710	0.503	--	0.717	0.514
x6 - mixing of people - problems	0.762	0.581	--	0.767	0.588
	δx ₂₁ = 0.241; δx ₃₁ = 0.218		φ ₂₁ = 0.573		
	df = 7; χ ² = 32.254 (P = 0.000)		δx ₃₂ = -0.166		
	RMSEA = 0.079; χ ² /df = 4.608		df = 7; χ ² = 9.875 (P = 0.196)		
			RMSEA = 0.027; χ ² /df = 1.411		

* Results are shown in a standardised version.

Table 1c: THE RESULTS OF CONFIRMATORY FACTOR ANALYSIS (CFA) – TEST OF THE SCALE OF ‘NATIONALISM’, RANDOM SEQUENCE (SJM 2003/3-4)*

statements:	one-factor model		two-factor model		
	lambda - x	R ²	lambda - x		R ²
	neg. nationalism (ξ ₁)		protect. (ξ ₁)	xenoph. (ξ ₂)	
x1 – no foreign products	0.594	0.352	0.554	--	0.307
x2 – no land for foreigners	0.623	0.388	0.684	--	0.468
x3 – precedence to Slovenian films...	0.491	0.241	0.528	--	0.278
x4 – more crime because of immigrants	0.621	0.385	--	0.658	0.433
x5 – no public posts for Non-Slovenes	0.691	0.477	--	0.702	0.493
x6 – mixing of people – problems	0.762	0.581	--	0.745	0.555
			ψ ₂₁ = 0.866		
	δx ₅₁ = -0.221; δx ₆₁ = -0.132		δx ₅₁ = -0.137		
	df = 7; χ ² = 21.090 (P = 0.004)		df = 7; χ ² = 19.950 (P = 0.006)		
	RMSEA = 0.058; χ ² /df = 3.013		RMSEA = 0.056; χ ² /df = 2.850		

* Results are shown in a standardised version.

Empirical Grounds for the Socio-Cultural Context Effect Hypothesis

In general, the results support the hypothesis of item order-effect (survey context), but some findings also give rise to certain doubts: a) to obtain a good fit for a two-factor model in both block sequences, a correlation between the factors was included in the model; b) to obtain good fits for all models, some correlations between measurement errors were necessary. Owing to these facts, we argue that the item-order effect (survey context) was corrupted with certain other effects. We have already mentioned the possible effect of irrelevant statements. But we are more interested in the possible effect of the varying cognitive sophistication of the respondents (socio-cultural context), which could result in varying abilities to identify the measured concept and (all) relevant items for the concept. This would mean that not only do the technical characteristics of the instrument (item-order) affect respondents' answers, but also the content of relevant (and irrelevant) items. Our experiment questionnaire also included questions which would allow us to test the abilities of respondents to identify relevant items from among all 12 relevant and irrelevant items. Respondents were asked to select 6 items which were relevant for the measured concept. The mean value of the number of correctly identified items is 4.20 and there are only minor (statistically non-significant) differences between the three methods of ordering: a) 4.13 for the block sequence 6-6; b) 4.20 for the block sequence 3-6-3, and; c) 4.24 for the random sequence.

We then tested the correlation between the respondents' level of education and their ability to identify the relevant items. For the purposes of our investigation, we take the level of education to be an indicator of general competence or cognitive sophistication as one of the dimensions of the socio-cultural context. The variations in the different education levels are statistically significant ($p < 0.01$) – a higher level of education indicates a higher level of identification: the mean value of correctly identified items within the group of respondents with elementary school (or less) is 4.03 (SEM = 0.079). This rises to 4.50 (SEM = 0.089) in the group of respondents with college or university education. To compare both effects (respondents' education and item-order), we carried out a linear regression analysis with the *number of correctly identified relevant items* as a dependent variable and a) *education level* (4-point scale) and b) *item-order binary variable* (non/random) as predictors. Only the *effect of respondents' education has proved to be statistically significant*, albeit weakly so⁵ (Beta = 0.10; $p < 0.01$) (see Table 2).

Table 2: THE EFFECTS OF EDUCATION AND QUESTION ORDER ON ITEM RECOGNITION – RESULTS OF MULTIPLE LINEAR REGRESSION ANALYSIS (SJM 2003/3-4)

	B	Beta	p
intercept	3.808		0.000
- education	0.150	0.100	0.000
- question order (non/random)	0.078	0.024	0.310
N = 1730, Adjusted R ² = 0.010, F = 9.354 ($p < 0.01$)			

Discussion

The general premise of our research was to test the presence of the context effect of survey questions on the responses obtained, as a consequence of the manipulation of the order in which they occur. The results we have obtained allow us to draw the following conclusions about the influence of context. We can confirm the work of authors who emphasise the importance of taking context into account when interpreting research findings (cf. Tourangeau et al., 2003), as well as those who assert that these effects are not as common as might be concluded on the basis of experimental findings (cf. Smith, 1988). Our findings suggest that, in specific situations,

⁵ When the level of education was increased by one degree, the level of correctly identified items increased by 0.15 (at the intercept value of 3.80).

the question-order (survey context) effect can be overridden by the socio-cultural context. The survey participants responded to the content of the questionnaire more than to its technical specificities and in doing so made use of their representations of the concept being measured. This can happen only when respondents identify the concept.

The results confirm both the hypotheses that: a) the technical properties of the instrument (local context) are likely to exert an influence on the participants' responses; and b) that cognitive sophistication (socio-cultural or global context) may influence the results. Regarding the local context: the two-dimensionality of the concept of 'negative nationalism' was formed when the structure of the questionnaire clearly showed it in the two block sequences. When the statements appeared in random order, the two-dimensionality was not 'identified'. We can confirm the general findings regarding the influence of the local context (already mentioned in our introduction), especially the following:

1. Respondents form representations of the measuring instrument, which has an influence on their response process or choice of modalities;
2. Above all, the findings confirm the influence of survey context or item order on the formation of dimensionality of the concept being measured.

Nevertheless, there are important additional effects which could not be simply ascribed to the context as it is usually understood in survey research. The understanding of the survey context should be broadened by applying the new dimension which we referred to as the socio-cultural or global context. In our research, this dimension is firstly present as an effect of the content of research (or the concept measured) which stimulates some responses according to the level of *cognitive sophistication* of respondents. This is also related to our key finding, that the *level of education of the respondents influences their recognition of the (content of the) concept being measured, while other classical context characteristics (such as item order) does not influence the recognition of the concept*. Based on the results of the analysis, we should stress that the effect of education is also rather weak.

This result also has direct research implications. We can see that respondents with a higher level of education are more likely to recognise the concept being measured, which means that they are also more capable of identifying the semantic variability of the different wording of the question. The opposite is true for less educated respondents, who express significantly lower semantic sensibility (cf. Uhan, 1998). The problem is that cognitive sophistication (in our case indicated by the level of education) affects the identification and interpretation of the researcher's intention (cf. Hippler, 1989). Semantic variations in conceptually and logically equivalent

questions trigger different perceptions and interpretations of these questions by different categories of respondents (see Uhan, 1998: 99). This consequently creates difficulties in drawing comparisons and equivalence in the measurement of results.

REFERENCES

- Hafner-Fink, Mitja and Samo Uhan (2013): The Bipolarity and/or Duality of Social Survey Measurement Scales and the Question-order Effect. *Quality & Quantity* 47 (2): 839-852.
- Hair, Mario (2005): Using an Individual Differences Perspective to Investigate Context Effects in Surveys: a Test of the Belief-Sampling Model. *International Journal of Public Opinion Research* 17: 213-225.
- Hippler, J. Hans (1989): Response Order Effects in Dichotomous Questions. The Impact of Administration Mode. Paper prepared for the meeting of American Association for Public Opinion Research.
- Johnson, P. Timothy, Diane O'Rourke, Elizabeth Severns (2009): Effects of Question Context and Response Order on Attitude Questions. In: 1998 Proceedings of the Section on Survey Research Methods pp: 857-860. American Statistical Association, Alexandria, VA. Dostopno preko <http://www.amstat.org/sections/srms/proceedings/>. 14. 11. 2009.
- Krosnick, A. Jon (1992): The Impact of Cognitive Sophistication and Attitude Importance on Response Order and Question-Order Effects. In: Schwartz, N., Sudman, S. (eds.) *Context Effects in Social and Psychological Research* pp: 203-218. New York: Springer - Verlag.
- Li, Tao, Kevin Wehr (2007): The Validity of the Factor Structure of the General Social Survey Environmentalism Scales across Gender and Ethnicity in the United States. *Organ. Environ.* doi:10.1177/1086026607306461
- MacIntosh, Randall (1998): A Confirmatory Factor Analysis of the Affect Balance Scale in 38 Nations: a research note. *Soc. Psychol. Q.* 61: 83-91.
- Marsh, W. Herbert, Dennis Hocevar (1985): Application of Confirmatory Factor Analysis to the Study of Self-Concept: First- and higher order factor models and their invariance across groups. *Psychol. Bull.* 97: 562-582.
- Ostrom, M. Thomas, M. Andrew Betz & J. John Skowronski (1992): Cognitive Representation of Bipolar Survey Items. In N. Schwarz & S. Sudman (Eds.), *Context Effects in Social and Psychological Research* pp: 297-315. New York: Springer-Verlag.
- Schuman, Herbert & Stanley Presser (1981): *Questions and Answers in Attitude Surveys: Experiments in Question Form, Wording, and Context*. New York: Academic Press.
- Smith, Tom W. (1988): *Ballot Position: An Analysis of Context Effects Related to Rotation Design* (GSS Methodological Report No. 55). Chicago: NORC.
- Sudman, Seymour, Norman M. Bradbourn & Norbert Schwartz (1996): *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey-Bass Publishers.

- Toš, Niko and group (2004a): Slovenian Public Opinion Survey 2003/3: National and International Security; Military Profession [codebook file]. Ljubljana: University of Ljubljana, Faculty of Social Sciences, Public Opinion and Mass Communication Research Centre [production]. Ljubljana: University of Ljubljana, Social Science Data Archive [distribution]. URL: <http://www.adp.fdv.uni-lj.si/opisi/sjm034.xml> (accessed: June 2005).
- Toš, Niko and group (2004b): Slovenian Public Opinion Survey 2003/4: Citizenship / International Survey ISSP [codebook file]. Ljubljana: University of Ljubljana, Faculty of Social Sciences, Public Opinion and Mass Communication Research Centre [production]. Ljubljana: University of Ljubljana, Social Science Data Archive [distribution]. URL: <http://www.adp.fdv.uni-lj.si/opisi/sjm034.xml>, (accessed: June 2005).
- Tourangeau, Roger & Kenneth Rasinski (1988): Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin* 103: 299-314.
- Tourangeau, Roger, Lance J. Rips & Kenneth Rasinski (2000): *The Psychology of Survey Response*. New York: Cambridge University Press.
- Tourangeau, Roger, Eleanor Singer & Stanley Presser (2003): Context Effects in Attitude Surveys. Effects on Remote Items and the Impact on Predictive Validity. *Sociological Methods & Research* 31: 486-513.
- Uhan, Samo (1998): *Prava in neprava mnenja*. Ljubljana: Znanstvena knjižnica FDV.
- Zaller, John & Stanley Feldman (1992): A Simple Theory of the Survey Response: Answering Questions Versus Revealing Preferences. *American Journal of Political Science* 36: 579-616.