Polarization of Tastes:
Stated Preference Stability in Sequential Discrete Choices

Tomasz Gajderowicz[1], Gabriela Grotkowska[2]

Abstarct:

Purpose: The aim of this study is to assess stated preference stability in long-format discrete choice experiments. As the number of choice situations increases, data reveal more precise information regarding preferences. However, there are many doubts concerning the incentives compatibility of long designs. Psychological effects such as respondents' learning, fatigue and decreasing concentration in successive choice situations can result in biased estimators of parameters of utility functions. It is, not clear which group of successive choices reveal the most trustworthy information: the initial choices (undistorted but potentially not robust) or a later set (consciously formed preferences but potentially under conditions of fatigue).

Design/Methodology/Approach: With the long-format (144 choice tasks) data concerning employment options, we estimated utility function parameters were estimated using MNL and MMNL models.To conduct inter- intra- respondent analysis we used imputed individual-level parameters of utility function.

Findings: We show that preferences are formulated at the intra-respondent level according to a specific pattern and, at the same time, the preferences of single respondents show lower variance across choice tasks than across populations. An increase in the standard deviation of parameters across the sample does not necessarily mean an inconsistency of preferences in this type of study. This can result from polarization of preferences in the population with simultaneous intra-respondent preference consistency.

Practical Implications: Long-format DCEs can reveal some of the behavioural mechanisms behind the decision-making process. We show that, using this kind of study, it is possible to observe preference formulation. In some specific cases obtaining accurate information, or even teaching respondents their preferences, can be of a substantial significance.

Originality/Value: An increase in the standard deviation of parameters across the population does not necessarily mean inconsistency resulting from the 'negative' consequences of ordering effects, contrary to the findings of Swait and Adamowicz (2001). Instead, this can result from the polarization of preferences in the population alongside the intra-respondent consistency.

Keywords: Stated preferences; DCE; mixed logit; intra-respondent heterogeneity.

JEL Codes: C9, D01, D91.

Paper type : Research article.

[1]Faculty of Economic Sciences, University of Warsaw e-mail: tgajderowicz@wne.uw.edu.pl
[2]Faculty of Economic Sciences, University of Warsaw e-mail: ggrotkowska@wne.uw.edu.pl

1. Introduction

Stated choice experiments have become a useful tool in the valuation of market goods across many fields of study. It is common practice in discrete choice experiments (DCE) to provide the respondent with several choice sets. Long-format surveys based on hypothetical choice situations offer a unique opportunity to extract precise information on respondents' preferences. The more choice sets one has, the more complete the information that can be extracted from the data. In some cases, when researchers have access to very limited numbers of respondents from the target population, it is worth considering how to extract as much information as possible from single respondents. In this context, questions regarding the stability and credibility of preferences in successive choice tasks become fundamental. Long-format DCEs are linked to several methodological questions. First of all, the behavioural nature of revealed preferences in choice sets must be considered. Second, issues regarding incentives compatibility must be discussed. Third, effects particularly connected to choice experiment methodology play an important role in the assessment of the credibility of study results.

Preferences are revealed by the choice-making of individuals, but it is not clear whether these preferences are a priori precisely formed and well known, learned through the experience of choice-making or internally coherent but dependent on exogenous anchors (Bateman *et al.,* 2008). The first interpretation is rather dominated by the others due to the reported inconsistency of preferences in successive decisions revealed by most DCE surveys (Hess and Train, 2011). The discovered preference hypothesis (DPH) offers a less restrictive view, allowing preferences to be discovered through the decision process (Plott, 1996). Preference formulation can occur via the respondent learning the choice task institution or values by repeating choices (Holmes and Boyle, 2005; Hanley *et al.,* 2005). When respondents face unfamiliar decisions, they will exhibit significant randomness.

However, as choices are repeated and respondents gain familiarity with their own preferences, the decisions progressively become more coherent and less random. This approach explains intra-respondent revealed preferences inconsistency in successive choice tasks. The third hypothesis goes even further, arguing that preferences are not discovered during subsequent decisions but are rather constructed (Slovic, 1995; Ariely *et al.,* 2006; 2003). For example, Ariely *et al.* (2003) argue that preferences depend on the design of the survey. They show that an initial value or exogenous shock leads to an anchoring effect through the sequence of choices. This theory predicts that respondents are unable to report their precise willingness-to-pay (WTP) for non-market goods and that their valuations are heavily dependent on several possible anchors. Their actual WTP varies as a function of the specific choice scenario. In recent years, a growing number of empirical works have proven that people are simply uncertain about their WTP and can merely report a WTP range rather than a point value (Ellingson *et al.,* 2009; Hanley *et al.,* 2009; Mahieu *et al.,* 2012).

Another approach, which can be seen as a development of the latter, assumes that preference formulation is based on the inverted relationship between choices and preferences. This hypothesis states that there are decisions that shape preferences (Brehm, 1956; Gerber and Jackson, 1993; Sharot, De Martino and Dolan, 2009), even in an unconscious way (Coppin *et al.,* 2010). Festinger (1957) proposed cognitive dissonance theory (CDT) to account for choice-induced preference formulation. He proved that making complicated choices regarding two goods between which an individual is initially indifferent have consequences for their valuation of such goods in future. Rejecting a favourite item induces a disutility called 'cognitive dissonance', which is unconsciously and automatically reduced by decreasing the valuation of the rejected item (Chen and Risen, 2010). This theory has been empirically proven by Sharot *et al.* (2010). They studied individuals who rated vacation offers both before and after making a blind choice (or one randomly performed by a computer) that could not be guided by pre-existing preferences. They found that the participants' preferences were altered by the blind choices, but not when the computer made the decision. This experiment suggests that there exists an inverted relationship between preferences and choices, which is in contradiction to the neoclassical microeconomic approach.

Moreover, choices alter preferences not just in short run, but even after a number of years (Sharot *et al.,* 2010). The issue, however, is yet more complicated when one analyses the biological foundation of the choice-making process. Recent neuroimaging studies suggest that the fact of a choice, or even the anticipation of a choice, recruits reward-related circuitry, such as the anterior and ventral striatum (Izuma *et al.,* 2010; Leotti and Delgado, 2011). From this medical perspective, choice may be intrinsically rewarding even when it is not preference-driven. Thus, the experience of making a choice will, in itself, influence preference for what is chosen (Tang, 2012).

After years of acceptance of the choice-induced preferences approach, it has been critically revisited by Chen and Risen (2010). They show that a free-choice paradigm (FCP) will produce spreading, even if individuals' preferences remain stable. If people's choices are an imperfect measure of their preferences and they are at least somewhat driven by preferences, then the FCP will measure spreading, even if people's preferences remain perfectly stable. Chen and Risen (2010) both proved a mathematical theorem that identifies a set of conditions under which the FCP will measure spreading and experimentally demonstrated that these conditions obtain and that the FCP measures a spread of alternatives, even when not caused by choice.

To summarize, biases are immanent characteristics of preferences mesurement. However, consumers' behaviour in many experiments often reveals consistent and well-defined preferences. The key to understanding the complexity of preference formulation is to link constructed preferences and well-defined values (Payne, Bettman and Schkade, 1999) and the demonstration of coherent arbitrariness (Ariely, Loewenstein and Prelec, 2007). This can be modelled as a kind of general

process whereby people construct preferences from a given starting point, as proposed by Barkan *et al.* (2016) who analysed how people extrapolate coherent preferences from relevant reminders. With the use of four empirical studies, they characterized the features of extrapolated preferences and compared them to preferences built from scratch. They demonstrated that the process of extrapolation leads to fewer errors, thus resulting in more consistent revealed preference estimates. Moreover, it reduces cognitive effort as the familiarity of the starting point increases the maintenance of transitivity (Barkan *et al.,* 2016).

2. Literature Review

The above-described issues suggest the conclusion that all choice experiments performed in order to reveal preferences in fact formulate them. On the other hand, if preferences are unstable across a survey it is crucial to assess which revealed preferences are the most trustworthy: the initial choices or those revealed towards the end of the study. In this context, long-format DCEs tend to reveal more realistic preferences and WTPs (revealing real complexity and assessments of stability), rather than a mere static representation of respondents' uncertainty that can be observed in single contingent valuation questions. In short, depending on the behavioural hypothesis one believes, one can expect that the instability of the estimated parameters of a utility function in a long-format DCE can be the result of either measurement errors (an a priori well-defined hypothesis) or the reduction in randomness of valuations with more decisions (suggesting the construction or learning of preferences). This phenomenon, over the set of choice scenarios, leads to systematic changes in both relative parameters (i.e., WTP) and absolute sensitivities (i.e., scale). In both cases, one would expect trends in estimated preferences parameters to be revealed in successive choice sets, rather than random fluctuations, and that these should be accommodated in a deterministic manner. On the basis of behavioural economics, it seems that constructing long-format choice experiments is a way to reveal real preferences or, more precisely, trends of uncertain preference behaviour.

The second fundamental problem connected with the stability of parameters across long-format DCE is incentives compatibility. This has been a topic of interest in a huge body of work since hypothetical choices became an increasingly popular method for the valuation of non-market goods. Carson and Groves (2007) formulated two conditions to maintain correct motivation: the survey should be associated with consequences in the real world and the space of choice should be binary (single choice between two alternatives). It should be noted, however, that these demands are so restrictive that the greatest advantage of the DCE is undermined – the possibility of obtaining a wealth of information about the preferences of one respondent using a questionnaire presenting sets of successive hypothetical situations.

The existing literature does not offer comprehensive investigation of the balance of these conditions' costs (loss of information on preferences by using just one single choice situation) and benefits (alleged theoretical incentives compatibility, though this has not yet been supported by empirical evidence). Leaving aside the problem of lacking motivation to reveal one's preferences, in successive choice situations respondents are exposed to a number of psychological effects connected with absorption of complex information. Preferences might be affected during the process of coding, combination, segregation, cancellation, simplification and detection of dominance (Kahneman and Tversky, 1979).

Those effects can perturb preferences, and more strongly the less known and more complicated the choice space is. According to the number of choice situations, decision-making becomes a more mechanized process and the time available for each decision decreases. As respondents select the best option of those presented, they pay attention mainly to the most important of each alternative's attributes. Such effects in DCE can thus lead to violation of some of the core assumptions of rational choice theory (Allingham, 2002).

Third, there are some psychological effects that are not necessarily to connected decision-making processes in general but rather to the DCE methodology in particular. These are generally known as ordering effects and refer to institutional learning, fatigue or boredom, and choice set order-dependence (Day et al., 2012). Institutional learning relates to the fact that most respondents participating in DCE surveys have never experienced this type of survey before. In experimental economics this has been described as confusion or failure of game design recognition (Andreoni, 1995; Chou *et al.,* 2009). The institutional learning hypothesis suggests that, in order to reduce uncertainty in DCEs, respondents should make repeated choices (Braga and Starmer, 2005).

On the other hand, repeating decisions often leads to fatigue or boredom. In this case, respondents' choices may exhibit increasing levels of randomness over the sequence of choice tasks (Swait and Adamowicz, 2001a; 2001b). It is important to notice that the fatigue effect strongly depends on the type of survey. Estimation results by Savage and Waldman (2008) suggest that, while online surveys provide benefits in terms of lower survey administration costs and reduced time between survey implementation and data analysis, these benefits may come at the cost of respondent fatigue and greater standard deviation in the estimation of utility.

However, Hess *et al.* (2012) provide strong evidence that concerns regarding fatigue are overstated in the literature, with no clear decreasing trend in scale across choice tasks. For the data sets tested, they find that accommodating any scale heterogeneity has little or no impact on substantive model results, that the role of constants generally decreases as the survey progresses, and that there is evidence of significant attribute-level (as opposed to scale) heterogeneity across choice tasks.

Choice set order-dependence refers to lexicographic preferences represented in some surveys. This happens when respondents tend to choose a particular alternative among others regardless of their preferences. On the other hand, this effect can also refer to the order of attributes when respondents tend to more favourably valuate attributes in a particular position of the design, regardless of their preferences (Day *et al.,* 2012). This problem, however, can be easily solved by employing randomization of attributes and alternatives in the design (Carlsson *et al.,* 2012).

Ordering effects, together with perceptions of high complexity in choice tasks or cumulative cognitive burden, result in changes in choice strategies, adopting non-compensatory decision rules. As a consequence, changes in the estimated marginal utilities of the attributes can occur (Czajkowski *et al.,* 2014). However, taking ordering effects on one hand and value learning on the other, it is again unclear which of the successive choices are most trustworthy (Carlsson *et al.,* 2012).

3. Empirical Strategy to Assess Preference Stability

As mentioned above, assessment of preference stability is a key methodological issue and has been a subject of interest in many methodological papers. A huge body of work shows that stability of preferences depends on the complexity of a choice task, incorporating the number of attributes, their levels, ranges and correlations, and the number of alternatives (Swait and Adamowicz, 2001; Caussade *et al.,* 2005; Day and Pinto Prades, 2010). It is, however, a very complicated endeavour to create an experimental design to reveal the psychological mechanisms behind indicated decisions. Most recent studies use the standard DCE approach and assess preference stability on the basis of a panel of 10-25 choice tasks. In order to account for preference stability, most researchers simply split panels into subsamples and assess differences in estimated parameters.

In order to assess preference stability and robustness in long-format DCEs in this paper, it was decided that a wide (four alternatives) and long (144 choice situations) DCE design would be employed. This empirical study focuses on a particular group of respondents: students and graduates (up to five years after graduation) of fields of social studies in Poland. For this group, it is relatively easy to ensure a high response rate, a relatively high level of homogeneity in terms of earlier labour market experience and (to a large extent) similarity in the types of jobs for which they can apply. The list of attributes identified as relevant include the following:

- Net salary (wage - overlapping);
- Type of contract (LCB, CivContr - overlapping);
- Overtime hours at work (time);
- Commuting time (ComTime);
- Fringe benefits (B_sport, B_med);
- Emotional attitude to work tasks (Task1, Task2);
- Type of position (Spec, Mngr);

- Career development possibilities (Dev1, Dev2);
- Remuneration system (Fix1, Fix2);
- Atmosphere in the workplace (Atm1, Atm2);
- Competition in the workplace (Comp1, Comp2).

The study included three continuous attributes: salary (PLN), overtime spent at work (hours) and commuting time (minutes). For each of the other attributes, three qualitative levels were defined. In the first step, the design (sets of job offers that are characterized by defined attribute levels) was created. This allowed for the estimation of utility function parameters. The selection of effective attributes sets required preliminary assumptions regarding these utility function parameters. Therefore, a pilot survey was carried out to estimate parameter values and to verify that respondents understood the attributes. The pilot survey was based on an orthogonal design (Street and Burgess, 2007), under which efficient experiment design is achieved when the sum of all parameters equals zero. Alternatively, the use of this approach is justified when an analyst does not have approximate information on the potential values of utility parameters.

The pilot survey was conducted on a sample of 67 representatives of the target group using printed copies of the choice sets. Each representative received 16 choice sets consisting of four job offers. Due to respondents' comments, we decided to reduce the number of attributes per choice task. For the sake of brevity, attributes were divided into three groups and presented separately. Two attributes (type of contract and wage) overlapped (they were in each and every choice situation).

The results of this pilot survey were as expected (all variables were significant and parameters returned expected signs). The parameters were then used as an assumption to construct a Bayesian efficient design. The Bayesian efficient design was generated by means of a numerical simulation in NGENE software. The generated Bayesian design was applied to the main survey, which was then carried out on the sample of 801 representatives of the target population of students and graduates of social sciences in Warsaw. Internet access was the only prerequisite to participation in the survey. Each participant was asked to take part in nine survey sessions and each session involved 16 choice situations. Each choice set covered four job offers, which participants were asked to rank from most to least attractive.

In addition to the four job offers, respondents were also provided with an option labelled 'none of the remaining offers' to indicating that they would not accept employment on the terms described by the job offers presented in the choice set. By ranking this option as the most attractive, a respondent could indicate that they would accept none of the presented offers. However, in such cases, respondents were still required to rank the remaining job offers in order of attractiveness below this first-place position. Adding the opt-out option potentially allows for the estimation of reservation wages with regard to particular job characteristics.

Finally, out of the 801 representatives who downloaded the online application and started the survey, 643 participants completed all nine sessions. In total, the database covered 513,760 observations, corresponding to 102,752 choice situations solved in 6,422 complete sessions. The process of collecting these data lasted from the 3rd of March to the 29th of May, 2014. The total time spent by all the respondents on making decisions and ranking job offers amounted to over 2,053 hours. Women accounted for 61% of the respondents. The ages of the respondents varied from 19 to 30 years old, and the average was just under 23 years of age. 25.2% of the respondents declared possessing work experience. The sample was relatively homogeneous and corresponded to the assumptions that were used to define the work attributes.

As for the estimation strategy we used both the multinomial logit (MNL) and the random parameter logit (MMNL) models (following Hess and Giergiczny, 2015). The MMNL model accommodates preference heterogeneity in a continuous specification, through integration of MNL choice probabilities over the assumed multivariate random distribution of the vector of preferences coefficients.This simple specification of the MMNL model is directly applicable to cross-sectional data. For the estimation of MMNL models on repeated choice data, the approach put forward by Revelt and Train (1998) has now become the state-of-the-art specification. This moves the integral from the level of individual choices to the level of the sequence of choices for individual respondents.

Due to the inclusion of 144 choice tasks per respondent, we decided to used mentioned above standard econometric procedures and statistical tests in subsamples in order to assess preference consistency over sequences choices.

4. Results

The first impression of how respondents behaved in this study is reflected in the time it takes respondents to make their decisions. The time between successive choices in the first session was, on average, more than twice as long as in the final, ninth session. For the first two cards in the first session a single decision took, on average, 15.49 seconds, while for the last two choices in the final session each decision took, on average, only 6.09 seconds. Analysis of the response time of the entire card selection provides similar results. The first two cards of each choice (choice sets) were each solved, on average, in 70.86 seconds (standard deviation: 52.05), while choice tasks ending the study (ninth session, choice sets 15 and 16) were solved, on average, in 27.58 seconds (standard deviation: 32.47).

In order to obtain parameters of the utility function, MNL and MMNL were calculated separately for the whole sample, for blocks (three sessions with the same set of attributes), for each session (16 choice tasks) and finally for every single choice situation across the 144 choice tasks. This last approach reveals the most interesting conclusions in addition to preference stability assessment.

To eliminate the scale parameter from the results, WTP was calculated for each level of attributes characterizing employment. Figure 1 presents WTP across 48 choice sets in every block of parameters. One can see that WTPs are relatively stable, but some trends are visible.

Figure 1. *Estimated WTPs from the MNL model across 48 choice sets (3*16 for 3 blocks)*
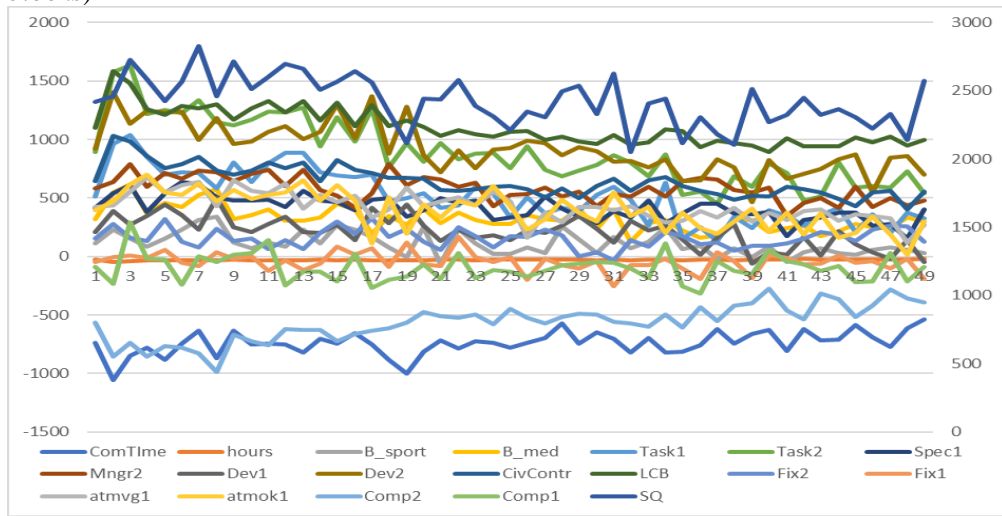
To assess preference heterogeneity across the whole panel of 144 choice sets, the overlapping attributes were taken into consideration. To avoid the influence of the scale parameter, we calculated WTP for the overlapping attributes (labour code-based employment (LCB), civil contract (CivContr) and the parameter of the opt-out alternative). In Figure 2 we can clearly see that WTP for the opt-out alternative is strongly block-sensitive. This is not a surprising result given that respondents can imagine different base levels of remaining job characteristics as different sets of attributes are presented. This is an interesting issue for further research and would seem to be crucial for economists calculating reservation valuations of whole programmes using opt-out alternatives. This is not, however, the issue addressed by this paper. For the remaining attributes, preferences (in the WTP space) demonstrate a consequent downward trend. This might support the DPH or FCP hypothesis.

In order to assess preference stability, formal statistical tests for differences in parameters were conducted. For each attribute we checked if WTP differed between sessions within the same block. It should be noted that sessions from each particular block never followed one another (Block 1: sessions 1, 4, 7; Block 2: sessions 2, 5, 8; Block 3: 3, 6, 9). The results indicate that parameters are statistically distinct between most sessions. Table 1 summarizes the results of the statistical tests indicating that WTPs are indifferent across sessions and across blocks. The shaded cells indicate that the difference between the WTPs is not significantly far from zero. In most cases, WTPs are significantly different. Both the DPH and FCP

hypothesis are in line with the conclusion of Table 1. This would, however, indicate that, while parameters in sessions and choice tasks that are separated from one another might be statistically different, successive sessions and choice tasks should reveal more consistent estimates. Table 2 presents statistical differences in estimated WTPs for overlapping attributes across 144 choice tasks.

Figure 2. *Estimated WTPs from the MNL model across 144 choice sets for overlapping attributes with 95% confidence intervals*
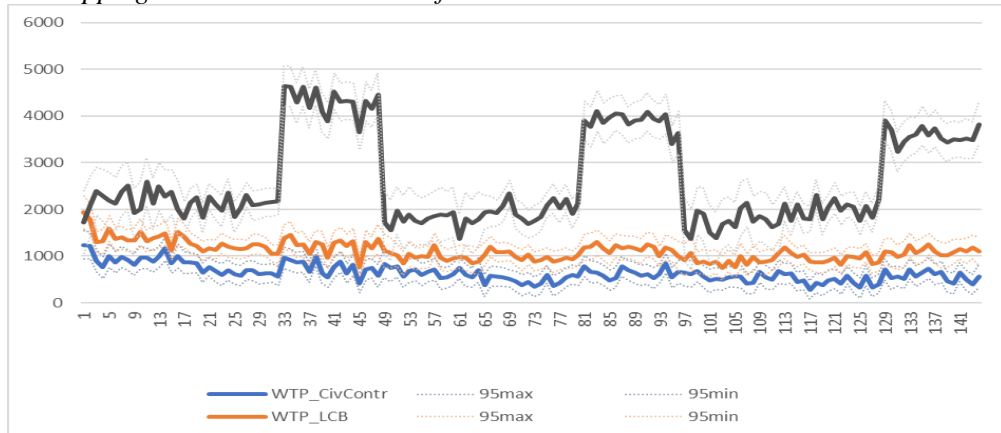

Table 1. *Tests for statistical difference in WTPs between sessions in each block of attributes*

	Block 1			Block 2			Block 3		
	1 vs. 4	4 vs. 7	1 vs. 7	2 vs. 5	5 vs. 8	2 vs. 8	3 vs. 6	6 vs. 9	3 vs. 7
WTP_1	-2.60	-2.29	4.81	-0.92	-2.73	3.66	3.68	5.18	-8.78
WTP_2	9.14	4.24	-13.26	1.66	1.00	-2.65	4.83	4.69	-9.53
WTP_3	13.23	5.96	-18.82	1.56	1.26	-2.81	2.79	1.49	-4.31
WTP_4	3.62	1.83	-5.41	5.36	0.77	-6.13	1.30	1.22	-2.54
WTP_5	3.91	1.67	-5.53	5.37	1.17	-6.51	2.39	0.00	-2.42
WTP_6	7.49	2.71	-10.11	-8.03	-4.74	12.68	2.03	1.33	-3.35
WTP_7	10.61	2.45	-13.02	0.90	-0.94	0.05	5.27	4.10	-9.34
WTP_8	4.69	0.62	-5.34	1.82	-0.09	-1.70	3.16	3.92	-7.07
WTP_9							5.48	4.43	-9.92

According to these results, estimated WTPs do not differ significantly in 81.3% of cases. It is worth noting that significant differences occur non-randomly between successive sessions (between the last choice in a particular session and the first in the following session). At the same time, the difference between WTPs calculated for choice tasks placed further from one another reveal increasing inconsistencies. Next, to assess intra-respondent heterogeneity, individual-level parameters were calculated using the method proposed by Revelt and Train (2000) and Train (2003).

Due to the number of variables and observations, individual-level parameters were calculated for every successive four choice task. As a result, for every respondent, 36 (144 divided by four) individual-level parameters were imputed. Some observations were excluded due to decision-time constraint in order to eliminate decisions performed by random clicking. We assumed that it is impossible to make a sensible decision (comparing five alternatives) in less than two seconds.

Table 2. *Tests for statistical difference in WTPs between successive 144 choice tasks for overlapping attributes*

		Choice task															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Session 1	WTP_CC	0.08	1.79	0.80	-1.37	0.75	-0.66	0.39	0.64	-0.88	-0.01	0.55	-0.65	-1.04	1.91	-0.88	0.80
	WTP_LCB	0.60	2.36	-0.12	-1.33	1.12	-0.16	0.36	-0.06	-0.98	1.08	-0.31	-0.20	-0.41	2.11	-2.21	0.55
	WTP_SQ	-0.72	-0.79	0.27	0.27	0.13	-0.64	-0.34	1.57	-0.25	-1.48	1.26	-0.95	0.51	-0.23	0.91	0.58
Session 2	WTP_CC	-0.01	0.08	1.38	-0.82	0.61	0.43	-0.62	0.56	0.21	-0.80	0.05	0.52	-0.05	-0.07	0.42	-2.60
	WTP_LCB	0.88	0.37	0.79	-0.52	0.25	-0.88	0.45	0.24	0.09	-0.12	-0.59	0.00	0.41	0.95	0.08	-1.89
	WTP_SQ	-1.34	-0.54	2.04	-2.11	0.73	0.61	-1.75	2.34	-0.86	-1.28	1.08	-0.12	-0.15	-0.05	-0.10	-9.51
Session 3	WTP_CC	0.30	0.38	-0.13	1.43	-2.21	2.35	0.56	-1.34	-0.72	1.58	-1.20	2.59	-1.90	-0.28	1.00	-1.48
	WTP_LCB	-0.35	1.08	-0.11	1.04	-1.30	0.21	1.66	-1.80	-0.30	0.64	-0.59	3.23	-3.02	0.68	-1.03	1.31
	WTP_SQ	0.02	1.12	-1.12	1.53	-1.60	1.85	0.69	-2.33	0.86	-0.13	0.04	2.38	-2.25	0.58	-1.08	7.70
Session 4	WTP_CC	0.58	-0.24	1.45	-0.78	-0.15	0.68	-0.39	-0.32	1.20	-0.10	-0.45	-0.92	0.96	0.41	-1.04	2.12
	WTP_LCB	0.32	0.27	1.25	-1.39	0.56	-0.26	0.20	-1.64	1.73	0.41	-0.37	-0.19	0.10	0.91	-0.34	-1.05
	WTP_SQ	0.44	-1.18	0.72	-0.45	0.37	0.11	-0.22	-0.17	-0.08	0.08	-0.21	1.61	-1.24	0.32	-0.26	-0.50
Session 5	WTP_CC	-1.31	0.13	0.06	0.20	0.43	0.49	-0.38	0.59	-0.38	-1.40	1.64	-0.40	-0.79	-0.48	0.24	-1.38
	WTP_LCB	-1.12	0.91	-0.07	-0.06	0.70	0.54	-0.83	1.04	-0.28	-0.45	0.75	-0.25	-0.37	0.31	-0.75	-1.00
	WTP_SQ	-0.06	0.15	-0.61	-1.26	1.98	0.38	0.42	-0.23	-0.36	-1.11	-0.66	1.03	-0.96	1.52	-1.10	-7.08
Session 6	WTP_CC	0.71	0.06	0.42	0.67	-0.37	-1.55	0.56	0.29	0.37	-0.14	0.52	-0.58	-1.55	1.98	-0.79	0.07
	WTP_LCB	-0.03	-0.51	0.67	0.56	-0.91	0.33	-0.12	0.17	0.27	-0.74	0.27	1.15	-1.12	0.26	0.77	0.52
	WTP_SQ	0.40	-1.09	0.98	-0.46	-0.38	0.13	0.77	-0.35	-0.07	-0.54	0.55	0.19	-0.62	2.18	-0.85	6.71
Session 7	WTP_CC	0.25	-0.42	0.76	0.64	-0.30	0.13	-0.28	-0.14	-0.05	1.10	-0.20	-1.54	0.90	0.30	-1.12	0.39
	WTP_LCB	-0.96	1.43	-0.22	0.39	-0.40	1.03	-1.16	1.03	-1.70	1.40	-1.19	0.85	-0.11	-0.33	-0.94	-0.82
	WTP_SQ	0.53	-1.76	0.23	1.20	0.34	-0.89	-0.20	0.39	-1.14	-0.31	1.15	-0.36	0.21	0.52	-0.21	-1.39
Session 8	WTP_CC	-0.13	1.26	-0.21	1.34	-1.05	0.33	-0.68	-0.26	0.77	-1.25	1.13	0.66	-1.71	1.66	-0.49	-2.22
	WTP_LCB	0.84	0.61	-0.39	1.04	0.17	0.02	-0.32	-0.51	1.22	-1.38	0.20	0.21	-0.88	1.67	-0.18	-1.59
	WTP_SQ	1.50	-1.49	1.25	0.03	-2.25	2.28	-1.03	-0.97	1.31	-0.59	0.16	1.23	-1.29	1.03	-1.61	-7.09
Session 9	WTP_CC	1.21	-0.23	0.34	-1.30	1.00	-0.58	-0.53	0.85	-0.40	1.44	0.27	-1.52	1.04	0.59	-1.09	-
	WTP_LCB	0.06	0.64	-0.34	-1.13	0.97	-0.33	-0.69	0.84	0.54	0.02	-0.44	-0.39	0.29	-0.41	0.43	-
	WTP_SQ	0.75	1.71	-0.77	-0.33	-0.14	-0.80	0.72	-0.49	0.77	0.31	-0.25	0.04	-0.14	0.17	-1.21	-

In Figure 3, imputed parameters are presented for 36 four-choice task packages. Most of the estimates of block-specific attributes remain relatively consistent across the study, despite the break between sessions in which respondent valuated those attributes. At the same time, for the overlapping attributes (LCB, CivContr, Wage and SQ), visible trends are observed. The parameter calculated for the opt-out alternative (SQ - right axis) varies due to the attributes block (see upper graph). The average parameter for wages increases across successive choice sets. This results in a decrease in the absolute value of most estimates calculated in the WTP space. This can be seen in Figure 3 (bottom graph).

Figure 3. *Imputed individual-level parameters (upper graph) and WTPs (bottom graph) across choice tasks*
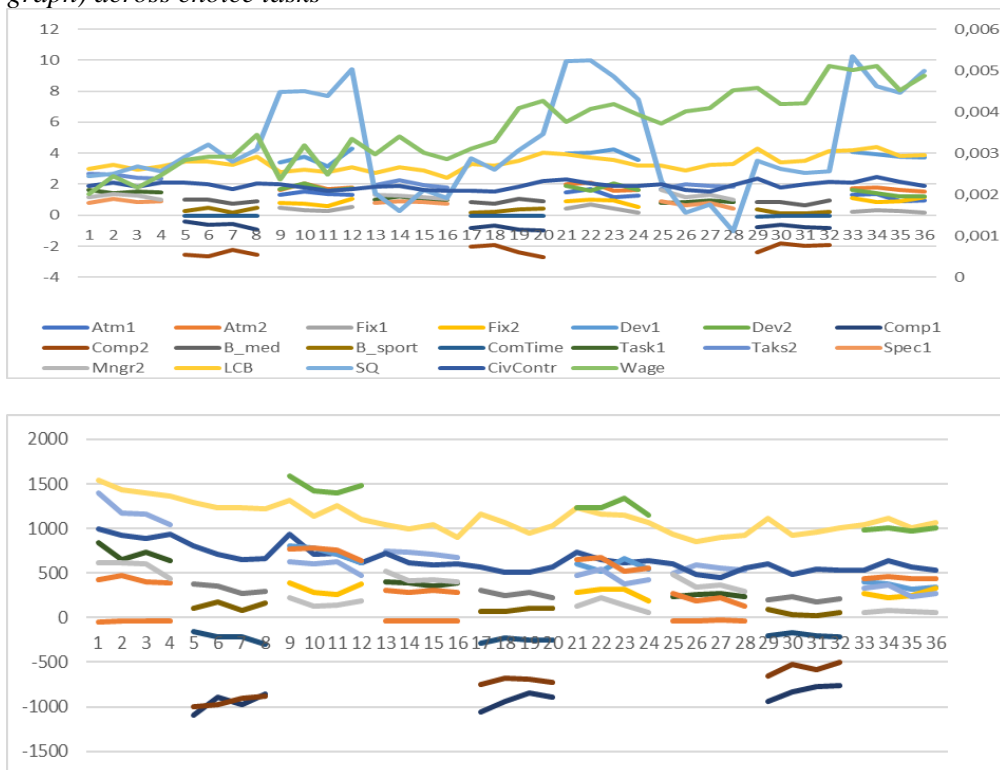
For the sake of brevity, WTPs and standard deviations for overlapping attributes are presented below (see the left graph). The mean WTP across choice tasks seem to manifest a downward (or u-shape) trend. At the same time, standard deviation increases slightly (for CivContr) or sharply (for LCB). This leads to the conclusion that preferences are somehow formulated during the process of choosing, with the simultaneous increasing variation in preferences. The coefficient of variation for those two variables increases, as can be seen in Figure 4 (see right graph). This effect might be linked to the fatigue effect in long-format DCEs.

As we observed considerable inter-respondent diversity in the data with respect to the choice set number, we assessed intra-individual heterogeneity. For each respondent, moving averages for WTP and standard deviation were calculated in four successive individual-level imputed parameters. As was expected, the average WTP for overlapping attributes across the population decreases and seems to asymptotically tend towards some target level (path marked with the dotted black line in Figure 5). This finding supports the DPH approach and choice-induced preferences hypothesis.

Figure 4. *Imputed individual-level parameters across choice tasks (left graph) and coefficient of variation (right graph)*
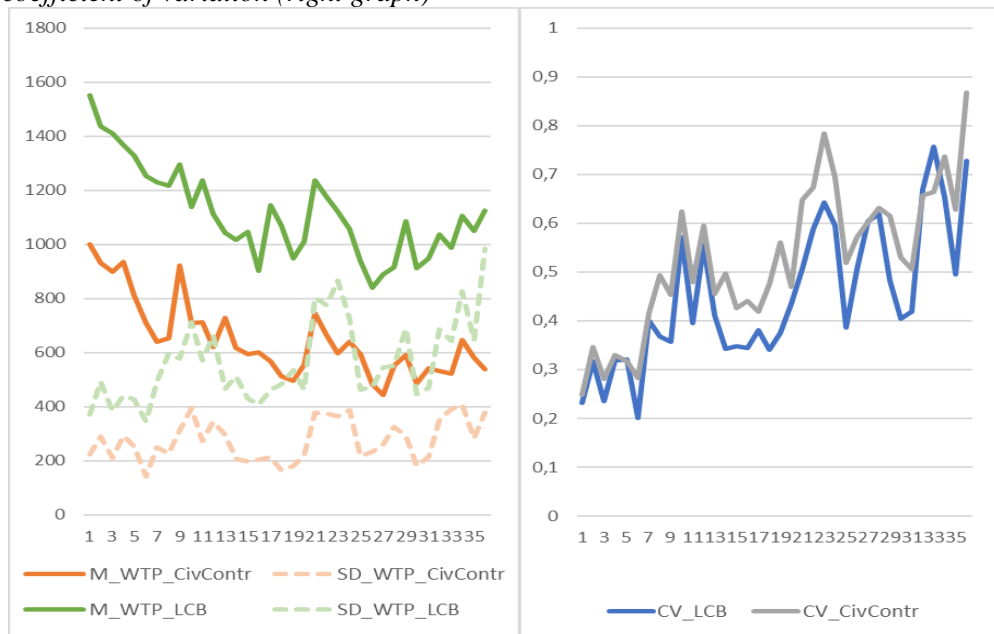

At the same time, moving standard deviation (calculated for successive sections of the panel) seems to be relatively stable (stable for CivContr attribute and slightly increasing for LCB) and is block-sensitive (Figure 5). The intra-respondent variance is much more stable than its inter-sample counterpart. This means that preferences revealed in successive choice tasks can systematically drift towards asymptotic values with stable variance. Therefore, the preference inconsistency that has been revealed in many studies can refer to sample polarization rather than to inconsistency of preferences.

5. Conclusions

Since DCE has become an accepted method in the valuation of non-market goods, it has been used widely to assess optimal policies across many fields of interest. Starting from the case of the Exxon Valdez tanker, the consequences of valuations

are achieved with microeconometric methods and thus methodological issues become extremely relevant (Brown, 2003). In particular, the question of the optimal number of choice sets per respondent has been widely discussed in literature. Many papers have found inconsistencies in preferences revealed in successive choice tasks, yet it is not clear which of these respondents' successive answers are the most credible and best reflect their preferences.

Figure 5. *Moving average and standard deviations for imputed individual-level WTPs for overlapping attributes*
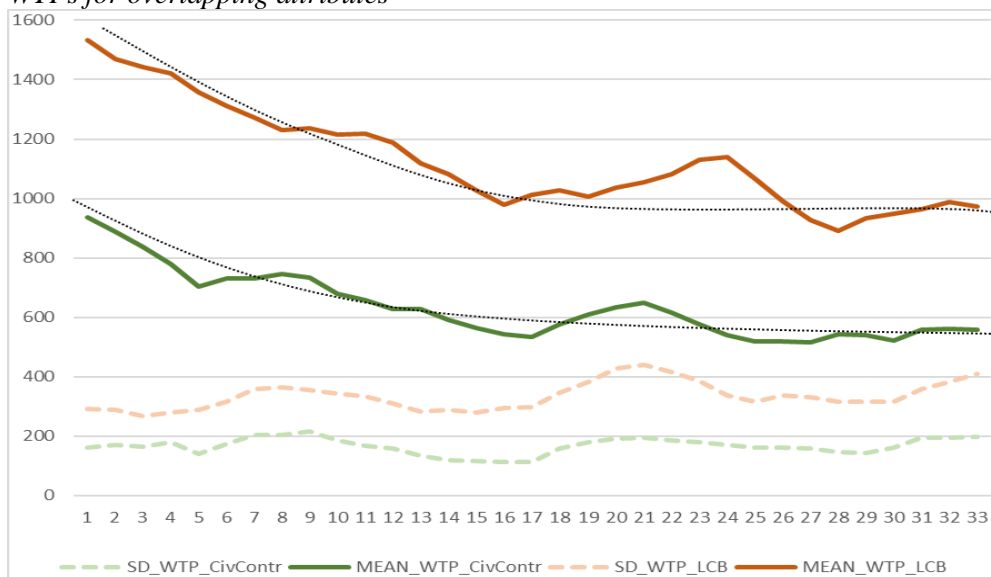
As is clear from this study, an appropriate answer to this question must be linked to the behavioural background of the decision-making process. As the DPH and choice-induced hypothesis state, preferences are not only revealed but also sometimes formulated during choice-making. From this point of view, to recognize well-formed, long-run preferences, one should observe as many choices as possible and assess the value towards which preferences converge asymptotically. On the other hand, ordering effects (such as fatigue) and survey length restrict multiplicity of choice tasks. Nevertheless, the issue of intra-respondent heterogeneity is fundamental for understanding preferences. Despite a massive body of work, most existing studies assume individual respondents' homogenous sensitivity across choice tasks, which we find to be an unrealistic assumption that can significantly bias the results of DCE studies.

In our investigation, we used 144 successive choice tasks for a labour supply empirical study. We observed substantial inconsistency in valuation of attributes in both preference and WTP spaces across choice sets, while other attributes were relatively stable. The estimates for different sessions are statistically different, but they do not differ significantly for the vast majority of successive choice tasks.

However, the farther away from one another the choice tasks were, the greater the observed difference between the estimates. In our results, therefore, a visible pattern of preference formulation is observed. At the same time, variance in the sample increases with the choice task number, which may be considered to be an indicator of respondents' fatigue. We propose a choice-making time instrument to partly censor the sample, thereby eliminating the fatigue effect.

Intra-respondent heterogeneity was assessed using imputed respondent-specific parameters. We found that the process of formulating preferences is accompanied by an increase in the diversity of the sample (which can be considered as an effect of fatigue) but, on the other hand and at the same time, intra-respondent preferences show a lower variance (than in the wider sample) across choice tasks. This leads to the conclusion that preferences are formulated at the respondent level according to a specific pattern. An increase in the standard deviation of parameters across the population does not necessarily mean inconsistency resulting from the 'negative' consequences of ordering effects, contrary to the findings of Swait and Adamowicz (2001). Instead, this can result from the polarization of preferences in the population alongside the intra-respondent consistency.

In this context, long-format DCEs can reveal some of the behavioural mechanisms behind the decision-making process. We show that, using this kind of study, it is possible to observe preference formulation. In our study, individual valuations of attributes seem to reach their asymptotically consistent values with relatively stable intra-respondent variance. Due to the cost and complexity of the conducted survey, this type of analysis is extremely difficult to run on a large sample. It seems, however, that in some specific cases obtaining accurate information, or even teaching respondents their preferences, can be a substantial result.

References:

Allingham, M. 2002. Choice theory: A very short introduction. OUP Oxford.

Andreoni, J. 1995. Cooperation in public-goods experiments: kindness or confusion? The American Economic Review, 85(4), 891-904.

Ariely, D., Loewenstein, G., Prelec, D. 2003. Coherent arbitrariness: Stable demand curves without stable preferences. The Quarterly Journal of Economics, 118(1), 73-106.

Ariely, D., Loewenstein, G., Prelec, D. 2006. Tom Sawyer and the construction of value. Journal of Economic Behavior & Organization, 60, 1-10.

Barkan, R., Ayal, S., Ariely, D. 2016. Revisiting Constructed Preferences: Extrapolating Preferences From Relevant Reminders. Decision, 3(4), 281-294.

Bateman, I., Burgess, D., Hutchinson, G., Matthews, D. 2008. Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. Journal of Environmental Economics and Management, 55, 127-141.

Braga, J., Starmer, C. 2005. Preference anomalies, preference elicitation and the discovered preference hypothesis. Environmental and Resource Economics, 32, 55-89.

Brehm, J.W. 1956. Postdecision changes in the desirability of alternatives. The Journal of Abnormal and Social Psychology, 52(3), 384-389.

Brown, T.C. 2003. Introduction to stated preference methods. [In:] A primer on nonmarket valuation, 99-110, Springer.

Carlsson, F., Mørkbak, M.R., Olsen, S.B. 2012. The first time is the hardest: A test of ordering effects in choice experiments. Journal of Choice Modelling, 5, 19-37.

Carson, R.T., Groves, T. 2007. Incentive and informational properties of preference questions. Environmental and resource economics, 37, 181-210.

Caussade, S., de Dios Ortúzar, J., Rizzi, L.I., Hensher, D.A. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation research part B: Methodological, 39, 621-640.

Chou, E., McConnell, M., Nagel, R., Plott, C.R. 2009. The control of game form recognition in experiments: Understanding dominant strategy failures in a simple two person "guessing" game. Experimental Economics, 12, 159-179.

Coppin, G., Delplanque, S., Cayeux, I., Porcherot, C., Sander, D. 2010. I'm No Longer Torn After Choice How Explicit Choices Implicitly Shape Preferences of Odors. Psychological Science, 21(4), 489-493.

Czajkowski, M., Giergiczny, M., Greene, W.H. 2014. Learning and fatigue effects revisited: Investigating the effects of accounting for unobservable preference and scale heterogeneity. Land Economics, 90, 324-351.

Day, B., Prades, J.L.P. 2010. Ordering anomalies in choice experiments. Journal of Environmental Economics and Management, 59, 271-285.

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., ... / Wang, P. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management, 63, 73-91.

Ellingson, L., Boman, M., Norman, J., Mattson, L. 2009. Valuation uncertainty and maximum willingness to pay: A nonparametric comparison between elicitation formats. Southern Swedish Forest Research Centre, Alnarp.

Festinger, L. 1957. Cognitive Dissonance Theory. Primary Prevention of HIV/AIDS: Psychological Approaches, Newbury Park, Sage Publications.

Hanley, N., Adamowicz, W., Wright, R.E. 2005. Price vector effects in choice experiments: an empirical test. Resource and Energy Economics, 27, 227-234.

Hanley, N., Kriström B., Shogren J.F. 2009. Coherent Arbitrariness: On Value Uncertainty for Environmental Goods. Land Economics, 85, 41-50.

Hess, S., Train, K.E. 2011. Recovery of inter- and intra-personal heterogeneity using mixed logit models. Transportation Research Part B, 45, 973-990.

Hess, S., Giergiczny, M. 2015. Intra-respondent heterogeneity in a stated choice survey on wetland conservation in Belarus: First steps towards creating a link with uncertainty in contingent valuation. Environmental and Resource Economics, 60, 327-347.

Hess, S., Hensher, D.A., Daly, A. 2012. Not bored yet–revisiting respondent fatigue in stated choice experiments. Transportation research part A: policy and practice, 46, 626-644.

Holmes, T.P., Boyle, K.J. 2005. Dynamic Learning and Context-Dependence in Sequential, Attribute-Based. Stated-Preference Valuation Questions, Land Economics, 81, 114-126.

Izuma, K., Saito, D.N., Sadato, N. 2010. Processing of the incentive for social approval in the ventral striatum during charitable donation. Journal of Cognitive Neuroscience, 22, 621-631.

Kahneman, D., Tversky, A. 1979. Prospect theory: An analysis of decision under risk. Econometrica, 47, 263-291.

Leotti, L.A., Delgado, M.R. 2011. The Inherent Reward of Choice. Psychological Science, 22, 1310-1318.

Mahieu, P.A., Riera, P., Giergiczny, M. 2012. The Influence of Cheap Talk on Willingness to Pay Range. Some Empirical Evidence. Journal of Environmental Planning and Management, 55, 753-763.

Payne, J.W., Bettman, J.R., Schkade, D.A. 1999. Measuring constructed preferences: Towards a building code. Journal of Risk and Uncertainty, 19, 243-270.

Plott, C.R. 1996. Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis. In Rational Foundations of Economic Behavior. K. Arrow, E. Colombatto, M. Perleman, and C. Schmidt, eds. London: Macmillan and NY: St. Martin's, 225-250.

Revelt, D., Train, K. 1998. Mixed logit with repeated choices: households' choices of appliance efficiency level. Review of economics and statistics, 80, 647-657.

Savage, S.J., Waldman, D.M. 2008. Learning and fatigue during choice experiments: a comparison of online and mail survey modes. Journal of Applied Econometrics, 23, 351-371.

Sharot, T., De Martino, B., Dolan, R.J. 2009. How choice reveals and shapes expected hedonic outcome. The Journal of Neuroscience, 29, 3760-3765.

Sharot, T., Fleming, S.M., Yu, X., Koster, R., Dolan, R.J. 2012. Is choice-induced preference change long lasting? Psychological science, 23, 1123-1129.

Sharot, T., Velasquez, C.M., Dolan, R.J. 2010. Do decisions shape preference? Evidence from blind choice. Psychological science, 21, 1231-1235.

Slovic, P. 1995. The construction of preference. American Psychologist 50, 364-371.

Street, D.J., Burgess, L. 2007. The construction of optimal stated choice experiments: theory and methods. John Wiley & Sons.

Swait, J., Adamowicz, W. 2001a. The influence of task complexity on consumer choice: a latent class model of decision strategy switching. Journal of Consumer Research, 28, 135-148.

Swait, J., Adamowicz, W. 2001b. Choice environment, market complexity, and consumer behavior: a theoretical and empirical approach for incorporating decision

    complexity into models of consumer choice. Organizational Behavior and Human Decision Processes, 86, 141-167.

Tang, C. 2012. Why do We Like What We Like? How Choice Shapes Preferences. Doctoral dissertation, Emory University.

Train, K. 2003. Discrete choice with simulation. Cambridge University Press.