IMPRECISE PREFERENCES AND ORDER EFFECTS IN TIME TRADE-OFF

Jose Luis Pinto Prades

Yunus Centre for Social Business & Health
Glasgow School for Business and Society
3rd Floor, Buchanan House
Glasgow Caledonian University
Cowcaddens Road
Glasgow, G4 0BA

joseluis.pinto@gcu.ac.uk

1. Introduction

One common finding in psychophysics is that measures of subjective sensations are heavily influenced by context. For example, if we ask two subjects who are in Glasgow on a sunny summer day to measure on a scale from 1 to 10 how hot a summer day is, we will probably receive two different responses if one subject is a native Glaswegian and another is a tourist from the south of Spain. The explanation is, obviously, that they may use different frames of reference. In general, when subjects are asked to translate into a number a subjective feeling they will frame their thoughts in a context to give meaning to their response.

This phenomenon has also been observed in the measure of preferences and it has been linked to the degree of imprecision/uncertainty in preferences. According to Simonson&Tversky[1] "if a consumer habitually purchases the same brand in a category, for example, context effects are unlikely to play a major role. In contrast, when people are uncertain about the values of options, they are more likely to use the context in determining the 'best buy'". It has been suggested [2] that a decision maker capable of meaningfully assessing absolute attribute values will conform to standard neoclassical preferences. However, this is not the most common situation especially when subjects have to evaluate unfamiliar objects like health states. If subjects are asked to provide an absolute value they will inevitable base their response on a relative and comparative judgment using an accessible reference point. Preferences become relative and the value of an object depends on the context.

Context effects represent a big challenge for the task of preference measurement. They put into question the principle of utility invariance that has been assumed to hold in health economics. For example, health technologies are evaluated using utilities from standard health profiles like the Euroqol or the Health Utility Index. They produce a unique, context-free, utility function that is used to evaluate health gains in very different contexts. Utility invariance is assumed when utilities are *elicited* and when utilities are *used*. In this paper, we will study to what extent utilities are context-dependent when they are elicited.

2. Context effects in the evaluation of health states

There is a contrast in the amount of literature on context effects in psychophysics, psychology and marketing compared to health economics. In health economics there has been more research about task effects, especially response mode effects [3-8], than about context effects (preference changes depending on the availability of other options or menu-dependence). We are only aware of two papers [9,10] that have analysed menu effects in the Visual Analogue Scale showing clear context effects. The most similar to our study is Schwarzinger et al[11] that found evidence of Assimiliation effects. The Time Trade Off (TTO) value of a moderate health state (M) was higher in the sequence Mild-Moderate than in the sequence Moderate-Severe. However, the limited sample size (namely 36 nurses divided in two subgroups) and the fact that they evaluated only one health state, makes it difficult to generalize these findings.

In this paper we will study context effects, more specifically sequence effects in the elicitation of preferences for health states using the TTO technique. We use the term "sequence effect" and not "order effect" because it has a wider meaning. When the same set of health states is presented in different order we can think on two kinds of order effects [12]. One is the position of the objects in the sequence (position dependent-order effects) and the other the nature of the options in preceding tasks (precedent-dependent order effects). For example, if we compare sequence A-B-C vs C-B-A the position of B does not change but the object evaluated previously changes. If object B is evaluated differently in both sequences we will attribute this to precedent-dependent order effects but not to position dependent-order effects.

We have chosen this specific type of context effect for several reasons. One is that in real life subjects usually evaluate several health states in a sequence. A second reason is that the study of sequence effects is useful to clarify the nature of preferences for health states. More specifically, given the evidence against the assumption that subjects have stable and well-developed preferences in many areas, it has been argued[13] that repetition generates preferences more in line with this assumption. Since sequences involve repetition, they can be used to analyze this issue. The third reason is that recent papers have shown that sequence effects can be an important problem. There are many elicitation methods (e.g. choice experiments) that rely on sequences of questions. These methods assume that each question is answered independently, that is, there are no sequence effects. However, some recent papers[14-16,] have shown that this assumption may have to be revised since there is evidence of order effects in sequences of choices. Finally, evidence of context-dependent preferences can have important implications on the way health benefits are estimated. The benefit of a health improvement is estimated as the distance between different points in an invariant scale. However, if the perception of the benefit is context dependent, this approach can be problematic.

In this study we provide evidence that preferences for health states are influenced by context. More specifically, we observe strong contrast effects. The perceived severity of a health state depends on the severity of health states evaluated previously. The evaluation of health seems to be highly relative.

3. Design and hypotheses

Assume we have a set of three rank-ordered health states such that X>Y>Z. We can build six different sequences with these three health states. The sequence X-Y-Z will be identified as "Descending", the sequence Z-Y-X will be identified as "Ascending", and the rest will be called "Mixed".

In our study, we use a combined within-between sample design (Table 1).

Table 1. Study design

		Order of TTO questions					
Block	Subgroups	1st	2nd	3rd	Change 1st to 2nd	Change 2nd to 3rd	Sequence
A	1	11211	22211	22222	-	-	Descending
	2	22211	11211	22222	+	-	Mixed
	3	22211	22222	11211	-	+	Mixed
	4	22222	22211	11211	+	+	Ascending
B	5	11311	22311	22322	-	-	Descending
	6	22311	11311	22322	+	-	Mixed
	7	22311	22322	11311	-	+	Mixed
	8	22322	22311	11311	+	+	Ascending
C	9	12211	12311	11311	-	+	Mixed
	10	12311	12211	11211	+	+	Ascending

Each subject evaluated three Euroqol health states using TTO. The 10 subgroups can be clustered in three blocks (A, B and C). Within each block all subjects evaluated the same health states. The only difference was the order of questions.

What kind of predictions can we make in relation to these sequences? The predictions depend on the assumptions we make about the nature of preferences:

1. Standard neoclassical preferences

If subjects have well-structured preferences (complete, coherent and stable) conforming to the standard neoclassical model:

a) The utility of a health state will be independent of the position in the sequence and on the nature of the previous health state, that is no sequence effects.
b) The ranking of utilities will correspond to the expect ordinal relationship between health states, that is U(X)>U(Y)>U(Z)

2. Discovered Preference Hypothesis (DPH)

Subjects may approach a TTO valuation task with some degree of imprecision but they learn about the valuation mechanism (institutional learning) and about the value of the good under investigation (value learning) with experience. If subjects learn through the elicitation process we could observe non-standard preferences at the beginning of the process and some convergence towards standard preferences later in the survey. For example, we could observe things such as respondent experiencing more problems discriminating between health states in the first question than in the third question. In the first question, subjects are still learning about the institution (the task) and about their own values. Utilities in the third question will look more in line with standard preferences.

3. Coherent Arbitrariness

Preferences are generally constructed at the time the valuation question and they are shaped by the interaction between the properties of the human information processing system and the properties of the decision task. Consistency is, to some extent, an artefact of the task. Under this approach subjects will be heavily influenced by previous evaluations and sequence effects will accumulate during the elicitation process. Two kinds of sequence effects have been observed in psychophysics; namely contrast effects and assimilation effects. Contrast effects imply that subjects evaluate two objects more differently when they are in a sequence than when they are evaluated in isolation. Assimilation effects predict the opposite. Contrast effects have been the main finding in the willingness to pay literature[17,18]. Apart from Contrast and Assimilation, a third potential context effect could be Loss Aversion. If a health state is evaluated after a better one, it could be perceived as a loss. However, if the health state is evaluated after a worse one, we should not observe any bias, at least not generated by Loss Aversion. Predictions according to these three effects are presented in Table 2.

Table 2. Patterns generated by different effects

Order of TTO questions			Contrast Effects		Assimilation effects		Loss Aversion	
1st	2nd	3rd	$^{*}U_{i2}- U_{i1}$	$U_{i3}- U_{i1}$	$U_{i2}- U_{i1}$	$U_{i3}- U_{i1}$	$U_{i2}- U_{i1}$	$U_{i3}- U_{i1}$
11211	22211	22222	-	-	+	+	-	-
22211	11211	22222	+	-	-	+	=	-
22211	22222	11211	-	+	+	-	-	=
22222	22211	11211	+	+	-	-	=	=
11311	22311	22322	-	-	+	+	-	-
22311	11311	22322	+	-	-	+	=	-
22311	22322	11311	-	+	+	-	-	=
22322	22311	11311	+	+	-	-	=	=
12211	12311	11311	-	+	+	-	-	=
12311	12211	11211	+	+	-	-	=	=

$^{*}U_{ij}$=utility of health state i in question j.
$U_{i2}- U_{i1}$=(-)/(+)/(=) the utility of the health state "i" elicited in 2nd position in the sequence is predicted to be lower/higher/the same than the utility of the same health state elicited in the 1st position in the sequence.

In summary, if we call the better health state (B) and the worst health state (W) we should find that:

a) According to Contrast Effects both U(B) and U(W) should be higher in the sequence W-B than in the sequence B-W.
b) According to Assimiliation Effects both U(B) and U(W) should be lower in the sequence W-B than in the sequence B-W.

c) According to Loss Aversion U(W) should be lower in the sequence B-W than in the sequence W-B while U(B) should be the same in both cases.

4. Methods

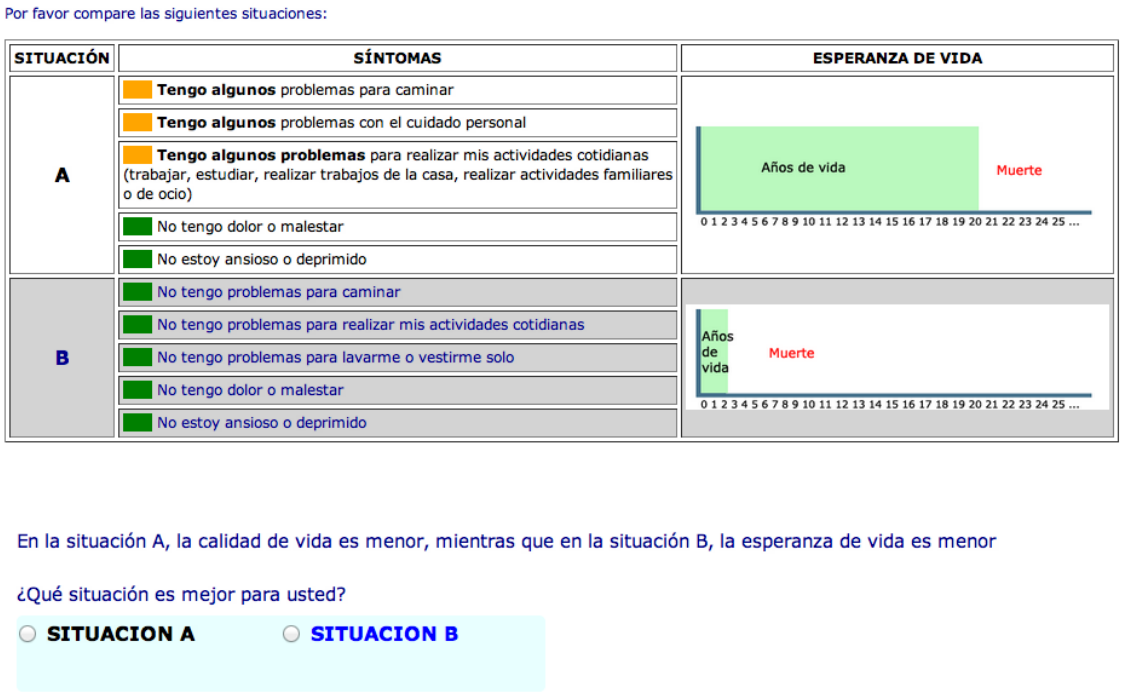
4.1. The survey

A web survey was conducted asking respondents to complete three TTO questions. Health states were described using the EQ-5D descriptive system which consists of five attributes (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), with three possible levels for each attribute. Level 1 refers to the best level in each attribute –so health state 11111 refers to full health and 33333 refers to the worst health state possible in the descriptive system. Subjects older than 65 years were excluded since we used duration of 20 years in the TTO and for subjects older than 65 this exceeded their life expectancy. The survey proceeded to explain subjects some details of the objectives of the study. It was explained that we wanted to know their perceptions about health problems and that there were no right or wrong answers. It was also explained that the survey was part of a research project of a Spanish university.

The survey started with an example of the kind of decisions they had to take. It was a choice between 20 years, in health state 22111 followed and 15 years in full health. They were told that any choice was legitimate. Subjects were randomized to one of the 10 treatments. All subgroups had the same structure (introduction, example, three TTO questions, personal characteristics) and the only difference between them was the health states evaluated and/or the order of questions. They were explained that they would have to respond to similar questions for three health states and they were shown those states.

A market research company was hired to recruit a sample of the Spanish population between 18 and 65 years of age. Initial contact was made via email. Individuals choosing to participate in the study were referred to the website that contained the survey. The market research company offers incentives to the subjects to complete questionnaires. Typically they receive points that they can change for some goods depending on the number of points they collect.

Figure 1. Display used



4.2. Elicitation procedure.

We used a choice-based procedure in order to estimate utilities. An example of the display used can be seen in Figure 1. Subjects were asked to choose between 20 years in bad health and fewer years in full health. Through a series of choices we estimated an interval where indifference between alternatives was located. The sequence of choices did not follow a clear pattern, e.g. a ping-pong or midvalue splitting technique. The number and nature of choices was independent of the previous response. The procedure always started with a choice between (2 years, full health) and (20 years, bad health). If subjects chose (20 years, bad health), they had to make four choices in a random order, namely they had to choose between (20 years bad health) and (6/10/14/18 years, full health). If there was a contradiction the system asked the subject to solve it. These choices made it possible to estimate an interval (of 2 or 4 years) where the indifference point was located. Once this first interval was estimated more choices followed (between one to three) in order to locate the interval within a one-year range. The process ended with an open question that asked subjects to state the number of months. For example, if choices revealed that subjects preferred (20 years, bad health) to (15 years, full health) and preferred (16 years, full health) to (20, bad health), we knew that indifference was between 15 and 16 years in full health. The open question asked about the number of months within this interval. If subjects chose (2 years, full health) in the first question they were asked if (20 years, bad health) was preferred to immediate death. If they preferred immediate death utilities were elicited asking people to choose between immediate death and several profiles where the duration in full health and bad health was changed until indifference was reached using a procedure similar as the one used for health states better than death.

5. Results

5.1. The sample

A sample of 6003 members of a market research panel was initially invited by email to participate in the survey. Of these, 2016 (33.6%) consented to begin the survey, 251 were excluded for excess quota leaving and 270 did not complete the survey. The final sample was then of 1495 subjects that were randomly allocated to 10 treatments according to the design of the study. All groups had similar characteristics.

5.2. Inconsistencies

Table 4 shows the consistency of responses. We say that there is an inconsistency when the utility of the dominant health state is lower than the utility of the dominated health state (and vice versa). We also have a column (Indif) for people who gave the same response for two different health states.

Table 3. Consistency of responses (%)

				1st vs 2nd valuation			2nd vs 3rd valuation		
Group	1st	2nd	3rd	CONSIS	INCON	INDIF	CONSIS	INCON	INDIF
1	11211	22211	22222	64.1	13.1	22.9	80.4	7.8	11.8
2	22211	11211	22222	62.9	14.6	22.5	66.9	15.2	17.9
3	22211	22222	11211	63.7	12.3	24.0	74.0	11.0	15.1
4	22222	22211	11211	70.1	9.6	20.4	80.9	4.5	14.6
5	11311	22311	22322	51.7	21.8	26.5	66.0	11.6	22.4
6	22311	11311	22322	53.4	24.3	22.3	75.0	8.1	16.9
7	22311	22322	11311	60.8	18.9	20.3	74.3	8.1	17.6
8	22322	22311	11311	56.5	23.1	20.4	53.7	19.7	26.5
9	12211	12311	11311	77.2	11.7	11.0	26.9	33.8	39.3
10	12311	12211	11211	81.0	6.5	12.4	54.2	5.2	40.5

We can observe that the degree of inconsistency is not random. There are some comparisons that are more difficult to make. We can see that:

a) In Block B (groups 5 to 8) subjects are less consistent than in Block A. Increasing the severity from level 2 to level 3 in the third dimension made the valuation task more difficult. In fact, at the end of the survey subjects were asked about the degree of difficulty of the task. While 12.0% considered the task difficult or very difficult in Block A, this increased to 18.5% in Block B.
b) There are more inconsistencies between the first and second question than between the second and third question. This could be interpreted as evidence in favour of the DPH, namely people understand the task better and clarify their minds about their values, giving more consistent responses as they respond to more questions. This interpretation is put into question by the next observation.
c) It is in the comparison between the second and third question of group 9

when we find the largest problem of inconsistency (33.8%). It could be thought that there is some problem with this group. This does not seem to be the explanation of this result since the number of inconsistencies was much smaller (11.7%) between questions 1 and 2 when, theoretically, their preferences were less well-structured. The explanation seems to be that subjects perceive health states 12311 and 11311 very similarly in terms of severity. Since both have in common the level 3 in the third dimension, an improvement from *Moderate* to *No Problems* in the second dimension does not modify the perception of severity, generating the highest level of errors in all groups and tasks.

d) This result is reversed for group 10. That is, it is in the comparison between the second and third task of group 10 when the percentage of inconsistencies is the smallest (5.2%). This is surprising since the "distance" between health states evaluated in the second and third place is the same in groups 9 and 10. In both cases, *moderate problems* are transformed into "*no problems*" in the second dimension. There seems to be something else apart from the distance between health states or the position in the survey that is generating such a different level of inconsistencies. This could be explained in terms of Evaluability/Saliency; namely the existence of one level "3" is the salient feature of the health states.

Following this analysis of inconsistencies we decided not to exclude any subject from further analysis for two reasons:

1. No subject gave inconsistent responses in all pairwise comparisons.
2. There are theoretical arguments to keep all subjects in the analysis[19].

In order to reduce the problems generated by inconsistencies, non-traders and extreme negative utilities we will use medians instead of means as measures of central tendency.

5.3. Main results

The utilities of the health states can be seen in Table 4. We use non-parametric tests to analyse statistical significance (Mann-Whitney). Results were internally consistent within each group. Median utilities were statistically different ($p<0.05$) within groups and were internally consistent; that is, they followed the logical order predicted by dominance.

Table 4. Median utilities

	Order of questions			Utilities (medians)		
	1st	2nd	3rd	U_1	U_2	U_3
1	11211	22211	22222	.896	.775	.496
2	22211	11211	22222	.825	.904	.654
3	22211	22222	11211	.796	.696	.946
4	22222	22211	11211	.737	.896	.975
5	11311	22311	22322	.633	.496	.125
6	22311	11311	22322	.694	.794	.492
7	22311	22322	11311	.675	.496	.896
8	22322	22311	11311	.542	.725	.846
9	12211	12311	11311	.875	.496	.546
10	12311	12211	11211	.696	.908	.996

5.3.1. Testing the hypothesis of well-structured preferences: the first response

If subjects have well defined preferences from the beginning of the survey we will find statistically significant differences in the direction predicted by dominance in the first valuation task. Before conducting these tests, we compared the utilities of the groups that evaluated the same health state in first place, namely groups 2-3 and 6-7. There were no statistically significant differences between groups 2 and 3 and between groups 6 and 7. For this reason, we pooled these four groups in two (2+3, 6+7) for the first question. The main results can be seen in Table 5.

Table 5. Tests of significance in first response: dominant vs dominated options

Comparison	Group	Health state	vs	Group	Health state	p-value
1	1	11211		9	12211	0.19
2	9	12211		2+3	22211	0.97
3	1	11211		2+3	22211	0.00
4	2+3	22211		4	22222	0.03
5	5	11311		10	12311	0.67
6	10	12311		6+7	22311	0.83
7	5	11311		6+7	22311	0.75
8	6+7	22311		8	22322	0.57
9	1	11211		5	11311	0.00
10	2+3	22211		6+7	22311	0.00
11	4	22222		8	22322	0.02
12	9	12211		10	12311	0.00

There are several things that we can be observed in Table 5:

1. A change from “1” to “2” in one single dimension did not produce any difference (comparisons 1, 2, 5, 6).
2. If the difference was bigger (two “2”) the result was mixed:
    a. If there was not a “level 3” present in the comparison, subjects perceived the health states as different (comparisons 3 and 4).
    b. If there was a “level 3” present in both arms of the comparison, subjects did not perceived the health states as different (comparisons 7 and 8). This reinforces the idea that the presence of

a level "3" seems to focus the attention of subjects and they fail to perceive the changes between levels "1" and "2".

3. In all cases subjects discriminated between health states when one health state had a "3" and the other did not (comparisons 9 to 12).

The first picture that we have is that health states that are different are not always perceived as such. Preferences do not seem well-defined from the beginning of the survey.

5.3.2. Preference adjustment: second and third responses

The question we address now is to what extent some of the problems we have seen in the first set of responses are reduced (or simply disappear) when subjects are more familiar with the tasks. From now on, the utility of a health state (X) will be identified as $U(X)_{ij}$ where i is the group and j is the order of the question.

We now compare utilities of health states evaluated in second place. Do we observe that subjects produce more consistent responses? In fact, we observe that some health states that were not perceived as different in the first question are now statistically different. This happens for 22322 and 11311, namely $U(22322)_{52} < U(11311)_{62}$ (p=0.00). However, we find new inconsistencies now, which were not observed in the first question. They are:

1. $U(22211)_{12}<U(22211)_{42}$ (p=0.00) and $U(22311)_{52}<U(22311)_{82}$ (p=0.00).
2. $U(22211)_{12}=U(22311)_{82}$ (p=0.14) and $U(11211)_{12}=U(22211)_{42}$ (p=0.54).

These problems are also observed in the third question. More specifically we observe different utilities for identical health states:

1. $U(22222)_{13}< U(22222)_{23}$ (p=0.052)
2. $U(11211)_{10,3}> U(11211)_{33}$ (p=0.057)
3. $U(11211)_{10,3}> U(11211)_{43}$ (p=0.01)
4. $U(22322)_{53}< U(22322)_{63}$ (p=0.03)
5. $U(11311)_{73}< U(11311)_{93}$ (p=0.00)
6. $U(11311)_{83}< U(11311)_{93}$ (p=0.00)

These results do not suggest that as the survey advances subjects understand better their preferences and utilities are more consistent with standard preferences. Given that Standard Preferences and DPH cannot explain our data we now turn to Context Effects.

5.3.3. Context Effects

We try to explain context effects according to the patterns developed above (Table 2). Results can be seen in Table 6. For example, let us focus on group 6. The health state they evaluated in second place was 11311 (second column, Table 6). The median utility was 0.794 (Group 6, Table 4). However, when this health state was evaluated in the first position the median utility was 0.633 (Group 5, Table 4). The

difference in medians was 0.161 (third column [U_2-U_1] Table 6). The health state that subjects of Group 6 evaluated in the first place was 22311 (a worst health state than 11311: W in column 4 of Table 6). According to Contrast Effects if health state 11311 is evaluated after a worse health state (222311) it will not look as bad as when it is evaluated in the first place. For this reason, Contrast Effects will predict that U(11311) in Group 6 will be higher than utility of U(11311) in Group 5 (then a "+" in column 'Contrast' of Table 6). Assimilation effects will predict the opposite and Loss Aversion will not predict any bias (then a "=" in column 'Loss Aversion'). In this case, the result supports Contrast as an explanation of our result.

Table 6. Testing context effects.

Group	Health state 2nd valuation	U_2-U_1	Previous health state	Contrast	Assimilation	Loss Aversion
1	22211	- 0.021^{ns}	B	-	+	-
2	11211	0.008**	W	+	-	=
3	22222	- 0.041^{ns}	B	-	+	-
4	22211	0.100***	W	+	-	=
5	22311	-0.230***	B	-	+	-
6	11311	0.161***	W	+	-	=
7	22322	- 0.046^{ns}	B	-	+	-
8	22311	0.031*	W	+	-	=
9	12311	- 0.200^{ns}	B	-	+	-
10	12211	0.033***	W	+	-	=
Group	Health state 3rd valuation	U_3-U_2	Previous health state	Contrast	Assimilation	Loss Aversion
1	22222	-0.241***	B	-	+	-
2	22222	-0.083^{ns}	B	-	+	-
3	11211	+0.05***	W	+	-	=
4	11211	+0.079***	W	+	-	=
5	22322	-0.417***	B	-	+	-
6	22322	-0.05**	B	-	+	-
7	11311	+0.263***	W	+	-	=
8	11311	+0.213***	W	+	-	=
9	11311	-0.087^{ns}	W	+	-	=
10	11211	+0.1***	B	+	-	=

*p<0.1. **p<0.05, ***p<0.001 Mann-Whitney Test

We can see in Table 6 that differences between utilities were always in the direction predicted by Contrast Effects. The severity of a health state is inversely related with the severity of the health state evaluated previously. If the previous health state is better, the utility of the health state is lower than if the health state is evaluated in the first place/position, when subjects are not influenced by any other previous valuation task. If the previous health state is worse, the utility of the health state is higher than if the health state is evaluated in the first place/position. Our results suggest that there is some degree of Coherent Arbitrariness in the valuations of health states. We do not mean by this that utilities are totally

arbitrary. For example, the utility of 11211 is always higher than the utility of 22311 or 22322. There does not seem to be any confusion here. What the principle of Coherent Arbitrariness implies is that the overall picture looks more consistent than it really is. To understand this better, let us use an example. Imagine a researcher that wants to estimate utilities for health states A, B and C where A>B>C. Our researcher is aware of the potential influence of order effects and she follows a similar strategy to ours, namely splits the sample in several subsamples and randomises the order of questions. She conducts the survey and pools together all data. She hopes that order effects may cancel each other out[1]. In our case, she would have got the following results (utilities in medians) for groups 1 to 8:

1. Block A: U(11211)=0.93, U(22211)=0.83, U(22222)=0.65
2. Block B: U(11311)=0.79, U(22311)=0.65, U(22322)=0.41

That would imply perfectly coherent results. It would look as if subjects can perfectly discriminate between all health states in the direction predicted by theory. However, that would not be the correct interpretation of the data. We know that part of this perfect discrimination could be attributed to Contrast Effects. In our case, these effects increase the "distance" between health states in the same direction as theory would predict. Contrast effects increase the utility of good health states and reduce the utility of bad health states increasing the distances in utilities in an arbitrary (coherent) way. In summary, part of the consistency in the pooled results could be attributed to Coherent Arbitrariness and not to well-structured preferences.

5.4. Context effects in more basic preferences

To end the analysis of our data we will see if Contrast Effects may also affect more "basic preferences". We analyse two kinds of 'basic preferences':

1. Non-traders: a health state is so mild that subjects are not willing to trade-off any length of life in order to improve quality of life.
2. Worse than death: a health state is so bad that subjects prefer to die immediately.

Can Contrast Effect modify these "basic preferences"? Results can be seen in Table 7.

[1] The researcher will make a mistake since randomization can only eliminate position-dependent order effect but it does not guarantee, as it happens in our case, that precedent-dependent order effects will cancel each other out.

Table 7. Contrast effects and basic preferences.

Group				Worse than death (%)			Non-traders (%)		
1	11211	22211	22222	2.6	5.2	19.0	18.3	13.7	7.2
2	22211	11211	22222	6.6	2.6	11.3	13.9	29.8	10.6
3	22211	22222	11211	8.2	7.5	3.4	13.7	8.9	37.7
4	22222	22211	11211	8.9	2.5	1.9	10.8	27.4	44.6
5	11311	22311	22322	13.6	19.0	29.3	8.2	4.8	4.1
6	22311	11311	22322	18.9	13.5	34.5	9.5	20.9	6.8
7	22311	22322	11311	15.5	17.6	10.1	6.1	4.7	21.6
8	22322	22311	11311	19.0	12.9	10.2	10.9	12.2	21.1
9	12211	12311	11311	5.5	13.1	13.8	13.8	6.9	6.2
10	12311	12211	11211	15.0	1.3	0.7	10.5	27.5	50.3

Contrast Effects also modify these basic attitudes although its influence is restricted to some cases. Contrast Effects change the consideration of a health state as better as or worse than death. This effect is concentrated in the worse health state in each block (22222 in Block A and 22322 in Block B). In Block A, the percentage of subjects who consider (22222) worse than death moves from 8.9% to 19% and in block B, the percentage of subjects who consider (22322) worse than death changes from 19% to 34.5%. In the case of non-trading preferences, Contrast Effects affect the mildest health states (11211 and 22211). Non-traders move from 18.3% to 50.3% for health state (11211) and from 13.7% to 27.4% for health state (22211).

In summary, in the case of bad health states, people seem to have problems in deciding if a health state is better or worse than death but they do not show any doubt about trading length expectancy and quality of life. In the case of the mildest health states, it is the opposite. They have problems in deciding if they have to trade of not length expectancy and quality of life but they seem to know very clearly that the health state is not worse than death. This reinforces the hypothesis that context effects are relevant when subjects have doubts about what to do (imprecision).

6. Discussion

The main result of this paper is that health state utilities are context dependent to some extent. Subjects have some basic intuitions about the severity of a health state but their preferences do not seem very precise. For this reason, they rely on relative comparisons in order to respond to TTO questions. More specifically contrast effects predominate over other effects. The question we address in this final section is what to do about it. The literature has identified several alternatives in order to deal with Context Effects in general and with Sequence Effects more specifically. We will see them in turn.

The first option in order to avoid Sequence Effects is to ask subjects to evaluate only one health state. However, this is not the perfect solution. One problem is that this is a very inefficient method both financially and statistically. The

second problem is that this is only valid if we assume that preferences are well defined and they are revealed the first time that subjects evaluate an object. The probability of revealing the true value will be higher when the evaluation of one object is not "contaminated" by anything else; that is, when it is evaluated at the beginning of the sequence. However, our results do not seem to support this perspective since utilities elicited at the beginning of the sequence in groups 5 to 8 do not discriminate between health states that are clearly different.

Another option is to ask several questions but to model context effects such that utility invariance is preserved. That is, given that utilities will change with the context, theory is used to disentangle context-effects and true preferences. This is approach was used by Birnbaum [20] who derived a context-free psychophysical function using range-frequency theory in order to explain context effects. This was applied to social issues[21] and to monetary lotteries[22] leading to the conclusion that utility invariance can be preserved if context effects can be modelled[23]. The Componential Context Model[24] also tries to separate these two components, one representing value maximization and another the context-dependent component of utility. This model has been recently applied in marketing[25].

However, some researchers have rejected utility invariance. Dellaert et al[26] show that preferences change with the complexity of the menu (number of alternatives, number of attributes and similarity). One interesting feature of this paper is that it is applied to a very simple and familiar product (yogurts) where the hypothesis of well-structured preferences should hold. If context changes the intrinsic utility all we can do is to estimate different utilities in different contexts.

In the case of sequence effects we can apply this dual approach depending on the perspective we take about preferences:

a) Subjects have well-defined preferences from the beginning but:

i. The reference point changes as we move along the sequence of tasks (reference point revision). The utility function does not change in the whole process (utility invariance). However subjects change the reference point and this translates into different weights for different characteristics of the problem. The solution to this problem is to debias all responses except the first one[27].
ii. They change during the sequence. The "true" preferences are really different at different stages. This has been shown to happen in cases where subjects actually use goods in a sequence. Previous consumption influences the utility of a good at a point in time. In this there is no context effect just a change in preference. Models of habit formation are particularly appropriate to model these effects. In this case, the solution is to estimate different utilities for different parts of the sequence[28].
iii. The task is new for them. As they become more familiar with the elicitation procedure they give responses that better reflect their preferences

(institutional learning). In this case, the solution is to give more weight to preferences elicited at the end of the sequence[27].

b) Subjects do not have well defined preferences when they start the survey and:

a. They refine and change their preferences (preference learning) as they respond to more questions until they become stable and well formed. In this case, the solution is to give more weight to preferences elicited at the end of the sequence.
b. The characteristics of the options evaluated at the beginning of the process completely determine how favourably or unfavourably each option is evaluated in relation to previous options. The information provided by the context shapes preferences. In this case, following Tversky and Simonson[24] we could make a distinction depending on whether it is or is not rational to infer relevant information from the context.

    i. Context-relevant information. Assume that a consumer has to choose between two objects (X and Y) that differ on two attributes, namely price (P) and quality (Q). The consumer is not sure if the difference in price $P_X > P_Y$ justifies the difference in quality $Q_X > Q_Y$. The consumer is not sure what to do and observes a third (Z) similar object (P_Z,Q_Z). This generates new comparisons and helps the subject to understand if paying a higher price for X is a 'good deal' or a 'bad deal'. The new information helps the consumer to shape her preferences. Although this process violates the principle of independence of irrelevant alternatives, it can help the consumer to take a good decision. This process can be rational in the sense that the context provides valid information in order to take a decision. If X, Y and Z are (e.g.) washing machines it is not a fundamental problem if we do not have well-established preferences between price and quality before visiting the shop. Our preferences will be built on the spot and they will be based on relative comparisons, but at the end of this process we can build a "strong" set of values about washing machines.
    ii. Context-irrelevant information. Assume that after buying the washing machine our consumer has to decide between two different brands of chocolate bars. Since the consumer has decided to pay 100€ more for the better machine, he/she has a tendency to buy the most expensive chocolate bar since it only costs 0.5€ more per unit. Paying 0.5€ more is nothing if one has paid 100€ more for another product. This kind of context effects could not be considered rational since the information provided by the context is not relevant.

Having said all that: what can we do in our case? If we could find a good model that explains contrast effects, we could try to disentangle the context-dependent and the context-independent components in order to keep the principle of utility invariance in the evaluation of health states. One candidate could be Adaptation Level Theory[30]. This theory claims that objects are evaluated in relation to some

adaptation level or norm that depends on the mean of previously evaluated objects. The lower the adaptation level (the lower the value of recently evaluated objects) the higher the value of the object evaluated. In our case, this translates into different perceptions of the severity of a health state according to the severity of previously evaluated health states. Wedell[31] models Adaptation Level theory as a weighted average of context-dependent and context-independents values.

This approach has two problems. The first one is that if the theory is not right the utilities will be biased. The second one is that when these context-free utilities are used in practice context-effects will again play a role. That is, we can get context-free utilities but we will find again the problem of context effects when we apply those utilities to take decisions in the real world.

Another approach could be to ask subjects to evaluate health states in what we have called a "relevant" context. If the evaluation of health states is largely governed by relative comparisons, if context effects are inevitable, if decisions are taken in a context, we could incorporate those effects when preferences are elicited and when they are used. For example, eliciting preferences in a joint evaluation mode instead of using a separate evaluation mode is a step in this direction [32]. The notion of "choice bracketing"[33] is also related to the idea that subjects need to use assess the consequences of their actions in a broad context in order to take good decisions[34]. Context effects are an inherent element in the way that subjects understand and interpret situations. Context not only helps people to understand some predetermined set of values (preferences), it can provide subjects with a wider perspective that helps to take better decisions. For this reason, methods to elicit preferences or to establish priorities can try to use "relevant" contexts in order to improve the way that public preferences shape public policy.

In summary, this paper makes a contribution to an area that has not received enough attention in health economics. It shows that sequence effects can be relevant. They have wide consequences on the methods that are used to elicit preferences and to establish priorities.

REFERENCES

1. Simonson I. Tversky A. Choice in context: Tradeoff contrast and extremeness aversion. Journal of Marketing Research 1992;29(3):281-95.
2. Simonson, I. Will I like a 'Medium Pillow'? Another look at Constructed and Inherent Preferences. Journal of Consumer Psychology, 2008, 18, 155-169.
3. Bleichrodt H. and Johannesson M. 1997. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. J Health Econ 16(2):155-75.
4. Lenert L. A., Cher D. J., Goldstein M. K., Bergen M. R. and Garber A. 1998. The Effect of Search Procedures on Utility Elicitations Medical Decision Making 18(1):76-83.

5. Stalmeier P., Wakker P. and Bezembinder T. 1997. Preference Reversals: Violations of Unidimensional Procedure Invariance Journal of Experimental Psychology:Human Perception and Performance 25(4):1196-205.
6. van Osch S. M. C., Wakker P. P., van den Hout W. B. and Stiggelbout A. M. 2004. Correcting Biases in Standard Gamble and Time Tradeoff Utilities. Medical Decision Making 24(5):511.
7. Sumner W. and Nease R. F. 2001. Choice-Matching Preference Reversals in Health Outcome Assessments Medical Decision Making 21(3):208-18.
8. Oliver A. 2006. Further evidence of preference reversals: Choice, valuation and ranking over distributions of life expectancy. Journal of health economics 25(5):803-20.
9. Robinson, A, G Loomes, and M Jones-Lee. 2001. Visual analog scales, standard gambles, and relative risk aversion. Med Decis Making 21 (1): 17-27.
10. Bleichrodt, H and M Johannesson. 1997. An experimental test of a theoretical foundation for rating-scale valuations. Med Decis Making 17 (2): 208-16.
11. Schwarzinger, M, T Herz, I Durand-Zaleski, M S de Pharmacologie Clinique, and P Sociale. 2008. The sequence effect in time trade-off elicitation . In Cost-Effectiveness Analysis to Aid or Not to Aid Health Policy Makers. Feasibility and Implications of Health Valuation Methods in France and Egypt. Unpublished Thesis.
12. Day, Brett, Ian J Bateman, Richard T Carson, Diane Dupont, Jordan J Louviere, Sanae Morimoto, Riccardo Scarpa, and Paul Wang. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management 63 (1): 73-91.
13. C.R. Plott, Rational individual behavior in markets and social choice processes: the discovered preference hypothesis, in: K. Arrow, E. Colombatto, M. Perleman, C. Schmidt (Eds.), Rational Foundations of Economic Behavior, 1996, Macmillan, London, St. Martin's, New York, pp. 225-250.
14. Day, Brett and Jose Luis Pinto-Prades. 2010. Ordering anomalies in choice experiments. Journal of Environmental Economics and Management 59:271-285.
15. Day, Brett, Ian J Bateman, Richard T Carson, Diane Dupont, Jordan J Louviere, Sanae Morimoto, Riccardo Scarpa, and Paul Wang. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management 63 (1): 73-9.
16. McNair, Ben J, David A Hensher, and Jeff Bennett. 2012. Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: A probabilistic decision process model. Environmental and Resource Economics 51 (4): 599-616.
17. Powe, N. A., & Bateman, I. J. (2003). Ordering effects in nested top-down and bottom-up contingent valuation designs. Ecological Economics, 45(2), 255-270
18. Clark, J., & Friesen, L. (2008). The causes of order effects in contingent valuation surveys: An experimental investigation. Journal of Environmental Economics and Management, 56(2), 195-206.

19. Lancsar, E. and JJ Louviere (2006). Deleting irrational responses from discrete choice experiments. Health Economics 15, 8, 797-811.
20. Birnbaum, M H. 1974. Using contextual effects to derive psychophysical scales. Attention, Perception, & Psychophysics 15 (1): 89-96.
21. Mellers, B A and M H Birnbaum. 1983. Contextual effects in social judgment. Journal of Experimental Social Psychology 19 (2): 157-171.
22. Mellers, B A, L D Ordoñez, and M H Birnbaum. 1992. A change-of-process theory for contextual effects and preference reversals in risky decision making. Organizational Behavior and Human Decision Processes 52 (3): 331-369.
23. Mellers, B A, L D Ordoñez, and M H Birnbaum. 1992. A change-of-process theory for contextual effects and preference reversals in risky decision making. Organizational Behavior and Human Decision Processes 52 (3): 331-369.
24. Tversky, Amos and Itamar Simonson. 1993. Context-Dependent preferences. Management Science 39 (10): 1179-1189.
25. Rooderkerk, R P, H J Van Heerde, and T H A Bijmolt. 2011. Incorporating context effects into a choice model. Journal of Marketing Research 48 (4): 767-780.
26. Dellaert, B, B Donkers, and A Soest. 2012. Complexity effects in choice experiment-based models. Journal of Marketing Research XLIX (ja): 424-434.
27. DeShazo, J R. 2002. Designing transactions without framing effects in iterative question formats. Journal of Environmental Economics and Management 43 (3): 360-385.
28. Swait, J, W Adamowicz, and M Bueren. 2004. Choice and temporal welfare impacts: Incorporating history into discrete choice models. Journal of Environmental Economics and Management 47 (1): 94-116.
29. Bateman, Ian J, Diane Burgess, W George Hutchinson, and David I Matthews. 2008. Learning design contingent valuation (LDCV): NOAA guidelines. Journal of Environmental Economics and Management 55 (2): 127 - 141.
30. Helson, H. 1964. Current trends and issues in adaptation-level theory. American Psychologist 19 (1): 26-38.
31. Wedell, D H. 1995. Contrast effects in paired comparisons: Evidence for both stimulus-based and response-based processes. Journal of Experimental Psychology: Human Perception and Performance 21 (5): 1158.
32. Hsee, C K and J Zhang. 2010. General evaluability theory. Perspectives on Psychological Science 5 (4): 343.
33. Rabin, Matthew, Daniel Read, and George Loewenstein. 1999. Choice bracketing. Journal of Risk and Uncertainty 19:171-197
34. Rabin and Weizsäcker. 2009. Narrow bracketing and dominated choices. American Economic Review 99 (4): 1508-1543.