Do we choose differently after a discussion?

Results from a deliberative valuation study in Ireland

Margrethe Aanesen, Thomas M. van Rensburg, Claire W. Armstrong

Professor Margrethe Aanesen, UiT Arctic University of Norway, PO Box 6050 Langnes, N-9037 Tromsø, Margrethe.aanesen@uit.no,

Professor Claire W.Armstrong, UiT Arctic University of Norway, PO Box 6050 Langnes, N-9037 Tromsø, Claire.armstrong@uit.no

Senior lecturer Thomas Van Rensburg, Dept. of Economics and Cedres, 1st floor, St.Anthony's, National University of Ireland (NUI), University Road, H91 TK33 Galway, Ireland, thomas.vanrensburg@nuigalway.ie

Abstract

A criticism against individual-based stated preference surveys is that people seldom make choices in social isolation. An approach used in response to this is deliberative valuation. We conduct a series of valuation workshops including a choice experiment. This involves providing respondents with information and opportunity to discuss. They made their choices individually both before and after the discussion. On average respondents stated different preferences after the discussion compared to before, and the stated preferences became more consistent after deliberation. Unexpectedly, persons stating high certainty of pre-deliberation choices changed stated preferences more after deliberation than persons stating lower certainty of choices.

JEL: Q51, Q57

Key words: stated preferences; choice experiment; deliberative methods; valuation workshops; group discussion; self-reported certainty; knowledge of good; unfamiliar goods; robustness of stated choices; consistency of choices

1 Introduction

The stated preference (SP) method is a non-market valuation method used extensively to value a variety of non-market goods and services (Haneman, 1994; Carson et al., 1998). It relies on the explanation of the hypothetical good being valued, and provision of information about this good and about the survey instrument is crucial for consistent and reliable data (Hoehn and Randall, 1987). Different formats have been used to convey such information including stand-alone individual surveys (Haneman, 1994), video-enhanced surveys (Sandorf et al, 2016; Mathews et al, 2017b) and deliberative techniques (Lienhoop and MacMillan, 2007, Aanesen et al., 2015). Advocates of conventional individual SP surveys argue that carefully crafted surveys provide consistent and reliable data (Haneman, 1994). However, their critics maintain that deliberative techniques, also known as deliberative monetary valuation (DMV), are superior because they allow respondents time to ask questions and thus construct or discover their preferences (Spash, 2008, Lo and Spash, 2013). While the constructed preferences hypothesis suggests that preferences, and thus mean willingness to pay (WTP) can shift, the discovered preferences hypothesis suggests that preferences are stable (Kingsley and Brown, 2010). Recent perspectives on discovered and constructed preferences have challenged the assumption that individuals retain well-formed, stable preferences (see e.g. Matthews et al., 2017a). This raises questions regarding the robustness of conventional individual SP surveys, and Völker and Lienhoop (2016) provide some evidence of preference adjustments after allowing survey participants time to discuss, but the changes were not unidirectional. They did not, however, demonstrate any increase in the consistency of choices made by the respondents, when consistency is defined in terms of the variance of the error term in a random utility model, which is inversely related to the scale parameter in a multinomial logit model (Brouwer et al., 2010, Kingsley and Brown., 2010). This research therefore aims to expand our understanding of how social interactions may influence public good valuation. This is important in order to

determine whether regular individual one shot choice experiment surveys that do not involve social interactions can provide results that emulate real world values.

We attend to these unresolved questions by testing the robustness of elicited individual preferences in the SP-setting across two elicitation situations separated by a group discussion. We also test whether such robustness varies according to knowledge levels or by how certain individuals are of their choices. As an empirical case we use results from a choice experiment (CE) to elicit a subject's willingness to pay (WTP) to protect cold-water coral (CWC), an unfamiliar public good in Ireland.

Our results indicate that stated preferences, and the consistency with which choices are made, change after deliberation, but only to a limited degree. The rest of the paper is organized as follows: Section 2 and 3 provide background on SP methods and the good to be values, section 4 presents method and data. In section 5 the results are given, and sections 6 and 7 discusses our findings and concludes.

2 Literature

A wide variety of DMV survey approaches have been deployed, including market stall (MacMillan et al, 2002, Lienhoop and MacMillan, 2007), valuation workshops (Hensher et al., 2011; Aanesen et al., 2015; Lienhoop and Volker, 2016), and repeated elicitation tasks (MacMillan et al, 2006). While valuation workshops are usually conducted in a single session over a few hours, the market stall approach includes two (or more) group sessions whereby respondents receive the survey at two different times, spaced some time apart, with time for discussion and reflection with family and friends and an opportunity to re-consider their choices.

Within the SP research there is a well-established "information effects" literature (Hoehn and Randall, 1987, Samples et al., 1986; Bergstrom et al. 1990) demonstrating that information presented in contingent markets results in more valid WTP estimates compared to when information is not provided. Economic theory suggests that individuals who are well informed about a good are less likely to change their preferences compared to less well-informed respondents, when provided with new information (Hoehn and Randall, 1987, Schläpfer and Fischhoff, 2012). While it is suggested that information presented in SP studies enhances the external validity of the estimated willingness to pay (WTP), and reduces respondent uncertainty and the divergence between reported WTP and true WTP, there are few empirical studies that demonstrate such effects (Hoehn and Randall, 1987; Paradiso and Trisorio, 2001). How information is provided may also matter, and the two main procedures for information provision are; i) to provide respondents with factual data about the good under consideration (Johnston et al., 2017), and ii) allowing the respondent to ask questions, and deliberate on the topic (Aanesen et al., 2015). While the former can be included in internet surveys, the latter demands in-person surveys. Needham et al. (2018) demonstrates that *ex ante* knowledge does matter for the value people place on non-market goods, in contrast to an exogenous increase in knowledge as part of the survey.

A number of studies have evaluated choice certainty[i] which is thought to be influenced by choice task learning, engagement, experience with the hypothetical good and group decision making (Brouwer et al., 2010; Hensher et al., 2011). However, with a few exceptions (LaRiviere et al., 2014), most empirical studies on respondent certainty tend to focus either on the effects of information provided by the SP analyst on consumer preferences (Shapansky et al., 2008; Ladenburg, 2013), or the implications of the information states – what a subject knows about

the good in question prior to being surveyed on WTP (Paradiso and Trisorio, 2001). LaRiviere et al., (2014) and Aanesen et al (2015)[ii] record the information state in their use of valuation workshops to elicit well-informed choices for the preservation of cold-water coral (CWC), but in these papers respondents do not exchange information. On the other hand, Lienhoop and Volker (2016) test the hypothesis that discussion amongst participants and time to reflect on information provided by the SP analyst stimulates preference discovery and leads to more consistent choices after deliberation. However, while self-reported choice certainty increased slightly after deliberation, this was not reflected in a statistically significant change in the scale parameter for choices made after deliberation relative to scale for choices made before.

3 Cold-water coral (CWC) – an unfamiliar public good

CWCs are found in all of the world's oceans and the Mediterranean Sea (Fosså and Buhl-Mortensen, 1999) with a global coverage that may exceed that of shallow water tropical coral reefs. Unlike tropical corals, CWCs do not need sunlight to survive, because they do not rely on the symbiotic relationship with photosynthetic algae (ibid), and so can live at depths below the photic zone. CWCs prey on zooplankton and are associated with strong currents which aid in the supply of food, egg and larval dispersal as well as removal of waste products. Reefs occur at depths of 39 – 3000 meters and a temperature range of 4° - 13°C. CWC reefs grow slowly at about 4-25mm per year and they are slow to recover from damage by human activity. Some of the larger reefs around the world are estimated to be more than 8000 years old (Freiwald et al, 2004). CWC reefs are impacted by bottom-trawling and other human sea-bed activities (Fossaa and Buhl-Mortensen, 1999), which is an argument for protection.

4 Methods and data

Sampling

Data was gathered using a stratified random sample of the general population with the aid of a professional survey company. Seven identical valuation workshops involving 20 individuals per workshop were conducted (excepting one with 19 respondents) in seven Irish municipalities and the subjects were representative regarding gender and age in the municipalities in which they were implemented. In total 139 individuals were used in the analysis. Respondents were informed that they would be offered a gift card amounting to €50 if they completed the survey. Table 1 indicates respondent characteristics.

Table 1 Socio-demographic variables for the sample and the Irish population

	Lowest value	*Highest value*	*Mean*	*National average*
Gender	0 (male)	1 (female)	0.489	0.494
Age	18 years	72 years	44.9 years	37.4
ENGO[a]	0 (not ENGO member)	1 (ENGO member)	0.02	
Education	1 (only obligatory)	4 (higher deg. Univ.)	3.45	1.85
Labor force participation	0	1 (in labor force)	64%	62.2
Working in the marine sector[b]	0 (other industries)	1 (the marine sector)	7.2%	1.5%
Household size (cont. var.)	1	8	3.34	2.75
Annual personal income	1 (below 7800 EUR)	9 (above 117k EUR)	3.63 (25,430)	36,508

Annual household income	1 (below 7800 EUR)	8 (above 117k EUR)	4.88 (43,991)	56,346

Source: Central Statistics Office (CSO), 2014, [a] ENGO = Environmental Non-Governmental Organization, [b] Vega and Hynes (2017)

Compared to the socio-demographics for the Irish population our respondents reside in larger households, are somewhat older, have lower income and are better educated. With respect to gender, labor force participation and household size, our sample is reasonably representative of the Irish population.

Survey

CWC has recently been identified across a significant number of sites in Irish waters and they represent important deep-sea marine biodiversity ecosystems. Specific threats include deep sea fishing and oil and gas extraction. Although protective measures for CWC reefs were enacted in 2006 by way of four Special Areas of Conservation (SACs) covering approximately 2,500 km^2 of protected seafloor, the Irish government is considering extending the number and size of protected areas. A choice experiment was used to gather information about respondents' preferences for protective measures.

The new hypothetical policy proposed in the survey is increased protection of CWC reefs in Irish waters. The survey instrument was originally developed in Norway. It was informed by five focus groups in Norway; three scientific groups with experts in marine biology, ecology, oceanography, resource economics and business/administration, and two public groups. Subsequently, this was adjusted to Irish conditions. The adjustments were based on three focus groups in Ireland (Galway), one among divers, one among sailors, and one involving

artists. Furthermore, the biological content of the survey was adjusted using the existing literature on CWC in Irish waters (Grehan et al, 2004; Guinan et al, 2009), with final verification by an expert (pers.comm. Anthony Grehan). The cost attribute was adjusted for differences in income and exchange rate between Ireland and Norway applying the PPP (purchasing power parity) index (https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm). CWC is little known by the public in both Ireland and Norway, both countries are traditional fishing nations with a recently developed petroleum industry, and both are part of a common western European culture, justifying the use of very similar survey instruments in the two countries.

Based on input from the groups and scientists, four attributes and their levels were chosen for the survey in Ireland. The *size* attribute refers to the total area protected, and takes three levels; the status quo (SQ) level, which was about 2500 km^2 in Ireland in 2014, 5000 km^2 and 10,000 km^2. The two increased protection areas were defined based on discussions with scientists, whereby the former area is the approximate identified CWC habitat in Irish waters, and the latter the hypothesized area based on habitat suitability modeling. There are a number of commercial offshore activities, fisheries and oil and gas exploration being the most important. Areas suggested for CWC protection may clash with areas where these activities are taking place, implying that industrial activities will be hampered if protection is extended. The attribute *commercial* distinguishes between whether areas eligible for CWC protection are either attractive fishing grounds, attractive areas for the oil and gas industry, for both, or for neither. The *habitat* attribute makes a distinction between areas considered very important as habitat for fish compared to areas of less importance. The *cost* attribute was described as a household federal tax contribution ring fenced as a trust fund exclusively for CWC protection in Ireland. With no increase in protection this additional tax will equal 0, while for alternatives

with increased CWC protection the tax will vary between 7 - 75 Euros per household per year. The attributes and the various levels they take are reported in Table A1 in the online supplementary material, together with an example of a choice card (Figure A1). The CE design included 12 choice cards. The original questionnaire was tested in a pilot survey, which provided the priors to inform the final design. We deployed a d-efficient design for a multinomial logit model, which has the property that it increases the precision of the parameter estimates and reduces the potential for parameter estimate misspecification (Scarpa and Rose, 2008). The original design for the Norwegian survey was updated after the pilot, and one more time throughout the study (Aanesen et al, 2015), and it is the last design of the Norwegian survey that was applied throughout all workshops in Ireland.

Each valuation workshop was conducted in two rounds as follows: It started with a power point presentation of CWC, lasting about 10 minutes and providing some basic information about the biology and distribution of CWC. Next, the participants were asked to fill in a CE survey, including 8 quiz questions on CWC and a stated certainty question for choices made on choice cards on a scale from 1 (not at all certain) to 10 (very certain). After the first elicitation exercise, respondents were asked to form groups of 5 individuals, to discuss CWC protection, in a systematic way for about 20 minutes. The groups were formed randomly by asking the five respondents sitting next to each other to form a group and all groups sat separately in a large room. The discussion procedure was systemized in the following way: First, each group were asked to form a circle and nominate a secretary. Next, going around the circle, each participant was asked to mention one association they had with CWC, and this continued until none of the subjects in the group had any more associations. The secretary noted each new association on a flip board, visible to the whole group. When finished, the participants were allowed to either ask a question if they did not understand the association or

if they felt strongly that the association was not relevant to CWC. After this deliberation, the participants were asked to individually fill in the choice cards and a stated certainty question from the original survey a second time. They were not able to see their first round answers in the second elicitation round. Each valuation workshop lasted about 2 hours, and the participants received the € 50 gift card when leaving the venue room.

Econometric model

Random utility theory suggests that the utility a person receives from a good can be divided into a determined part, which can be observed by the researcher, and a random part (Train, 2009). The utility to person *n* of choosing alternative *j* in choice situation *t* is thus given by;

$$U_{njt} = \boldsymbol{b} \cdot \boldsymbol{X}_{jt} + e_{njt} \quad [1]$$

where $\mathbf{X_{njt}}$ is a vector of attributes specifying the good, $\mathbf{b}$ is a vector of estimated coefficients for the attributes, and e_{njt} is an i.i.d. extreme value (usually Gumbel) error term.

The variance of the error term is given by $\sigma_n(\frac{\pi^2}{6})$, where σ is a scale parameter, possibly individual specific. The scale parameter is inversely correlated with the variance of the error term, and can be interpreted as how consistent respondents are when making their choices (Czajkowski et al., 2014a). Dividing (1) through by σ we get;

$$U^*_{njt} = \frac{\boldsymbol{b}}{\sigma} \cdot \boldsymbol{X_{jt}} + \varepsilon_{njt} \quad [2]$$

where $\varepsilon_{njt} = \frac{e_{njt}}{\sigma}$, and this new error term has constant variance given by $\frac{\pi^2}{6}$.

Equation (2) assumes constant preferences and scale across individuals, and the two are confounded and cannot be estimated separately. We modify the model in two respects. First; we allow preferences to be heterogeneous across individuals, assuming them to be randomly and continually allocated according to a well-specified distribution. Throughout the paper we use the normal distribution. This implies that the vector of attribute coefficients, **b**, is respondent dependent, i.e. $\mathbf{b_n}$. Hence,

$$\mathbf{b_n} = \mathbf{b} + \boldsymbol{\mu}_\mathbf{n} \cdot \mathbf{C} \qquad [3]$$

where **b** is the mean estimated coefficient for the specified attribute and $\boldsymbol{\mu}_\mathbf{n}$ is the vector of person *n* specific deviations from the mean, and **C** is draws from a specified distribution. Second, we allow scale to differ according to whether a choice was made before or after the group discussion. As it is not possible to estimate scale parameters separately for choices made before and after the discussion, we normalise scale for choices made before the group discussion, denoted σ_B, and set it equal to 1. Then we can estimate σ_A , which is the scale for choices made after the group discussion relative to scale for choices made before. Note that this implies that we assume the scale to be equal for all choices made before deliberation, and for all choices made after deliberation, but not across the two elicitation situations.

When the error terms are distributed extreme value and independent over *n*, *j* and *t*, the difference between two extreme value distributed error terms is logistically distributed, such that the choice probability when preferences are individual is given by;

$$P(j|\boldsymbol{X_{nt}}) = \frac{1}{D}\sum_{d=1}^{D} \frac{exp^{(\frac{\boldsymbol{bn}}{\sigma A})*\boldsymbol{Xjt})}}{\sum_k exp^{(\frac{\boldsymbol{bn}}{\sigma A})*\boldsymbol{Xkt})}} \qquad [4]$$

where *d* is an index for the numbers of draws from a specified (the normal) distribution. The simulated probability of observing person *n* choosing a sequence of *t* choices is given by;

$$P(j|\boldsymbol{X_{nt}}) = \frac{1}{D}\sum_{d=1}^{D}\prod_{t=1}^{T}\prod_{j}\left[\frac{exp^{(\frac{\boldsymbol{bn}}{\sigma A})*\boldsymbol{Xjt})}}{\sum_k exp^{(\frac{\boldsymbol{bn}}{\sigma A})*\boldsymbol{Xkt})}}\right]^{ynjt} \quad [5]$$

where y_{njt} is a dummy variable taking the value 1 if alternative *j* is chosen and 0 otherwise. Summarising over all respondents and alternatives, and taking the log of (5) yields the log likelihood function.

$$LL(\boldsymbol{b}) = \sum_{n=1}^{N}\sum_{j} y_{njt} ln P_{njt} \quad [6]$$

(6) is maximised w.r.t. the **X** vector in order to find estimates of the vector of attribute preferences, $\mathbf{b_n}$. The parameter vector, $\mathbf{b_n}$, and the relative scale, σ_A, are estimated by simulation, using the maximum likelihood procedure, as the derivative of the expression in 6 w.r.t. $\mathbf{X_{jt}}$ does not have a closed form solution (Train, 2009).

To test for statistically significant deviations in stated preferences based on choices made before and after the group discussion, we use interaction terms for the attribute vector in (2);

$$U^{*}_{njt} = \frac{\boldsymbol{bn}}{\sigma A}\cdot \boldsymbol{X_{njt}} + \frac{\boldsymbol{gn}}{\sigma A}\cdot \boldsymbol{X_{njt}}\cdot D + \varepsilon^{T}_{njt} \quad [7]$$

where *D* is a dummy taking the value 1 for choices made after the group discussion and 0 for other choices, $\mathbf{g_n}$ and $\mathbf{b_n}$ are vectors of parameters to be estimated, σ_A is the relative scale parameter for choices made after the discussion, and ε^{T}_{njt} is an error term. Estimated

coefficients of $\mathbf{g_n}$ significantly different from 0 indicate that stated preferences after the discussion are significantly different from stated preferences before the discussion. This procedure corresponds to a t-test for parameter equality. Estimated σ_{A} significantly different from 1 indicates that the consistency with which choices are made differs between the two elicitation situations.

Finally, given the interest in establishing estimates of WTP for the non-monetary attributes, it is convenient to introduce the following modification, which is equivalent to estimating the parameters in WTP space (Train and Weeks, 2005)

$$U_{njt}^{**} = \frac{bc}{\sigma A}(p_{njt} + \boldsymbol{\beta_n W_{njt}}) + \varepsilon_{njt}^{WTP} \qquad [8]$$

where $\mathbf{W_{njt}}$ is a vector of non-cost attributes. Note that under this specification the vector of parameters $\boldsymbol{\beta}_{\mathbf{n}} = \frac{\mathbf{bn}}{\mathrm{bc}}$ is now (1) scale-free and (2) can be directly interpreted as a vector of implicit prices for the non-cost attributes. p_{njt} is the cost attribute.

5 Results

The 139 respondents made altogether 1668 choices before the group discussion. The same number of choices were made after the discussion. After the discussion, 120 respondents (86%) changed at least one of their choices, while 19 (14%) kept all choices unchanged. Out of the 1668 choices before the discussion, 417 (25%) were altered after the discussion. Although many respondents changed their choices, the distribution of choices across the three alternative scenarios on the choice cards remained largely unaffected. Before deliberation the status quo accounted for 41.1% of choices, and after deliberation this share represented 42.1%. A similar pattern was apparent for alternative 1 which represented 30% of choices

before the discussion and 27.5% after. Table A3 in the online supplementary material shows that in most of the groups, members make the same number of similar choices before and after deliberation, and the deliberation does not seem to impact choices. The second largest category consists of groups that make more similar choices after deliberation compared to before, indicating that stated preferences become more homogenous. The smallest category are those groups where members' choices diverge after the deliberation, indicating that preferences have become more heterogeneous.

The crucial question is whether these changes imply that different preferences were stated before and after the group discussion, and whether the consistency of the choices changed after the discussion. Table 2 yields results from the mixed MNL model for each elicitation situation separately and for the pooled model assuming unequal scale in the two elicitation situations. The models are estimated in R, using the apollo package (Hess and Palma, 2019) and its forerunner cmcRcode. Normal distribution is used for all non-cost attributes whereas a lognormal distribution is used for the cost attribute. Starting on 500 and gradually increasing the number of draws, we ended up using 1000 Halton draws as higher number of draws (or lower) did not improve the likelihood value substantially. Results for corresponding MNL models in preference and WTP-space can be found in the supplementary material (Tables A5-A7). Note that two of the original attributes have been split into two, returning altogether 6 attributes.[iii]

Table 2 Estimated attribute coefficients in preference space (t-values in brackets) based on individual choices before and after deliberation, and for the pooled model when assuming endogenous relative scale for choices made after deliberation

Attribute	*Mixed MNL model*	*Mixed MNL model*	*Pooled model, common attribute coefficients an*

	for choices made Before deliberation	*for choices made After deliberation*	*Endogenous Relative scale for choices after deliberation*
Small (mean)	0.27 (0.78)	0.32 (0.86)	0.38 (1.48)
Large (mean)	0.12 (0.41)	0.16 (0.39)	0.27 (1.26)
Oil (mean)	-0.80 (-3.35)***	-0.52 (-1.46)	-0.62 (-1.73)*
Fish (mean)	-0.12 (-0.57)	-0.25 (-1.06)	-0.19 (-1.24)
Habitat (mean)	1.93 (7.31)***	2.31 (8.30)***	1.96 (8.73)***
Cost (mean)	0.03 (4.08)***	0.02 (3.27)***	0.03 (4.87)***
Small (st.dev.)	-1.28 (-3.19)***	1.36 (4.10)***	-1.30 (-2.89)***
Large (st.dev.)	-1.32 (-3.38)***	-1.96 (-6.11)***	1.31 (8.52)***
Oil (st.dev.)	2.05 (7.29)***	2.41 (7.45)***	-1.73 (-8.60)***
Fish (st.dev.)	1.43 (6.82)***	1.70 (5.68)***	-1.32 (-8.36)***
Habitat (st.dev.)	2.00 (7.98)***	3.25 (7.80)***	2.20 (9.43)***
Cost (std.dev)	-0.21 (-5.68)***	-0.27 (-6.95)***	-0.22 (-8.94)***
Relative scale			1.25 (1.70)*
Log-Likelihood value	-1232.7	-1104.8	-2342.0
N, k	1668, 12	1668, 12	3336, 13
Adj. R squared	0.31	0.38	0.35

***, **, * means significant at 1%, 5% and 10% respectively

Independent of elicitation situation, Table 2 shows that marginal preference sensitivities are heterogeneous across the respondents. For choices made before deliberation, preferences for the oil and habitat attributes, in addition to cost, are significantly different from zero. The negative sign of the oil parameter implies that people are not willing to protect CWC if this hampers oil and gas activities. The positive sign of the fish parameter indicates that people are

more likely to protect CWC if the reefs are important fishing grounds, but the estimate of this attribute is not significant. The two size attributes capture non-use values and measure preferences for protecting CWC-reefs *per se*, but neither of them is significant. The positive sign of the habitat attribute means that if a coral reef is supposed to serve as habitat for fish it is more likely that a respondent is willing to pay to protect the coral area. Habitat is the single most important attribute to explain why respondents choose to protect coral reefs. The cost parameter is significant, indicating that on average respondents did take the cost of protecting further coral reefs into consideration when making their choices. For choices made after deliberation, only preferences for the habitat attribute, in addition to cost, is significantly different from zero.

Table 2 also shows that relative scale for choices made after deliberation is significantly higher than 1, which is the normalized scale for choices made before deliberation. As scale is inversely related to the variance of the error term of the model, a higher scale indicates a lower error variance, which in turn can be interpreted as if more of the variation in the data is explained by the systematic part of the model (Swait and Louviere, 1993). As preference sensitivity is confounded with scale, a difference in scale across the two elicitation situations implies that we cannot compare estimated attribute coefficients for the two situations directly, but must either correct for scale or use scale-free estimates for marginal preference sensitivities for the attributes.

A preliminary for model equality, using the standard likelihood ratio statistics, which are chi-squared distributed, shows that we can reject the null of equal scale across the two elicitation situations (test-stat=12, 90% critical value=2.70), but not the null of equal mean parameters (test-stat=9, 90% critical value=19.8).

Estimating the model in WTP-space will produce scale free attribute coefficients (Rose and Masiero, 2010). Table 3 report results from the model in WTP-space, using the pooled dataset and separate coefficients for choices made before and after deliberation. Distributions and draws are the same as for the model in preference-space. Table 3 yields mean WTP and 90% confidence interval and standard deviation for the WTP for each of the attributes in the two elicitation situations. A student's t-test, comparing two means by the use of standard errors by which the two sample means are separated, is used to test for mean coefficient similarity across the two elicitation situations. The test statistic shows that stated preferences do not change significantly for any of the attributes. For the habitat attribute, however, there is a positive change from the first to the second round, but only significant at 12% level (t-value= 1.59).

Table 3 Mean willingness to pay (WTP) estimates, 90% confidence intervals, standard deviation, Euro, and test-statistic for difference of mean coefficients

	WTP in Euro, Before	*90% CI/ t-value*	*WTP in Euro, After*	*90% CI/ t-value*	*Test-statistic*[iv]
Mean					
Small	-22.2*	(-40, -3.5)	-15.1	(-35, 5)	-0.42
Large	-25.9***	(-43, -9)	-13.9	(-30, 2.5)	-0.83
Oil	3.7	(-7.5, 15)	-3.3	(-15, 8)	0.70
Fish	7.6	(-1, 16)	1.00	(-5, 7)	1.00
Habitat	34.7***	(24, 45)	54.8***	(37, 72)	1.59
Cost[a]	-0.05***	(0.01)	-0.03***	(0.01)	1.41

Std.deviation				
Small	37.2***	(5.24)	28.98***	(3.97)
Large	36.7 ***	(5.69)	47.72***	(5.02)
Oil	36.2 ***	(7.09)	40.33***	(4.54)
Fish	35.3 ***	(6.76)	26.58**	(4.02)
Habitat	52.1 ***	(8.18)	46.80***	(5.41)
Cost[a]	0.21***	(0.03)	0.27***	(0.05)
LL-value	-2437.0			
Adj. R2	0.32			
N, k	3336, 24			

[a] In preference space, std.error in brackets, ***, **, * means significant at 1%, 5% and 10% level respectively

To further elaborate potential changes in stated preferences across the two elicitation situations, we ran the model in preference-space on the pooled dataset including a treatment dummy. This dummy took the value 1 for choices made after deliberation and 0 otherwise, and was interacted with each of the attributes. We ran two versions of this model, one assuming a homogenous treatment dummy and one assuming a heterogeneous treatment dummy. The results are given in Table 4.

Table 4 Mean attribute coefficients for choices made before deliberation, mean treatment coefficients, and relative scale (t-values in brackets) for treatment coefficients for model with fixed and random treatment coefficients

	Model with fixed treatment coefficients	**Model with random treatment coefficients**

Attribute	Main attribute coefficient for choices made before deliberation	Treatment coefficient for choices made after deliberation	Main attribute coefficient for choices made before deliberation	Treatment coefficient for choices made after deliberation
Cost_mean	0.03 (5.50)***	-0.003 (-0.04)	-0.187 (-6.64)***	-0.167 (-1.37)
Small_mean	0.36 (1.12)	-0.11 (-0.26)	0.288 (0.88)	0.11 (0.18)
Large_mean	0.15 (0.53)	0.14 (0.34)	0.128 (0.45)	-0.086 (-0.14)
Oil_mean	-0.66 (-2.81)***	0.17 (0.42)	-0.597 (-2.76)***	-0.347 (-0.84)
Fish_mean	-0.13 (-0.66)	0.004 (0.01)	-0.21 (-1.02)	-0.012 (-0.03)
Habitat_mean	2.13 (6.20)***	-0.1 (-0.24)	2.014 (8.09)***	1.774 (1.88)*
Cost_std.dev	-0.23 (-8.33)***		0.22 (7.22)***	0.263 (1.63)
Small_std.dev.	1.19 (4.62)***		-1.55 (-2.49)***	-0.534 (-0.65)
Large_std.dev.	-1.53 (-4.50)***		1.367 (5.48)***	-2.339 (-3.04)***
Oil_std.dev.	1.84 (8.73)***		-1.763 (-6.78)***	-3.237 (-4.04)***
Fish_std.dev.	1.37 (7.64)***		1.485 (5.35)***	-1.675 (-2.32)***
Habitat_std.dev	-2.00 (-9.71)***		1.909 (7.41)***	-4.293 (-3.38)***
Scale for treatment interactions	1.25 (1.57)		0.6811 (-1.61)	
LL-value	-2342		-2339	
N, k	3336, 19		3336, 25	
Adj. R-squared	0.35		0.35	

***, **, * means significant at 1%, 5% and 10% respectively

Table 4 confirms that preferences are heterogeneous. In addition, changes in stated preferences after the discussion are heterogeneous. Allowing the treatment to vary across respondents, yields a significant treatment coefficient for the habitat attribute, which means

that marginal preference sensitivity for the habitat attribute is higher after deliberation than before. This result was also present in the WTP-space model, although not significant. Standard deviations have decreased significantly for 4 attributes, indicating reduced heterogeneity in stated preferences after the discussion. Note that relative scale for treatment interactions are substantially different in the two models.

Having indicated a slight change in stated preferences after deliberation for the sample, we turn to four sub-groups to test the robustness of this result. As part of the survey, respondents were asked the following question:

> Thinking about the choice you made on the last choice card (Choice card 12 on page 19). On a scale from 1-10, please indicate how certain you were of the answer to this question. Please circle the number that best represents your answer if 1 = not at all certain and 10 = very certain.

We define this as self-reported certainty. They were asked this in both elicitation rounds directly after having made their choices. Furthermore, we included the quiz score, i.e. how many correct answers out of 8 did the respondent receive on the quiz that was implemented prior to the first elicitation situation. On average the certainty-score was 7.52 (95% confidence interval: (7.19, 7.84)) for choices made before the group discussion, and 7.57 (95% confidence interval: (7.26, 7.88)) for choices made after the group discussion. Defining those with stated certainty below the mean, i.e. those who answered 1-7 on the certainty-question, as “uncertain” and those with stated certainty above this mean, i.e. those who answered 8-10 to the certainty question, as “certain”, we divided all respondents with an answer to the question into the two categories. We did this based on their pre-deliberation statement. Of the 28 discussion groups, 24 groups had members that were both above and

below average certainty, while 3 groups contained only members with stated certainty above average and 1 group contained only members with below. We take the quiz score as a proxy for the respondents' knowledge level of the issue under consideration at the time of the survey. The mean number of correct answers out of 8 was 6.24. Respondents with scores higher than the mean, i.e. 7 or 8 correct answers, are henceforth denoted high-score, whereas the remaining are defined as low-score. All 28 discussion groups contained a mix of high-score and low-score members.

Continuing with the four subsets, Table A2 in the online supplementary material shows that the subsets do differ when it comes to socio-demographic characteristics. Respondents certain of their choices and high-score respondents are on average younger, better educated, have higher personal and household income, and are more likely to be in the labor force compared to respondents more uncertain of their choices and low score respondents. They also belong to slightly larger households, and there is a female majority in both these sub-groups. Respondents less certain of their responses, together with high-score respondents, are more likely to be a member of an Environmental Non-Governmental Organization (ENGO). Respondents certain of their choices and low-score respondents are more likely to work in the marine sector (fisheries and oil and gas activities).

As the pooled model indicated heterogeneous treatment effect we run the heterogeneous treatment model on the four subsets. The results are given in Tables 5 and 6.

Table 5 Mean attribute coefficients for choices made before deliberation, mean treatment coefficients, and relative scale for treatment coefficients (t-values in brackets) for respondents with above and below average self-reported certainty of choices

	Model for respondents with above average self-reported certainty of answers		**Model for respondents with below average self-reported certainty of answers**	
Attribute	Main attribute coefficients for choices made before deliberation	Random treatment coefficient for choices made after deliberation	Main attribute coefficients for choices made before deliberation	Random treatment coefficient for choices made after deliberation
Cost_mean	-0.21 (5.50)***	-0.003 (3.42)***	-0.21 (-7.94)***	-0.00 (0.26)
Small_mean	0.14 (0.29)	-1.58 (-1.52)	0.68 (1.65)*	0.39 (0.69)
Large_mean	0.10 (0.24)	-0.32 (-0.34)	0.44 (1.18)	-0.02 (-0.03)
Oil_mean	-0.17 (-0.42)	-0.55 (-0.52)	-0.98 (-3.30)***	0.32 (0.81)
Fish_mean	-0.12 (-0.38)	0.34 (0.47)	-0.22 (-0.95)	-0.16 (-0.42)
Habitat_mean	2.75 (7.36)***	3.84 (3.17)***	1.21 (4.73)***	0.09 (0.26)
Cost_std.dev	-0.26 (-4.26)***	0.31 (4.45)***	0.25 (8.38)***	-0.03 (-2.66)***
Small_std.dev.	1.58 (5.38)***	2.95 (4.33)***	-0.98 (-4.31)***	-1.61 (-4.64)***
Large_std.dev.	1.27 (3.84)***	6.10 (3.85)***	1.10 (5.33)***	1.42 (4.27)***
Oil_std.dev.	-2.49 (-6.27)***	-5.90 (-3.28)***	-1.53 (-6.74)***	0.15 (0.57)
Fish_std.dev.	1.46 (5.73)***	2.65 (-2.13)***	-1.21 (-4.44)***	-1.53 (-4.12)***
Habitat_std.dev.	3.15 (4.78)***	-5.20 (-4.22)***	0.96 (5.50)***	1.34 (5.77)***
Scale for treatment interactions		0.62 (2.99)***		1.39 (1.46)
LL-value	-1255.6		-1045.8	
N, k	1944, 25		1392, 25	
Adj. R-squared	0.40		0.28	

***, **, * means significant at 1%, 5% and 10% respectively

Respondents in the "certain" sub-group exhibit the larger changes as revealed by Table 5. For two of the attributes, cost and habitat, the treatment coefficient is significantly different from

zero, indicating that stated preferences changed across the two elicitation situations. Furthermore, standard deviation has changed for all attributes, indicating more homogenous preferences for the oil and habitat attributes, and less homogenous preferences for the other attributes. Finally, their change in choices after the discussion is less consistent compared to the choices made before, as shown by the relative scale being significantly smaller than one. Subjects more uncertain of their choices do not change stated preferences for any attribute. However, they have become more homogenous with regard to stated preferences for the large, oil and habitat attributes, and less homogenous for the other attributes.

Table 6 *Mean attribute coefficients for choices made before deliberation, mean treatment coefficients, and relative scale (t-values in brackets) for high-score and low-score respondents*

	Model for high-score respondents		**Model for low-score respondents**	
Attribute	Main attribute coefficients for choices made before deliberation	Treatment coefficient for choices made after deliberation	Main attribute coefficients for choices made before deliberation	Treatment coefficient for choices made after deliberation
Cost_mean	-0.19 (-7.03)***	0.09 (5.64)***	0.004 (4.74)***	0.00 (0.14)
Small_mean	0.86 (-0.33)	-0.24 (-1.37)	-0.28 (0.57)	0.57 (0.97)
Large_mean	0.34 (0.93)	-0.04 (-0.06)	0.04 (0.09)	-0.37 (0.77)
Oil_mean	-0.71 (-1.69)*	0.31 (0.47)	-0.57 (-2.21)***	0.14 (0.45)
Fish_mean	-0.33 (-0.91)	-0.22 (-0.46)	0.12 (0.46)	-0.20 (-0.57)
Habitat_mean	2.83 (7.81)***	0.72 (0.51)	1.44 (4.97)***	-0.14 (-0.31)
Cost_std.dev		-0.17 (-2.84)***		-0.04 (-2.74)***
Small_std.dev.	0.96 (1.71)*	-1.57 (-1.25)	0.60 (3.16)***	-0.10 (-0.34)
Large_std.dev.	1.04 (-4.22)***	-1.69 (-1.75)*	-1.23 (-5.60)***	1.07 (5.27)***

Oil_std.dev.	-2.47 (-5.15)***	-2.28 (-2.24)***	-1.36 (-5.18)***	-0.36 (-1.60)
Fish_std.dev.	1.68 (6.41)***	0.69 (2.41)***	-1.06 (-4.69)***	0.54 (1.95)*
Habitat_std.dev.	2.71 (5.51)***	-0.07 (-0.16)	1.71 (6.52)***	1.47 (4.44)***
Scale for treatment interactions	0.95 (0.21)		1.52 (1.77)*	
LL-value	-1116.2		-1182.9	
N, k	1608, 25		1728, 25	
Adj. R-squared	0.35		0.34	

***, **, * means significant at 1%, 5% and 10% respectively

Table 6 shows that high-score respondents exhibit the larger changes in stated preferences. The positive treatment coefficient for the cost attribute, rendering the absolute value of this coefficient smaller, indicates that high-score respondents put less emphasis on this attribute after deliberation. Low-score respondents do not change stated preferences for any of the attributes. However, the relative scale is significantly above 1, indicating that changes in stated preferences across the two rounds are more consistent compared to pre-deliberation stated preferences. The interaction term standard deviations are mainly negative for both groups, indicating more homogenous stated preferences after the group discussion. The exception is the fish attribute, for which preferences have become more heterogeneous after deliberation.

Table 7 sums up the results from the treatment interaction model in table 6.

Table 7 Summary of mean treatment coefficient and relative scale parameter for treatment coefficients compared to scale for choices made before deliberation

	Pooled dataset	*High-score subset*	*Low-score subset*	*Certain subset*	*Uncertain subset*
Relative scale (t-value)	0.68 (-1.61)	0.95 (-0.21)	1.52 (1.77)	0.62 (-2.99)	1.39 (1.46)
Small_mean	No change	No change	No change	No change	No change
Large_mean	No change	No change	No change	No change	No change
Oil_mean	No change	No change	No change	No change	No change
Fish_mean	No change	No change	No change	No change	No change
Habitat_mean	Positive change	No change	No change	Positive change	No change
Cost_mean	No change	Positive change	No change	Negative change	No change
Small_std.dev.	No change	No change	No change	Positive change	Negative change
Large_std.dev	Negative change	Negative change	Positive change	Positive change	Positive change
Oil_std.dev	Negative change	Negative change	No change	Negative change	No change
Fish_std.dev.	Negative change	Positive change	Positive change	Negative change	Negative change
Habitat_std.dev	Negative change	No change	Positive change	Negative change	Positive change
Cost_std.dev	Positive change	Negative change	Negative change	Positive change	Negative change

6 Discussion

Stated preferences for increased CWC protection among the Irish population are heterogeneous. This is the case in both the pre- and post-deliberation choices, as well as for changes in choices across the two situations. Focusing on models allowing for preference heterogeneity we demonstrate that stated preferences after deliberation differ from those stated before, rejecting the hypothesis of parameter equality across the two elicitation rounds.

However, stated preferences were altered during the second elicitation situation with the smallest possible margin as they changed for only one out of six attributes. The attribute for which participants changed stated preferences is habitat. This is the most important attribute in explaining choices to protect CWC in both elicitation situations, and in all four sub-groups. After deliberation workshop participants on average stated stronger preferences for the habitat attribute compared to before. This result is consistent with Völker and Lienhoop (2016) showing that after deliberation participants in homogenous groups were willing to pay more for all ecosystem services, while participants in heterogeneous groups were willing to pay more for two out of three ecosystem services. Furthermore, they found that participants in heterogeneous groups put more emphasis on the cost attribute after deliberation, but they did not find this for participants in homogenous groups. In our survey, on average, stated preferences for the cost attribute did not change after deliberation, but splitting the sample on sub-groups shows that while high score respondents significantly decreased (the absolute value of) stated preferences for the cost attribute, respondents certain of their choices increased it. On the one hand, a stable cost attribute parameter is important given this attribute's role in the calculation of willingness-to pay (WTP) estimates. An unstable cost attribute parameter will jeopardize WTP calculations. On the other hand, one may argue that an increase in (the absolute value of) the cost attribute parameter, like for "certain" respondents in our survey, indicates that respondents take this attribute more seriously after deliberation, which is positive for the reliance of SP-survey results (McMillan et al., 2002).

Allowing for a separate scale parameter in the pooled model shows that relative scale for choices made after deliberation is significantly higher than one. This suggests choices are made with higher consistency after deliberation, in the sense that a larger part of the variation in choices is now captured by the systematic part of the model and less by the random component.

The result is robust across models assuming common or separate attribute coefficients for the two elicitation situations.[v] Hence, we can refute the hypothesis of equal consistency of choices in the two elicitation situations. Neither Lienhoop et al. (2016) nor Matthews et al. (2017a) were able to demonstrate this effect, although their studies also indicated somewhat more consistent choices after social interaction and time to reflect. Allowing for heterogeneity in treatment effects, the interaction model yields a relative scale for treatment coefficients below one, indicating less consistent changes, but the result is not significant. One may argue that such a result is reasonable, as the way people change a previously stated preference usually is more random compared to the way they state (the same) preference in the first place.

A few studies have explored the relationship between self-reported certainty of choices and statistical consistency of choices. For example, Brouwer et al. (2010) demonstrates that although respondents tend to self-report higher certainty for choices made later in the sequence (of altogether 5 choices), the actual certainty, measured as the relative scale parameter, did not change significantly. The same is true for Lienhoop and Völker (2016), who report a tendency that self-reported certainty of choices increases after deliberation, though without it corresponding with a higher scale parameter. Matthews et al (2017b), on the other hand, find a higher scale parameter when using a video-enhanced approach compared to a traditional survey, but this does not correspond with higher self-reported certainty. Given our findings, these highly ambiguous results give reason to question the use of self-reported statements of certainty of choices as signals of informed or discovered preferences in CEs. On the other hand, it could be argued in our survey that for the subjects involved the good under consideration has been changed from the first to the second elicitation situation, due to the discussion involving the CWC associations. For this reason, we acknowledge that we cannot prove that self-reported certainty is not a reliable measure of informed or discovered preferences.

Turning to our treatment design, it can be claimed that in real life social interaction is restricted to a person's social network, which will confine the information input achieved by social interaction. This means that a more proper treatment would be to assure homogenous groups, as in the study by Lienhoop and Völker (2016). However, we make the point that social interaction is not necessarily restricted to a person's social network, but may also include colleagues at work, which most people don't choose, people in the same recreational activity as themselves, which people do choose but not necessarily according to their political or moral viewpoints, and neighbors, which may be more or less random. The presence of social interaction with this broader set of people makes our treatment's resemblance with reality more credible, and our groups are probably more like the heterogeneous groups reported in the study of Lienhoop and Völker (2016). Nonetheless, as Lienhoop and Völker (2016) demonstrate significantly different effects of group interactions on choices in a CE-survey for heterogeneous and homogeneous groups, we do recognize the importance of controlling for group type in future studies involving group interaction in SP-surveys.

We cannot establish a causal link between differences in stated preferences before and after deliberation and the group discussion. The reason is that the observed changes may be due to two reasons; 1) institutional learning, and 2) value learning. While the former implies that respondents gain experience with the choice tasks through practice and repetitions, the latter indicates that people discover their own preferences (Plott 1996). Although it is tempting to ascribe changes in stated preferences to the group discussion, we have to acknowledge that gaining experience in answering hypothetical questions provides a better understanding of the survey instrument which may also explain the changes in stated preferences. Our data does not provide sufficient information to disentangle these various possible explanations. Brouwer et

al. (2010), letting participants rate how certain they were of their choice after each of five consecutive choices in a CE-survey, without any social interaction between the choices, were not able to demonstrate a significantly higher scale parameter for choices made later in the sequence. The participants did, however, state higher average self-reported certainty for choices later in the sequence. Hence, Brouwer et al. (2010) lends little support to the institutional learning interpretation, which in turn can be seen as supporting an assumption that the identified changes in our study are due to the group discussion.

Another possible explanation for changes in stated preferences, or indeed a lack of such change, could be the fact that in a treatment like ours some people will try to answer consistently, by making similar choices in the two elicitation situations (Day et al., 2012). While this may be an argument in simpler and more transparent surveys, in our survey it is arguable whether the respondents remembered all of their 12 choices from the first round when making their second-round choices. Those following this strategy, 19 out of 139 respondents, were mainly respondents who persistently chose one and the same alternative on all choice cards in the first round.

Observing that most participants did not succeed in following a strategy of making similar choices in the two elicitation situations, a worthy question is why respondents adjust their choices? Our findings from the public focus groups during the early stages of the study show that virtually none of the participants had heard of CWC and did not realize they exist in Irish waters. Many were enthusiastic about supporting potential measures for the conservation of this species despite the associated costs of protection. This may also have been the case for the workshop participants in the first elicitation situation. In the second round, the "certain" sub-group on average increased the (absolute value) of the parameter for the cost attribute,

indicating that they had become more aware of costs of protection. Furthermore, a topic that featured in the majority of the groups was the environmental aspects of CWC and its role as an important habitat for fish. Given that this attribute got even more important after deliberation may be explained by the impact of the discussion.

There is an ongoing debate, especially in the DMV literature on whether preferences are constructed or discovered (Plott 1996), the main difference connected to the question of the assumed stability of estimated preferences. While it is suggested that under constructed preferences mean WTP is labile and can shift, under discovered preferences the mean WTP is more stable and less inclined to change. Although our results show a change in stated preferences for a unfamiliar public good in two subsequent elicitation situations, divided by a group discussion, the change comes with the smallest possible margin, as mean WTP changes for only one out of six attributes, and the change is significant only at 10% level. Hence, we are inclined to interpret our results as support for the discovered preferences hypothesis. Schaafsma et al. (2014) reveals a similar conclusion when testing temporal stability of WTP estimates in a CE-survey. Repeating an identical CE-survey with the same sample a year apart, they demonstrate that the parameter estimates are not temporally consistent, i.e. not equal across the two elicitation situations. However, they conclude that the WTP estimates for the attributes that are significant are mostly robust with respect to transfers over time.

A novel aspect of this survey is that we are able to distinguish between sub-groups based on both their knowledge of the good and stated certainty of their choices. Here, some of our results are unexpected. For example; people who state above average certainty for choices made in the first elicitation situation tend to change their stated preferences more than people less certain of their initial choices, and high-score respondents change stated preferences more than do low-

score respondents. Possible explanations for these results are that a high score may signal that respondents are taking a greater interest in the issue, which in turn lead to a higher involvement in the group discussion making these respondents more open to changes of their original choices due to arguments generated in the discussion. Low score and "uncertain" respondents, on the other hand, can be taken as signals of lower familiarity with the survey instrument and the good under consideration. These respondents may to a larger degree try to keep to their original choices, which is possible if, f ex., they have persistently chosen one and the same alternative on all choice cards in the first elicitation round. Such behavior can also explain that low-score respondents made changes in stated preferences after deliberation with higher consistency than their original choices.

A comment must be made regarding the unequal distribution of women and men in the sub-groups. Women are over-represented in two sub-groups; respondents certain of their initial choices and high-score respondents. Running the pooled model on female and male respondents separately shows that there is a gender bias when it comes to consistency in choices made after the group discussion, but not regarding whether respondents change stated preferences after the discussion. The change in stated preferences after deliberation is less consistent for women compared to for men, but this difference is not statistically significant. Women also score somewhat higher on the quiz, but the difference is not significant. Regarding stated preferences, while women change stated preferences for the cost and the habitat attribute, men change stated preferences for the cost and the fish attribute. While these differences are worth noticing, we cannot see that they change any of the arguments above.

7 Conclusions

On average people do state different preferences after a discussion compared to before, and they tend to be more consistent in making their choices after the discussion. This said, the conclusions are made with the smallest possible margin, as stated preferences only changed for one out of six attributes, and results were significant only at 10%-level. Such modest changes support the assumption of discovered preferences rather than constructed preferences. The results are in line with results from Schaafsma et al. (2014) on temporal stability demonstrating that although parameter estimates were not temporally consistent, mean WTP estimates were mostly robust to transfers over time.

The valuation workshop format allowed us to inform respondents about the good under consideration as well as the format of the survey instrument. It also enabled us to distinguish between various sub-groups according to their knowledge of the good and familiarity with the survey instrument. Thus, we were able to analyze whether groups of individuals behaved differently after the discussion, and we did find differences regarding preference adjustments between the pooled data and the sub groups. High score respondents changed their stated preferences more after the discussion compared to those with lower score, and participants certain of their initial choices changed stated preferences more after the discussion compared to respondents less certain of their choices. Although we did not control for attitudinal variables in the discussion groups, our findings are consistent with Volker and Lienhoop (2016) who reveal that individuals in homogenous and heterogenous groups react quite differently to a discussion.

So, what do these results imply for choice experiments as a whole? Our results indicate that on average stated preferences are relatively stable with respect to deliberation in groups, and this holds even for preferences for unfamiliar public goods like CWC. The latter is important as previous literature indicates that information and deliberation may have larger effects in the

elicitation of unfamiliar goods (Fischoff et al. 2013). Hence, we are not suggesting that ideal SP-practice needs to involve deliberation. On the other hand, the information provided as part of our study may have been crucial for participants to be able to discover their preferences prior to the initial elicitation. This study did not test for effects of the pre-survey information, but Mathews et al. (2017) do demonstrate using a CE-survey that using more advanced technology, like 3D computer generated models to present the good to be valued reduces choice error, and they reveal a change in stated WTP. This suggests that focusing on the informational part of the survey rather than time to deliberate may be a more productive way of improving SP-surveys. Finally, given the discrepancy noted above between self-reported certainty and the scale parameter, our results make us question the use of self-reported certainty of choices in CE-surveys.

Acknowledgement

We would like to thank Pål Buhl-Mortensen and Jan Fosså, IMR, and Anthony Grehan for informing the survey regarding the biology of cold-water coral. We also would like to thank the 139 participants in the workshops contributing with the data to this paper. Stephane Hess and David Palma has contributed valuable assistance regarding the use of the Apollo package in R to estimate the models. Finally, we would like to thank two anonymous referees for helpful comments. All content and results, including any errors are the responsibility of the authors.

References

Aanesen, Margrethe, Claire Armstrong, Mikolaj Czajkowski, Jannike Falk-Petersen, Nick Hanley and Ståle Navrud. 2015. "Willingness to pay for unfamiliar public goods: Preserving cold-water coral in Norway." *Ecological Economics* 112: 53-67.

Bergstrom, John C., John R.Stoll, and Allan Randall. 1990. "The impact of information on environmental commodity valuation decisions." *American Journal of Agricultural Economics* 72 (3): 614-621.

Brouwer, Roy, Thijs Dekker, John Rolfe, and Jill Windle. 2010. "Choice certainty and inconsistency in repeated choice experiments." *Environmental and Resource Economics* 46 (1): 93-109.

Carson, Richard T., W.Michael Hanemann, Raymond J. Kopp, John A. Krosnick, J.A., Robert C. Mitchell, Stanley Presser, Paul A. Ruud, V. Kerry Smith, Michael Conaway, and Kerry Martin. 1998. "Referendum design and contingent valuation: The NOAA panel's no vote recommendation." *Review of Economics and Statistics* 80 (2): 335-338.

Central Statistics Office (CSO), 2014 Statistics, (Available at): http://www.cso.ie/en/index.html.

Czajkowski, Mikolaj, Marek Giergizny, and William M Greene. 2014. "Learning and fatigue effects revisited: Investigating the effects of accounting for unobservable preferences and scale heterogeneity." *Land Economics* 90 (2): 324-351.

Day, Brett, Ian Bateman, Richard T. Carson, Diane Dupont, Jordan J. Louviere, Sanae Morimoto, Ricardo Scarpa, and Paul Wang. 2012. "Ordering effects and choice set awareness in repeat-response stated preference studies." *Journal of Envirnmental Economics and Management* 63: 73-91.

Fosså, Jan H. and Pål Buhl-Mortensen. 1999. "*Kaldtvannskoraller i Norge*", Universitetsforlaget, Oslo, Norway (only in Norwegian).

Freiwald, Andre, Jan H.Fossa, Anthony Grehan, Tony Koslow, and J.Murray Roberts. 2004. *"Cold-water coral reefs: Out of sight - no longer out of mind*." UNEP-WCMC Biodiversity Series No 22.

Grehan, A., V. Unnithan, A. Wheeler, X. Monteys, T. Beck, M. Wilson, J. Guinan, A. Foubert, M. Klages, and J. Thiede. 2004. *"Evidence of major fisheries impact on cold-water corals in the deep waters off the Porcupine Bank, west coast of Ireland: are interim management measures required?*" ICES Annual Report CM 2004/AA:07.

Guinan, J., A. J. Grehan, M. F. J. Dolan, and C. Brown. 2009. "Quantifying relationships between video observations of cold-water coral cover and seafloor features in Rockall Trough, west of Ireland." *Marine Ecology Progress Series* 375: 125-138.

Haneman, Michael W. 1994. "Valuing the environment through Contingent Valuation." *The Journal of Economic Perspectives* 8 (4): 19-43.

Hensher, David A., Matthew J. Beck, and John M. Rose. 2011. "Accounting for preference and scale heterogeneity in establishing whether it matters who is interviewed to reveal household automobile purchase preferences." *Environmental and Resource Economics* 49 (1): 1-22.

Hess, Stephane, and David Palma. *"Package Apollo. Tools for choice models estimation and applications.*" Available at https://cran.r-project.org/web/packages/apollo/apollo.pdf

Hoehn, John P, and Allan Randall. 1987. "A satisfactory benefit-cost indicator from contingent valuation." *Journal of Environmental Economics and Management* 14 (3): 226-247.

Johnston, Robert J., Kevin J.Boyle, Wiktor Adamowicz, Jeff Bennett, Roy Brouwer, Trudy A. Cameron, Michael W. Hanemann, Nick Hanley, Mandy Ryan, Ricardo Scarpa, Roger Tourangeau, and Christian A. Vossler, 2017. "Contemporary guidance for stated preference studies." *Journal of American Environmental and Resource Economics*, 4 (2): 319-405.

Kingsley, David. C. and Thomas C. Brown. 2010. “Preference uncertainty, preference learning, and paired comparison experiments.” *Land Economics* 86 (3): 530-544

Ladenburg, Jacob. 2013. “Does gender specific starting point bias in choice experiments prevail among well-informed respondents: evidence from an empirical study.” *Applied Economic Letters*, 20 (17): 1527-1530.

LaRiviere, Jacob, Mikolaj Czajkowski, Nick Hanley, Margrethe Aanesen, Jannike Falk-Petersen, and Dugald Tinch. 2014. “The value of familiarity: Effects of knowledge and objective signals on willingness to pay for a public good.” *Journal of Environmental Economics and Management* 68 (2): 376-389.

Lienhoop, Nele, and Douglas MacMillan. 2007. “Valuing wilderness in Iceland: Estimation of WTA and WTP using the market-stall approach to contingent valuation.” *Land Use Policy* 24 (1): 289-295.

Lienhoop, Nele, and Mark Völker. 2016. “Preference refinement in deliberative choice experiments for ecosystem service valuation,” *Land Economics* 92 (3): 555-577.

Lo, Alex Y. and Clive L. Spash. 2013. “Deliberative monetary valuation: In search of a democratic and value plural approach to environmental policy.” *Journal of Economic surveys* 27 (4): 768-789.

Macmillan, Douglas C., Lorna Philip, Nick Hanley, and Begona Álvarez-Farizo. 2002. “Valuing the non-market benefits of wild goose conservation: a comparison of interview and group-based approaches.” *Ecological Economics* 43 (1): 49-59.

Macmillan, Douglas, Nick Hanley, and Nele Lienhoop. 2006. “Contingent valuation: Environmental polling or preference engine?” *Ecological Economics* 60 (1): 299-307.

Matthews, Yvonne, Ricardo Scarpa, and Dan Marsh. 2017a. “Stability of willingness to pay for coastal management: a choice experiment across three time periods.” *Ecological Economics* 138: 64-73.

Matthews, Yvonne, Ricardo Scarpa and Dan Marsh. 2017b. “Using virtual environments to improve the realism of choice experiments: A case study about coastal erosion management.” *Journal of Environmental Economics and Management* 81: 193-208.

Needham, Kathrine, Mikolaj Czajkowski, Nick Hanley, and Jacob LaRiviere. 2018. “What is the causal impact of information and knowledge in stated preference studies?” *Resource and Energy Economics* 54: 69-89.

Paradiso, Massimo, and Antonella Trisorio. 2001. “The effect of knowledge on the disparity between hypothetical and real willingness to pay.” *Applied Economics* 33 (11): 1359-1364.

Plott, Charles. 1996. “Rational individual behavior in markets and social processes: the discovered preferences hypothesis”. In *The Rational Foundations of Economic behavior,* ed. Ken J. Arrow, E.Colombatt, M.Pearlman, C.Schmidt. 225-250. London: McMillan and New York: St.Martin’s Press.

Rose, John M. and Lorenzo Masiero. 2010. “A Comparison of the Impacts of Aspects of Prospect Theory on WTP/WTA Estimated in Preference and WTP/WTA Space.” *European Journal of Transport and Infrastructure Research*, 10 (4): 1567-7141

Samples, Karl C., John A.Dixon and Marcia M. Gowen. 1986. “Information disclosure and endangered species valuation.” *Land Economics* 62 (3): 306-312.

Sandorf, Erlend D., Margrethe Aanesen and Ståle Navrud. 2016. “Valuing unfamiliar and complex environmental goods: A comparison of valuation workshops and internet panel survey with videos.” *Ecological Economics*, 129: 50-61

Scarpa, Ricardo, and John M. Rose. 2008. “Design efficiency for non-market valuation with choice modeling: How to measure it, what to report and why.” *Australian Journal of Agricultural and Resource Economics* 52: 253-282.

Schläpfer, Felix, and Baruch Fischhoff. 2012. "Task familiarity and contextual cues predict hypothetical bias in a meta-analysis of stated preference studies." *Ecological Economics* 81: 44-47.

Shapansky, Bradford, Wiktor L. Adamowicz, and Peter C. Boxall. 2008. "Assessing information provision and respondent involvement effects on preferences." *Ecological Economics* 65 (3): 626-635.

Spash, Clive L. 2008. "Deliberative Monetary Valuation and the evidence for a new value theory." *Land Economics* 84 (3): 469-488.

Train, Kenneth and Michael Weeks. 2005. "Discrete choice models in preference space and willingness to pay space., In: *Applications of Simulation Methods in Environmental and Resource Economics,* eds. Ricardo Scarpa and Anna Alberini. 1-16, Springer, Dordrecht.

Train, Kenneth. 2009. "*Discrete choice methods with simulations*". 2nd ed., New York: Cambridge University Press

Vega, Amaya, and Steven Hynes. 2017. "*Ireland's ocean economy*," Socio-Economic Marine Research Unit (SEMRU) Report series ISSN 2009-6933, National University of Ireland, Galway.

Völker, Mark, and Nele Lienhoop. 2016. "Exploring group dynamics in deliberative choice experiments." *Ecological Economics* 123: 57-67.

1 Choice certainty can either be self-reported in the survey or derived statistically in terms of the variance of the error term in a random utility model as measured by the scale parameter (Brouwer et al 2010).
2 These two papers are based on the same dataset.
3 Size is split into small and large, and commercial is split into fish and oil. These four new attributes are all dummies.
4 This is given by z=sqrt((mean(before) – mean(after))/std.error of the difference of the means).
5 See table A4 in supplementary material for the model when run on each of the sub-sets and with separate scale for choices in the two elicitation situations.