NONE-OF-THESE BIAS IN STATED CHOICE EXPERIMENTS

Frode Alfnes[1] and Gro Steine[1,2]

[1]Norwegian University of Life Sciences

[2]AKVAFORSK

Paper prepared for presentation at the

XIth International Congress of European Association of Agricultural Economists,

Copenhagen, Denmark, August 24-27, 2005

Copyright 2005 by Frode Alfnes and Gro Steine. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

NONE-OF-THESE BIAS IN STATED CHOICE EXPERIMENTS

Frode Alfnes and Gro Steine

Date: July 14, 2005

Corresponding author: Frode Alfnes

Department of Economics and Resource Management

Norwegian University of Life Sciences

P.O. Box 5003, 1432 Aas, Norway.

E-mail: frode.alfnes@umb.no

Frode Alfnes is a postdoctoral fellow, Department of Economics and Resource Management, Norwegian University of Life Sciences, and Gro Steine is a PhD student at AKVAFORSK, Institute of Aquaculture Research AS. Frode Alfnes was a visiting scientist at Iowa State University when the paper was written. The authors would like to thank Atle Guttormsen, Kari Kolstad, Mia Bencze Rørå, Ingrid Olesen, Bjarne Gjerde, Hans Jørgen Whist, and Juan Francisco Aldunce Meyer for valuable discussions and assistance in data collection. The Research Council of Norway (grant nos. 143045/140 and 159523/110) financed this research.

NONE-OF-THESE BIAS IN STATED CHOICE EXPERIMENTS

Abstract
We conduct a within-sample test of hypothetical bias and parameter equality between a hypothetical stated choice (SC) experiment using pictures and a real choice (RC) experiments using products. With exception of the none-of-these alternative-specific dummy, we cannot reject parameter equality between the two datasets. However, when we estimate the models separately with no parameter restrictions, the SC experiment gave WTP estimates that were approximately 50 percent higher and marginal WTP estimates that were almost two times as high as the corresponding estimates from the RC experiment. However, even though the monetary value of the WTP disparity was large, the disparity between the WTP estimates from the two data sets was not statistically significant.

Key words: Choice experiment, color, hypothetical bias, mixed logit, salmon, willingness to pay
JEL: C81, C93, D12, Q22

Introduction

Hypothetical stated choice (SC) experiments are increasingly used to investigate consumer preferences for new and unavailable products, and to investigate socioeconomic differences in product preferences. Recent examples of SC experiments used to study food demand include Burton et al. (2001), Lusk (2003), and Lusk et al. (2004), who investigated the demand for GMO food; Blend and van Ravenswaay (1999), Wessells, Johnston and Donath (1999), and Loureiro, McCluskey and Mittelhammer (2001), all of whom studied the demand for ecolabeled food; Holland and Wessells (1998), who investigated preferences for safety inspection and production method of salmon; Steine, Alfnes and Rørå (2005), who investigated the effect of salmon color on WTP; Enneking (2004), who analyzed the WTP for safety improvement in the meat sector; Unterschultz et al. (1998), Quagrainie, Unterschultz and Veeman (1998), and Alfnes (2004), all of whom analyzed preferences for country-of-origin for beef; and Lusk, Roosen and Fox (2003), who analyzed international preferences for hormone-treated beef.[1]

Common for all these SC experiments is that the presentations of the alternatives are done with product descriptions or pictures and that there are no real economic incentives. In fact, not relying on real products is one of the major strength of hypothetical studies and allows researchers to investigate consumer preferences for hypothetical products and products still on the drawing board. On the other hand, lack of economic incentives and real consequences is also a problem for SC experiments.

The empirical evidence strongly suggests that participants in most hypothetical studies exaggerate their WTP for private as well as public goods; see List and Gallet (2001), List (2001), List and Shogren (2002), Harrison and Rutström (forthcoming), and Murphy et al. (2005). However, there is mixed evidence regarding the importance of real economic incentives in choice experiments. Carlsson and Martinsson (2001) used a within-subject design with 34 participants to test for differences in marginal WTP between a SC and a non-hypothetical choice experiment, hereafter referred to as a real choice (RC) experiment, for donations to public goods. The goods were three different environmental projects managed by the World Wildlife Foundation. Due to their choice of payment mechanism, a none-of-these (NOT) option would have been strictly dominated and they did not include one in their design. They found an insignificant discrepancy between marginal WTP for donations in the SC and the RC experiment. Lusk and Schroeder used a between-subject design with a total of 104 participants to test for differences in WTP for private goods. They estimated WTP for quality-differentiated beef steaks using both a SC and a RC experiments. Each choice scenario had five product alternatives and a NOT option. Lusk and Schroeder (2004) confirmed Carlsson and Martinsson (2001) results of an insignificant discrepancy in marginal WTP estimates. In addition, Lusk and Schroeder (2004) found that the results of the SC experiment overestimated the total WTP by underestimating the popularity of the NOT option.

Both Carlsson and Martinsson (2001) and Lusk and Schroeder (2004) deviates on major points from the majority of the food demand SC experiments discussed above. Carlsson and Martinsson (2001) elicit WTP for a public good, and Lusk and Schroeder (2004) uses real products in the SC experiment. Neither of them investigates the properties of the typical SC experiment for private goods where the products are presented by pictures or written product descriptions. In most cases, the reason to do a hypothetical study instead of a non-hypothetical study is that the products are not available. The effect of no economic incentives and products presented with pictures or written product descriptions are typically confounded in SC experiments, and it is important to study the combined effect of these features.

Is the product presentation important? In the only published comparison of hypothetical markets using pictures and non-hypothetical markets using real products, that we are aware of, Melton et al. (1996) used Vickrey auctions to elicit WTP for quality attributes of pork shops. They conclude that the results of hypothetical auction using pictures where "an especially unreliable source of market information" (p. 923). In light of Melton et al.'s (1996) conclusion, it is important to extend the Lusk and Schroeder (2004) analysis to SC experiments not using real products.

Participants in both SC and RC experiments are asked to make choices in a series of choice scenarios. In each scenario, two or more product alternatives are presented with their attributes. A NOT alternative is commonly included in each choice scenario to avoid forced choices between the product alternatives. To induce real economic incentives in the RC experiment, one of the choice scenarios is randomly drawn as binding at the end of the experiment. The alternative that each participant has chosen in the binding RC scenario is implemented for the respective participants. In the SC experiment the choice has no economic consequences, whereas in the RC experiment the choice is nonhypothetical. It is in the RC experiment participants' own interests to choose the alternative they prefer in each scenario, and their incentive to reveal true preferences is relatively transparent.

By varying the product attributes between the choice scenarios in SC and RC experiments, the researchers can estimate logit or probit models with sensitivity parameters for the various attributes. If a variable, e.g., price, is negatively correlated with the choice frequency, every thing else kept constant, the respective parameter will be negative, and if a variable is positively correlated with the choice frequency, then the respective parameter will be positive. The size of the individual parameters has no economic interpretation in itself, but the ratio between two parameters is equal to the marginal rate of substitution between the two attributes in the choice experiment. Marginal willingness to pay (WTP) for the product attributes is found as the marginal rate of substitution between the respective product attributes and money (usually represented by the negative of the price). The total WTP for a product alternative is found by subtracting the utility of the NOT alternative from the total utility of the product alternative, and dividing the sum by the negative of the price sensitivity parameter.

The objective of this paper is to conduct a within-sample test of hypothetical bias and parameter equality between a SC experiment using pictures and a RC experiments using real products. The analysis is done on the results of a choice experiment conducted to investigate consumers' WTP for salmon color. The rest of the paper is as follows. First, we give some background information on salmon color, followed by the experimental procedure, products, design choice scenarios, sample, econometric model, results, and concluding remarks.

Background on Salmon Color

Salmon are recognized for their pink-red flesh color, which distinguishes them from other species. Consumers use intrinsic cues such as color to infer the quality of food products. In surveys as well as focus groups, consumers have stated that they see the color of salmon as an indicator of flavor and freshness. For a review of the literature on consumers' attitude toward salmon color, see Steine, Alfnes and Rørå (2005).

The internationally recognized method for salmon color measurement is comparing the salmon fillet flesh with the colors in the *SalmoFan*™. The SalmoFan is a color fan developed on the basis of the color of salmonid flesh pigmented with astaxanthin. The color of conventional farmed salmon fillets sold in the Norwegian market normally range from 23 to 30 on the SalmoFan, and most common are fillets ranging from 25 to 27.

Experimental Procedure

The experimental session included a survey, an SC experiment with pictures, and an RC experiment. The SC experiment consisted of 10 choice scenarios and the RC experiment consisted of three sets of 10 choice scenarios. The last 10 RC scenarios, focusing on organic and ecolabeled salmon, will not be analyzed in this article. In each choice scenario, the participants chose between two salmon fillets with posted prices. After the first 10 RC scenarios we conducted the SC experiment. The setup of the SC experiment was the same as the setup of the RC experiment, but the salmon fillets were replaced with pictures of salmon fillets, and the choices were hypothetical. See Table 1 for an example of the choice scheme used in both the RC and the SC experiment.

In each of the RC scenarios we displayed two 400-gram consumer packages of salmon fillets placed on a box filled with ice. The prices of the two alternatives in each scenario were posted on laminated paper in the back of the boxes. The participants chose one of the two salmon fillets, or a NOT alternative, in each scenario. To induce real economic incentives, each participant drew one binding scenario; the participants then had to buy the salmon fillet they had chosen in their binding scenario. This setup was very flexible and allowed us to vary not only the price, as done in earlier non-hypothetical choice-based market experiments for food products (Lusk and Schroeder, 2004), but also to vary the products among the scenarios according to a statistical design.

The RC experiment had nine steps. Step 1: The experimental procedure was explained to the participants. Step 2: The participants studied the alternatives in scenarios 1 to 10, and marked on a choice scheme which of the alternatives in each scenario they wanted to buy. Step 3: The participants were informed about the origin of the color. Step 4: The participants studied the alternatives in scenarios 11 to 20, and (as in Step 2) marked on a choice scheme which of the alternatives in each scenario they wanted to buy. Step 5: The participants were informed about organic and ecolabeled salmon. Step 6: The participants studied the alternatives in scenarios 21 to 30, and (as in Steps 2 and 4) marked on a choice scheme which of the alternatives in each scenario they wanted to buy. Step 7: After all participants had completed all scenarios, each participant drew one card determining his or her binding scenario. The drawing was done without replacement, so that only one participant was assigned to each scenario. Step 8: Each participant got the salmon fillet he or she had chosen in his or her binding scenario. Step 9: The participants went to the cashier and paid for their salmon fillets.

In half the sessions, the participants conducted the SC experiment before Step 3 in the RC experiment and in the other half after Step 3. The proportion of the choices in the SC experiment and the RC experiment done before Step 3 was the same. None of the parameters changed significantly after the information was supplied to the participant, and we have therefore not included any information dummies in this paper. For a discussion of the effect of the information in the SC experiment, see Steine, Alfnes and Rørå (2005), and for a discussion of the effect of the information in the RC experiment, see Alfnes et al (2005).

The test of hypothetical bias will be based on a within-subject design. All subjects conduct both hypothetical and nonhypothetical choices. The primary disadvantage with a within-subject design is that participation in a nonhypothetical treatment can alter the behavior in a hypothetical treatment or vice versa. Therefore, we are more likely to find no hypothetical bias than if we had used a between-subject design. On the other hand, the within-subject design gives more observations in each treatment and the within-subject design generates statistical tests with greater power since we know that the differences we find are not related to differences in the sample.

Products

For the RC color experiment we bought salmon from three conventional and one organic salmon farm. The salmon fillets were cut into portions weighing approximately 400 grams, put into packaging familiar to consumers, exactly weighted, and we recorded if the fillet portions were from the front or tail of the fillet. The fillet colors were determined using a SalmoFan, and colors ranged from 20 to 30 on this scale. The fillets were grouped into five color categories; hereafter referred to as alternatives R21, R23, R25, R27, and R29. The mean weight of the fillets was 400.28 grams with a standard deviation of 40.25 grams. To prevent weight playing an important role in the choices, we imposed a 40-gram upper limit of by how much a choice pair were allowed to differ.

For the SC experiment, we used 18 cm x 13 cm pictures of salmon fillets obtained from the salmon photo library of Photofish, a subsidiary company of AKVAFORSK. We developed pictures of salmon fillets with a flesh color that were equal to alternatives R21, R23, R25, R27, and R29. In addition, we also developed pictures of an extremely red salmon, matching 32 on the SalmoFan. We will hereafter refer to this last alternative as R32. The color of the pictures was checked after development. Fillets as pale as 21 are only available from organic salmon producers and are never seen in the Norwegian market, whereas fillets as red as 32 are usually seen only for rainbow trout.

Price was a five-level attribute, with the levels NOK 24, 30, 36, 42, and 48 per 400 grams in both the RC and SC experiments.[2] This corresponds to NOK 60, 75, 90, 105, and 120 per kilogram. The week before the SC experiment, the price of salmon fillets in the three largest grocery stores in the area were NOK 79, 89, and 119 per kilo. Thus, all prices except for that of NOK 24 are within a familiar price range for salmon fillets in Norway.

Design Choice Scenarios

We used a SAS macro to generate the statistical design of the choice scenarios. We generated two fractional factorial designs with 40 choice scenarios in each. Each scenario had two alternatives described by color and price. In both designs price was a five-level attribute, while color was a five-level attribute in the RC design and a six-level attribute in the SC design. To avoid clearly dominated alternatives we limited the design to scenarios where the color of the two alternatives differed. There were, however, no limitations on the price attribute, and several scenarios had the same price for both alternatives. Both designs were divided into four blocks of scenarios, and the scenarios were randomly arranged within the blocks. SAS reported a D-efficiency of 96.85 for the RC design and a D-efficiency of 95.70 for the SC design. Each block of scenarios in the RC design was used once as scenarios 1 to 10, and in another session as scenarios 11 to 20. For a description of the SAS macro, see Kuhfeld (2001).

Sample

The experiment was conducted at MATFORSK, The Norwegian Food Research Institute, during four nights in February 2004. We conducted two sessions each night, and the sessions lasted approximately one and a half hours each. In total, 115 participants were recruited through various local organizations in southeastern Norway. Between 13 and 16 persons participated in each session. The organizations were given NOK 200 for each participant they recruited, and the participants were given NOK 300 to take part in the experiment. One participant who said that he did not eat fish was excluded from the sample.

Table 2 presents the descriptive statistics for the sample. The participants' ages ranged from 20 to 63 years, with an average of 39 years. Fifty-eight percent of the participants were women. The average household income was NOK 562,000.

Econometric Model

In this study the participants made discrete choices among a set of exclusive alternatives. We follow the recent trend in the choice literature and analyze the data with a mixed logit model (also known as a random parameter logit model). Let us assume that the individual's utility from each alternative can be decomposed into a linear-in-parameters part that depends on observable variables, and an error term that is an independently and identically distributed (iid) extreme value. Given these assumptions, the utility of individual *n* from alternative *i* in choice scenario *s* is denoted by:

(1) $$U_{nis} = \beta' x_{nis} + \sigma_s^{-1} * \varepsilon_{nis}$$

where x_{nis} is a vector of observed variables relating to alternative *i*; β is a vector of fixed and random parameters; σ_s is a scaling parameter that takes the value one when there is no parameter restrictions across data sets, and one for the RC data when there is parameter restrictions across data sets; and ε_{nis} is an iid extreme value error term. The standard logit is a special case of the mixed logit where the variance of all the random parameters is zero, i.e., all parameters are fixed.

The choice data from the RC and SC experiments were analyzed with the following mixed logit model:

(2) $$U_{nis} = (\beta_{0i} + \beta_1 Tail_{nis}) Weight_{nis} + \beta_2 Price_{nis} + \beta_3 Price24_{nis} + \sigma_s^{-1} * \varepsilon_{nis}$$

where β_{0i} is the alternative specific constant for color i, in other words, there is a constant for each color; $Tail_{nis}$ is a dummy taking the value one if the product is a fillet tail, and zero otherwise; $Weight_{nis}$ is the exact weight of the alternative i in kilograms[3]; $Price_{nis}$ is the price of alternative i in NOK 100; $Price24_{nis}$ is a dummy taking the value one if the price is NOK 24, and zero otherwise; and σ_s a scaling parameter taking the value one when the two data sets is estimated separately, and is taking the value 1 for the RC data in joint estimation of the two data sets. The random parameters β_{0i} are triangularly distributed,[4] and constant over choice made by the same individual. We have restricted the standard deviations of the random parameter to be identical for all color alternatives and to be identical for the two NOT alternatives. For identification, the alternative-specific parameters for the palest alternative, R21, is normalized to zero. For estimation purposes, the weight of the NOT alternative is set to one.

We inspected the data and found that the choice probability increased as we reduced the price down to NOK 30. However, in both the SC and RC data we found that the choice probability significantly decreased from NOK 30 to NOK 24. Since we were interested in the price sensitivity for price levels found in the market (i.e., prices higher than NOK 30), we have included a dummy for the NOK 24 price. For a further discussion of the NOK24 dummy, see Alfnes et al. (2005).

The mean WTP per kilogram of alternative i can be calculated by dividing the utility difference between one kilogram of the varieties and the NOT alternative, with the negative of the price sensitivity parameter. Since the price sensitivity parameter measures the utility of the price in NOK 100, we must multiply the result by 100 to get the WTP in NOK:

(3) $$WTP_{is} = -100 * \frac{(\beta_{0i} + \beta_1 Tail_{is}) - (\beta_{0NOT})}{\beta_2}$$

where WTP_{is} is the estimated mean WTP per kilogram of alternative i in scenario s; and all other variables and parameters are as described in equation (2).

All the currently available software packages assume a fixed scale for the iid-residual in mixed logit models, and therefore we cannot estimate the scaling factor together with the other parameters in the joint estimation of the RC and the SC data sets. To overcome this problem we conducted a manual search for the optimal scaling factor. This was done by multiplying all the variables in the SC dataset with a scaling factor, and then estimating the mixed logit model for the combined dataset. We started out with a scaling factor of 1.0, and then gradually decreased or increased the factor by 0.1 until we reached a point where the last two decreases or increases reduced the log likelihood. We then narrowed the search for the optimal scaling factor around the scaling factor that gave the highest log likelihood in the first round of searches. The reported models are the models with the optimal scaling factor, i.e., having the highest log likelihood. This procedure was originally proposed for multinomial logit models by Swait and Louviere (1993), and is suggested as a manual method by Louviere, Hensher and Swait (2000, page 238). This is a time consuming procedure for the mixed logit model, and we do not obtain a standard error estimate for the scaling factor. For a general discussion of the scale parameter in logit models, see Train (2003, pp. 44–46) or Louviere, Hensher and Swait (2000, pp. 234–236).

Results

Table 3 presents the estimated parameters for a multinomial and a mixed logit model for both the RC and SC data. First, we can see that the price parameters are negative and significant, meaning that increased price reduce the utility and the likelihood of choosing an alternative. Second, the utility is increasing in color. The consumers preferred redder salmon to paler salmon. Third, the NOT alternative specific parameters are negative in all models and significant in the mixed logit estimations (model 2 and 4). The NOT alternative is inferior to all the salmon alternatives at zero price. Fourth, the tail dummy is negative but not significantly. The consumers slightly prefer the thicker front part of the fillets to the tails of the fillets. Fifth, the Price 24 parameter is negative. Offering the salmon at a price that was lower than what is usually seen in the market had an adverse effect on the choice probability. Sixth, the standard deviations for the alternative specific constantans in the mixed logit model are positive, significant and not homogenous. This is inconsistent with independent of irrelevant alternatives assumption underlying the multinomial logit model. Seventh, the absolute value of the parameter estimates in the SC models are, with exception of the price sensitivity parameter, larger than the parameters in the RC models. The variables summarized in the residuals are less important in the SC model then in the RC. For a more thorough discussion of these results, see Alfnes et al. (2005) and Steine, Alfnes and Rørå (2005).

Table 4 presents the WTP estimates for four models presented in table 3. The differences in the NOT parameter between the multinomial and the mixed logit models and the difference in relative size of the price parameter between the RC and SC estimates are evident in the WTP estimates. First, we can see that the level of the WTP estimates in the mixed logit models is higher than the level in the multinomial logit models. This is mainly due to the fact that the standard error of the error term in the multinomial logit model is homogenous, while the standard error of the NOT alternative is higher than the standard error of the salmon alternatives in the mixed logit model. The increase in standard error allows the mean of the NOT parameter to be lower, and still correspond to the same number of NOT choices. For the RC data the mixed logit model estimates a WTP that is between NOK 19.92 and NOK 22.66 higher than the multinomial logit model. The variation between the models is bigger for the SC data, with a minimum difference of NOK 20.93 and a maximum difference of NOK 34.45. Second, the WTP estimates from the SC data are higher than the WTP estimates in the RC data. The mean calibration factor between the two multinomial logit models is 1.56 and between the two mixed logit models is 1.53, and the mean calibration factor for the marginal WTP between the two multinomial logit models is 1.96 and between the two mixed logit models is 1.94.

To test if the differences in WTP between the SC and RC data are statistically significant, we conducted a series of Wald tests. We cannot reject the null-hypotheses of equal WTP between the RC and the SC data set for R21 (p-value = 0.23), R23 (p-value = 0.26), R25 (p-value = 0.14), R27 (p-value = 0.16) and R29 (p-value = 0.13). Furthermore, we cannot reject the null-hypotheses of equal marginal WTP between the RC and the SC data set for R29-R27 (p-value = 0.56), R29-R25 (p-value = 0.53), R29-R23 (p-value = 0.09), R29-R21 (p-value = 0.19), R27-R25 (p-value = 0.89), R27-R23 (p-value = 0.15), R27-R21 (p-value = 0.24), R25-R23 (p-value = 0.13), R25-R21 (p-value = 0.22), and R23-R21 (p-value = 0.44). Even though the hypothetical bias is large in monetary terms, it is not statistically significant for these models.

To further explore the differences between the RC and SC responses we have estimates the mixed logit model on the joint RC and SC dataset. Table 5 presents the results for three different samples. The first sample includes the RC and SC data from all 114 fish-eating participants. The second sample includes the RC data from the 99 participants who chose one of the product alternatives in at least one of the 20 RC scenarios and the SC data from the 108 participants who chose one of the product alternatives in at least one of the 10 SC scenarios. The third sample includes the RC and SC data from the 99 participants who chose one of the product alternatives in at least one of the 20 RC scenarios. The joint mixed logit model has restrictions on the four color-specific parameters, the price parameter, the Price24-dummy parameter, and the two standard deviation parameters. By estimating the mixed logit model for the joint RC and SC data, as well as separately for the RC and SC data, we can test for preference regularity with respect to these variables. We use a log ratio test, $LR = -2 [LL^{Joint} - (LL^{RC} + LL^{SC})]$, which is χ^2 distributed with seven degrees of freedom. The test statistics for samples 1, 2, and 3 are LR = 12.46 (*p-value* = 0.09), LR = 10.70 (*p-value* = 0.15), and LR = 11.84 (*p-value* = 0.11), respectively. Parameter equality is not rejected at a 5% level for any of the three samples.

The results of the joint model presented in tables 5, confirms that the price sensitivity is slightly, but not significantly, lower in the SC data than in the RC data. We cannot reject the null hypothesis of equal price sensitivity in the SC and RC data, Wald = 0.41 (p = 0.52), Wald = 0.35 (p = 0.55), and Wald = 0.00 (p = 0.97), for the three samples, respectively. For the NOT parameter, however, the parameter value from the hypothetical SC data was significantly lower than the NOT parameter from the non-hypothetical RC data in all three samples, Wald = 9.76 (p = 0.00), Wald = 6.15 (p = 0.01), and Wald = 7.16 (p = 0.01), respectively. Using sample 2 (which correspond to the samples used in table 3), we find only weak evidences of hypothetical bias in total WTP, and cannot reject the null-hypotheses of equal WTP between the RC and the SC data set for R23 (p-value = 0.07), R25 (p-value = 0.11), R27 (p-value = 0.14) and R29 (p-value = 0.14), but we reject the null hypothesis for R21 (p-value = 0.00). We cannot reject the null-hypotheses of equal marginal WTP between the RC and the SC data set for R29-R27 (p-value = 0.97), R29-R25 (p-value = 0.59), R29-R23 (p-value = 0.57), R29-R21 (p-value = 0.56), R27-R25 (p-value = 0.57), R27-R23 (p-value = 0.57), R27-R21 (p-value = 0.56), R25-R23 (p-value = 0.57), R25-R21 (p-value = 0.56), and R23-R21 (p-value = 0.57).

These results is consistent with the findings of Lusk and Schroeder (2004), who found that hypothetical choices overestimated total WTP for beef steaks, but that marginal WTP for a change in steak quality was, in general, not statistically different across hypothetical and actual payment settings. This means that even in a within-subject test of hypothetical bias, participants choose the NOT alternative more frequently in the hypothetical SC experiment than in the nonhypothetical RC experiment. However, given that they chose one of the alternatives, their response to the price variable was not significantly different between the two settings. In terms of WTP, this means that the level of the WTP estimates relative to the NOT alternative was higher in the hypothetical than in the nonhypothetical data. However, the marginal difference in WTP between the salmon color alternatives are not significant different between the two datasets.

The disparity in the NOT estimate is a result of the frequency the participants chose the NOT alternative in the SC and RC experiment. Table 6 shows the frequency with which the participants chose the NOT alternative in the two series of RC scenarios and the series of SC scenarios. The participants are significantly less likely to choose the NOT alternative in the SC scenarios. In the first 10 RC scenarios, 54 (47.37%) of the participants chose the NOT alternative less than three times, and in the last 10 RC scenarios 59 (51.75%) participants chose the NOT alternative less than three times. In the 10 SC scenarios, on the other hand, 84 (73.68%) of the participants chose the NOT alternative less than three times. Furthermore, in the first 10 RC scenarios, 32 (28.07%) of the participants chose the NOT alternative at least eight times, and in the last 10 RC scenarios 24 (21.05%) participants chose the NOT alternative at least eight times. In the 10 SC scenarios, on the other hand, only seven (6.14%) of the participants chose the NOT alternative at least eight times.

To illuminate the response patterns that gave the disparity in the NOT estimates, we investigated each participant's series of responses over the SC and RC experiment. Table 7 presents a cross table for the frequency of the NOT choice in the SC and RC experiments. From the first row, we can see that 58 participants never chose the NOT alternative in the SC experiment. Of these, only 30 never chose the NOT alternative in the RC experiment, and nine chose the NOT alternative 18 or more times in the RC experiment. Furthermore, we notice that the lower left triangle of Table 7 is filled with zeros. In fact, only 21 (18.42%) of the participants had a higher percentage of NOT choices in the SC experiment than in the RC experiment, while 61 (53.51%) had a higher percentage of NOT choices in the RC experiment than in the SC experiment. In a reduced sample where we excluded the 15 participants who chose the NOT alternative in all 20 RC scenarios, the corresponding numbers are 21 (21.21%) and 52 (52.52%). This shows that a significant number of the participants that never chose a product alternative in the RC experiment chose product alternatives in the SC experiment. Furthermore, even among those who chose product alternative in some RC scenarios there is a hypothetical bias with respect to the NOT alternative. They chose the NOT alternative more often in the non-hypothetical RC experiment than in the hypothetical SC experiment.

Concluding Remarks

The SC experiment gave WTP estimates that were approximately 50 percent higher and marginal WTP estimates that were almost two times as high as the corresponding estimates from the RC experiment. These calibration factors are in line with what is found using other types of valuation mechanisms. However, even though the monetary level of the disparity in WTP estimates was large, the disparity between the WTP estimates from the two data sets was not statistically significant.

In contrast to Melton et al. (1996), we cannot reject parameter equality between the hypothetical market using pictures and the non-hypothetical market using products, except for the NOT alternative-specific dummy when we estimate the two data sets together. This is very positive with respect to the possibility of using SC data to extrapolating RC results to product and subjects not included in the RC experiment.

There are several aspects that are not included in either this study or the Lusk and Schroeder (2004) study that may very well result in a significant hypothetical bias in both marginal and total WTP estimates for private goods. First, as discussed by Johansson-Stenman and Svedsäter (2004), product attributes with an ethical dimension such as fair trade, environmental friendly, or animal friendly are much more likely to have a hypothetical bias than uncontroversial products attributes such as beef tenderness and the color of salmon. Second, some product attributes have important externalities that people are more likely to take into account in a hypothetical setting. An example of such externalities can be seen in Alfnes (2004) and Alfnes and Rickertsen (2003). Alfnes (2004), using an SC experiment, found much higher WTP differences between domestic and imported beef than did Alfnes and Rickertsen (2003), using an experimental auction with real economic incentives. Most participants expressed strong support for the current Norwegian small-unit, high-cost agricultural policy (Alfnes, 2004). Since the SC is hypothetical, the SC participants can consider any externalities from their choices without any costs. In such cases, the SC experiment will come closer to reflecting a referendum over the agricultural politics than reflect what would happen if the imported alternatives were available in the stores. Third, in experimental markets, real physical goods are used, whereas in many SC surveys, only a list of attributes is used. This difference in product representation is an unexplored source of hypothetical bias. These three potential sources of hypothetical bias need further investigation before we can draw any general conclusions about hypothetical bias in choice experiments.

Footnotes

[1] For a thorough survey of SC methods and applications in other fields, see Louviere, Hensher and Swait (2000).
[2] NOK 1 = EUR 0.1144 = US$ 0.1434. February 4, 2004. http://www.oanda.com.
[3] Multiplying $Weight_{nis}$ with both the alternative specific constants and the tail variable, imply a linear increase in the utility from an increase in weight for both regular and tail fillets. In other words, the utility of a 420 grams regular fillet is assumed to be 5% higher than the utility of a 400 grams regular fillet with the same color, and the utility of a 420 grams tail fillet is assumed to be 5% higher than the utility of a 400 grams tail fillet with the same color.
[4] The triangular distribution is chosen over the normal distribution, since the former is a limited distribution while the latter is not. Unlimited distributions imply unlimited WTP. For a discussion of various distributions on the non-iid error term, see Hensher and Greene (2003).

References

Alfnes, F. "Stated Preferences for Imported and Hormone-Treated Beef: An Application of the Mixed Logit Model." European Review of Agricultural Economics 31(2004):19–37.

Alfnes, F., A. Guttormsen, G. Steine, and K. Kolstad. "Consumers' Willingness to Pay for the Color of Salmon: A Choice Experiment with Real Economic Incentives." Paper presented at the American Agricultural Economic Association's annual meeting in Providence, July 24-27, 2005.

Alfnes, F., and Rickertsen, K. "European Consumers' Willingness to Pay for U.S. Beef in Experimental Auction Markets." American Journal of Agricultural Economics 85(2003):396–405.

Blend, J.R. and van Ravenswaay, E.O. "Measuring Consumers' Demand for Ecolabeled Apples." American Journal of Agricultural Economics 81(1999):1072–1077.

Burton, M., Rigby, D., Young T. and James, S. "Consumer Attitudes to Genetically Modified Organisms in Food in the UK." European Review of Agricultural Economics 28(2001):479–98.

Carlsson, F. and Martinsson P. "Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?" Journal of Environmental Economics and Management 41(2001):179–92.

Enneking, U. "Willingness-to-Pay for Safety Improvements in the German Meat Sector: The Case of the Q&S Label." European Review of Agricultural Economics 31(2004):205–23.

Harrison, G.W. and Rutström, E.E. "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Experiments." In Handbook of Experimental Economics Results, eds., C.R. Plott and V.L. Smith. New York: Elsevier Press, forthcoming.

Hensher, D.A., and Greene W.H. "The Mixed Logit Model: The State of Practice." Transportation 30(2003):133–76.

Holland, D., and Wessells, C.R. "Predicting Consumer Preferences for Fresh Salmon: The Influence of Safety Inspection and Production Method Attributes." Agricultural and Resource Economics Review 27(1998):1–14.

Johansson-Stenman, O. and Svedsäter, H. "Cheap Talk Does not Always Come Cheap: Hypothetical Bias in Choice Experiments." Unpublished paper, Department of Economics, Göteborg University, Göteborg, Sweden, 2004.

Kuhfeld, W.F. "Standard Logit, Discrete Choice Modeling: An Introduction to Designing Choice Experiments, and Collecting, Processing, and Analyzing Choice Data with SAS." TS-643. Cary, NC: SAS Institute Inc., 2001.

List, J.A. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportcards." American Economic Review 91(2001):1498–507.

List, J.A. and Gallet, G.A. "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?" Environmental and Resource Economics 20(2001):241–54.

List, J.A. and Shogren, J.F. "Calibration of Willingness to Accept." Journal of Environmental Economics and Management 43(2002):219–233.

Loureiro, M.L., McCluskey, J.J. and Mittelhammer, R.C. "Assessing Consumer Preferences for Organic, Eco-Labeled, and Regular Apples." Journal of Agricultural and Resource Economics 26(2001):404–16.

Louviere, J.J., D.A. Hensher, and J.D. Swait. Stated Choice Methods: Analysis and Application. Cambridge: Cambridge University Press, 2000.

Lusk, J.L. "Effect of Cheap Talk on Consumer Willingness-to-Pay for Golden Rice." American Journal of Agricultural Economics 85(2003):840–856.

Lusk, J.L., House, L.O., Valli, C., Jaeger, S.R., Moore, M., Morrow, B. and Traill, W.B. "Effect of Information about Benefits of Biotechnology on Consumer Acceptance of Genetically Modified Food: Evidence from Experimental Auctions in United States, England, and France." European Review of Agricultural Economics 31(2004):179–204.

Lusk, J.L., Roosen, J. and Fox, J.A. "Demand for Beef from Cattle Administered Growth Hormones or Fed Genetically Modified Corn: A Comparison of Consumers in France, Germany, the United

Kingdom, and the United States." American Journal of Agricultural Economics 85(2003):16–29.
Lusk, J.L., and Schroeder T.C. "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks." American Journal of Agricultural Economics 86(2004):467–82.
Melton, B.E. W.E. Huffman, J.F. Shogren and J.A. Fox. "Consumer Preferences for Fresh Food Items with Multiple Quality Attributes: Evidence From an Experimental Auction of Pork Chops." American Journal of Agricultural Economics. 78(1996):916-923.
Murphy, J.J., Allen P.G., Stevens, T.H. and Weatherhead, D. "A Meta Analysis of Hypothetical Bias in Stated Preference Valuation." Environmental and Resource Economics 30(2005):313-25.
Quagrainie, K.K., Unterschultz, J. and Veeman, M. "Effects of Product Origin and Selected Demographics on Consumer Choice of Red Meats." Canadian Journal of Agricultural Economics 46(1998):201–19.
Steine, G., Alfnes, F. and Rørå, M.B. "The Effect of Color on Consumer WTP for Farmed Salmon." Marine Resource Economics 20(2005):211-19.
Swait, J.D. and Louviere, J.J. "The Role of the Scale Parameter in the Estimation and Use of Multinomial Logit Models." Journal of Marketing Research 30(1993):305–14.
Train, K. Discrete Choice Methods with Simulation. Cambridge: Cambridge University Press, 2003.
Unterschultz, J., Quagrainie, K.K., Veeman, M. and Kim, R.B. "South Korean Hotel Meat Buyers' Perceptions of Australian, Canadian and US Beef." Canadian Journal of Agricultural Economics 46(1998):53–68.
Wessells, C.R., Johnston, R.J. and Donath, H. "Assessing Consumers Preferences for Ecolabeled Seafood: The Influence of Species, Certifier, and Household Attributes." American Journal of Agricultural Economics 81(1999):1084–9.

Table 1. Example of choice scheme

Scenario 1	400 grams of farmed salmon		None of these
	Alternative 1 NOK 36	Alternative 2 NOK 48	
I would choose (check ✓one)	☐	☐	☐

Table 2. Descriptive statistics for the sample

Variable	Definition	Mean	St.dev.[a]
Gender	Gender of participant (Female = 1; Male = 2)	1.43	0.49
Age	Age of participant	38.81	10.29
Income	Total income of household[a] (in NOK 100 000)	5.62	2.63
Education	Highest completed education (Elementary school = 1; High school = 2; College/University = 3)	2.54	0.67

[a]The income question had six classes. The midpoints of the classes were used in the estimation.

Table 3. Estimated parameters for the logit models

Variables	Model 1 (RC, MNL)[a]	Model 2 (RC, MXL)[a]	Model 3 (SC, MNL)[b]	Model 4 (SC, MXL)[b]
Price sensitivity parameters				
Price[c]	-3.25	–3.59	-2.87	–3.27
	(0.58)	(0.62)	(0.82)	(0.92)
Alternative specific constant for the none-of-these alternative, NOT				
NOT	-0.45	–1.28	-0.49	–1.67
	(0.25)	(0.33)	(0.38)	(0.51)
Alternative specific constant for the colors				
R23	2.63	2.86	3.53	3.55
	(0.33)	(0.36)	(0.53)	(0.58)
R25	3.13	3.49	4.86	5.20
	(0.33)	(0.36)	(0.53)	(0.58)
R27	3.58	3.96	5.36	5.69
	(0.34)	(0.38)	(0.52)	(0.58)
R29	3.51	3.82	5.49	5.87
	(0.38)	(0.42)	(0.53)	(0.59)
R32			5.80	6.15
			(0.54)	(0.61)
Generic parameters				
Tail dummy	–0.40	–0.42		
	(0.28)	(0.31)		
Price24 dummy	–0.45	–0.51	-0.32	–0.39
	(0.12)	(0.14)	(0.19)	(0.21)
Standard deviation parameters				
R23-R32		2.09		3.78
		(0.63)		(0.77)
NOT		5.35		4.72
		(0.44)		(0.59)
Summary statistics				
# Participants	99	99	108	108
# Choice observations	1980	1980	1080	1080
Log likelihood	-2060.66	–1816.81	-955.27	-887.60

Notes: Estimated with Nlogit 3.0. 250 Halton draws.
[a]RC data excluding the 15 participants who chose the NOT alternative in all 20 RC scenarios. [b]SC data excluding the six participants who chose the NOT alternative in all 10 SC scenarios. [c]Price in NOK 100.

Table 4. Estimated willingness to pay

Variables	**Model 1** (RC, MNL)[a]	**Model 2** (RC, MXL)[a]	**Model 3** (SC, MNL)[b]	**Model 4** (SC, MXL)[b]
R21	13.86	35.62	17.18	51.63
	(5.63)	(6.57)	(9.39)	(11.93)
R23	94.72	115.09	140.47	161.40
	(11.93)	(15.53)	(32.27)	(38.57)
R25	109.98	132.64	186.99	212.12
	(14.14)	(18.04)	(43.46)	(50.78)
R27	123.85	145.71	204.06	227.28
	(16.62)	(20.44)	(47.61)	(54.29)
R29	121.90	141.82	209.01	232.84
	(17.18)	(20.79)	(49.58)	(56.22)
R32			219.43	241.73
			(55.10)	(61.32)

Notes: Estimated with Nlogit 3.0.
[a]RC data excluding the 15 participants who chose the NOT alternative in all 20 RC scenarios. [b]SC data excluding the six participants who chose the NOT alternative in all 10 SC scenarios.

Table 5. Estimated parameters for the pooled mixed logit model

Variables	Sample 1[a]	Sample 2[b]	Sample 3[c]
Price sensitivity parameters			
Price (RC) [d]	–3.35	–3.39	–3.46
	(0.56)	(0.57)	(0.57)
Price (SC)	–2.92	–2.98	–3.47
	(0.60)	(0.62)	(0.65)
Alternative specific constant for the none-of-these alternative, NOT			
NOT (RC)	–0.41	–0.98	–0.99
	(0.33)	(0.29)	(0.29)
NOT (SC)	–1.95	–1.95	–2.07
	(0.42)	(0.37)	(0.38)
Alternative specific constant for the colors			
R23	2.88	2.94	2.98
	(0.28)	(0.29)	(0.29)
R25	3.76	3.82	3.81
	(0.29)	(0.29)	(0.29)
R27	4.28	4.34	4.37
	(0.29)	(0.30)	(0.30)
R29	4.27	4.33	4.38
	(0.31)	(0.32)	(0.32)
R32	4.66	4.77	4.79
	(0.36	(0.37)	(0.39)
Generic parameters			
Tail dummy	–0.69	–0.70	–0.72
	(0.28)	(0.28)	(0.28)
Price24 dummy	–0.41	–0.41	–0.44
	(0.11)	(0.11)	(0.11)
Standard deviation parameters			
R23-R32	2.58	2.60	2.57
	(0.44)	(0.44)	(0.48)
NOT	7.38	4.95	4.88
	(0.50)	(0.34)	(0.32)
Scaling factor SC	1.23	1.18	1.14
Summary statistics			
# Participants RC/SC	114/114	99/108	99/99
Log likelihood Joint	–2818.68	–2709.86	–2649.28
Log likelihood RC only	–1895.60	–1816.81	–1816.81
Log likelihood SC only	–916.85	–887.60	–826.55

Notes: Estimated with Nlogit 3.0. 250 Halton draws.
[a]All 114 fish-eating participants. [b]RC data excluding the 15 participants who chose the NOT alternative in all 20 RC scenarios. SC data excluding the six participants who chose the NOT alternative in all 10 SC scenarios. [c]RC and SC data excluding the 15 participants who chose the NOT alternative in all 20 RC scenarios. [d]Price in NOK 100.

Tabel 6. Number of none-of-these choices made by the participants

# NOT Choices	RC Scenario 1 to 10 Frequency	RC Scenario 1 to 10 Percent	RC Scenario 11 to 20 Frequency	RC Scenario 11 to 20 Percent	SC Scenario 1 to 10 Frequency	SC Scenario 1 to 10 Percent
0	42	36.84	35	30.70	58	50.87
1	4	3.51	13	11.40	14	12.28
2	8	7.02	11	9.65	12	10.53
3	12	10.53	11	9.65	8	7.02
4	7	6.14	4	3.51	7	6.14
5	8	7.02	6	5.26	2	1.75
6	2	1.75	6	5.26	4	3.51
7	8	7.02	4	3.51	2	1.75
8	10	8.77	5	4.39	1	0.88
9	4	3.51	3	2.63	0	0.00
10	18	15.79	16	14.04	6	5.26
Total	114	100.00	114	100.00	114	100.00

Table 7. Cross-table for frequency of none-of-these choices in SC and RC

	# RC									
# SC	0-2	3-5	6-8	9-11	12-14	15-17	18-20	Total	Mean[a]	Std.err
0	30	6	5	3	3	2	9	58	5.93	7.35
1	6	5	1	0	2	0	0	14	3.86	4.40
2	4	4	3	1	0	0	0	12	4.50	3.12
3	1	2	1	2	0	1	1	8	8.87	6.13
4	1	1	2	1	2	0	0	7	7.69	4.17
5	0	0	1	0	0	0	1	2	12.00	6.02
6	0	0	0	1	0	2	1	4	14.25	3.27
7	0	0	1	0	0	1	0	2	11.50	3.51
8	0	0	0	0	0	0	1	1	20.00	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	6	6	20.00	0
Total	42	18	14	8	7	6	19	114		
Mean[b]	0.50	1.28	2.14	2.27	1.42	3.67	4.32			
Std.err	0.93	1.19	2.13	2.04	1.67	1.67	4.47			

[a]Mean number of non-of-these choices in RC experiment.
[b]Mean number of non-of-these choices in SC experiment.