If you have to eat it, will you tell me the truth?

Darla Hatton MacDonald, John M. Rose, Haidee J. Lease and David N. Cox

Abstract (143 words, limit of 150 words for intended journal)

A vexing problem in the design of discrete choice experiments is how to create incentives for participants to reveal their true preferences. In this study, methods for introducing incentive compatibility are demonstrated as part of a food experiment where the novel contribution is the use of tasting across two samples. The 16 choice sets consisted of the attributes (levels): price (8), meat quality (2), fat content (3) and how recycled water was used in the production process (3 plus reference of tap water). Participants were randomly assigned to group 1 and 2 and halfway through each experiment, either told they would be eating random selections from their next 8 choices or asked to taste the product. Model results reveal preference and scale differences between the pre-and post intervention choice responses as well as between both intervention datasets suggesting anticipation is different from experience.

Word count of main body of paper: 8,089 words excluding tables and figures

Number of Tables: 5 (~1250 words)

Number of Figures: 2 (~500 words)

Total: 9,839 words

1. Introduction

For many urban areas, water supplies are becoming increasingly scarce. In the Australian context, physical scarcity of water during the Millennium drought (1997-2009) led to urban water utilities exploring the acceptability of desalination and different forms of water recycling such as wastewater and stormwater recycling. For food processors, closed loop production processes involving water recycling has the potential for reducing costs in two ways: by diverting wastewater to recycling and avoiding the disposal of trade waste (and the fee) as well as the potential reduction in input costs. However, consumer acceptability of water recycling in food production is unknown. Evidence from wastewater and stormwater recycling in urban water settings indicates that communities may support the idea of wastewater and stormwater recycling conceptually but projects have become derailed in the actual implementation (Marks et al. 2006; Nancarrow et al. 2008). For this reason, true consumer acceptability remains a major hurdle. Consumers are very conscious of food safety, technology in food production and have varying perceptions of risk (Sparks and Shephard 2006; Dosman et al. 2002). Further, the oral ingestion of substances that may cause illness may be a deeply ingrained evolutionary response (Curtis et al. 2004). Further once the association between a technology/product and a visceral "disgust" response, it is very hard to shake as the recent case of lean finely textured beef demonstrated in the United States.

In this paper, we use the opportunity presented by a food experiment to devise incentives through tasting to encourage participants to reveal their true preferences concerning non-price attributes. We build on the existing literature by exploring the effect of a "mostly" binding threat[1] of having to taste a product versus the actual experience of tasting on the stated preferences expressed by participants. We contribute to the literature on consequentiality and binding constraints in stated choice (SC) experiments by comparing a binding constraint and the actual experience of a commodity. The within-subject design of the overall experiment tests whether a binding constraint or experience affects participant's preferences relative to when they believe that no such binding constraint exists (that is, prior to actually experiencing the product). In the experiment reported herein, participants are presented with eight choice tasks, after which they are assigned to one of two experimental groups where they complete eight additional tasks.

In section 2, we place our experiment in the context of an interdisciplinary historical literature. The survey and sampling plan are outlined in section 3, after which the modelling approach used is discussed in section 4, Section 5 provides the model results after which discussion is given and conclusions drawn in Section 6.

2. The Literature

The use of stated choice (SC) techniques to gather data to model preferences is not a new phenomenon. The first SC experiment is thought to have been conducted by Thurstone (1931) who, using a crude form of experimental design, estimated indifference curves by asking a single participant to make choices between different combinations of coats, hats and shoes. Despite reported success, this early work was derided by many economists, chiefly by Wallis and Friedman (1942), who criticized the methodology on the basis of a lack of realism which they argued was likely to give rise to spurious results, stating:

"*[f]or a satisfactory experiment it is essential that the subject give actual reactions to actual stimuli... Questionnaires or other devices based on conjectural responses to hypothetical stimuli do not satisfy this requirement. The responses are valueless because the subject cannot know how he would react*."

In response, Rousseas and Hart (1951) undertook a study where participants were asked to make a single choice (subsequently repeated a month later) from a breakfast menu and participants were obliged

[1] Ethics approval required that participants could refuse to eat the product.

to eat what they had chosen. The experimental approach of Rousseas and Hart (1951) required participants to actually experience their choices was quickly adopted by other researchers, most notably by MacCrimmon and Toda (1969) who had participants trade-off gifted money for French pastries.

Despite the work of Rousseas and Hart (1951) and others working in behavioural economics and the then nascent field of mathematical psychology, the same criticism remains with respect to the use of SC methods sixty years later (e.g., List 2001; Camerer and Hogarth 1999; Diamond and Hausman 1994). Recent attempts to make SC choice tasks more realistic and less prone to hypothetical bias have taken many forms. Making SC choice tasks more incentive-compatible (e.g., Lusk et al. 2008; Ding 2007; Alfnes et al. 2006) and individual customisation of SC choice tasks to participant specific experiences (e.g., Rose et al. 2008; Train and Wilson 2008) represent just two approaches researchers employ. It is with the former approach that this paper is concerned.

Encapsulated by the concept of incentive compatibility, researchers have been concerned with detecting and potentially measuring whether participants act in a strategic manner rather than reveal their true preferences when answering stated preference surveys (Bateman et al. 2008; Carson and Groves 2007). Rather than abandon SC experiments, researchers have explored alternative strategies to either encourage participants to act as they would in real markets, or minimise any biases that may arise if they do not make choices that reflect their true preferences. These attempts may be categorised into pre- and post-data collection strategies.

Pre-data collection strategies include innovations such as the introduction of methods such as information acceleration (Urban et al. 1997) where researchers create a choice environment that mimics the context in which future consumption will be made, the use of cheap talk (Ladenburg et al. 2010; Carlsson et al. 2005; Champ et al. 2004) to emphasise the importance of the answers participant give and how these should reflect their true preferences, the use of certainty scales (e.g., Ready et al. 2010; Garcia et al. 2008; Norwood 2005) designed to ask participants to reflect on the choices they make, or giving each participant only a solitary binary choice task involving the choice between a status quo and a single hypothetical alternative (Farquharson 1969) or informing participants that only one choice task out of the sequence will be randomly selected and used for modelling purposes (McNair et al. 2011; Carson and Groves 2007) so that they believe that their specific choices matter. More recently, although not a SC experiment, Jacquemet et al. (2013) applied social psychology theories on the oath-taking as a truth-telling-commitment device to assist in mitigating hypothetical bias in an auction-based stated preference study. Other methods have focused on increasing the consequentiality of the choices participants make when undertaking SC experiments (Herriges et al. 2010; Vossler and Evans 2009) by ensuring that participants face some real outcome or consequence from the choices they make. Typically such studies involve participants being informed that at least one of the choices they make whilst undertaking a SC experiment will be selected at random and be binding (see e.g., Carlsson et al. 2010; Moser et al. 2010; Ding 2007; List et al. 2006; Alfnes et al. 2006; Lusk and Schroeder 2004).

Post-data collection methods have tended to focus on improving the econometric modelling of discrete choices, which may require additional data be collected during the survey task. For example, the use of *a priori* information on acceptable alternatives or attribute levels such as questions on attribute level thresholds have been used in modelling discrete choices (see e.g., Cantollo and Ortuzar 2006 and Swait 2001) whilst *ex post* data such based on questions as to what attributes where not considered or ignored during the SC survey have also been used (see e.g., Hensher et al. 2005; Rose et al. 2005). Other researchers have applied econometric models such as the latent class model in an attempt to better recover the true choice behaviour of participants without having to resort to asking participants what they did during the survey (see e.g., Campbell et al., 2011; Hess and Rose 2007).

3. Survey design, Sampling and Data Structure

The sampling plan and experimental design methodology

Participants were recruited to participate in this in-person experiment using by an accredited market research company. Participants were 18 to 65, not pregnant, without a food allergy, the main shopper in their household and a regular consumer of beef products. Participants received $40 as compensation for their time and expenses. Up to seven participants at a time were brought into the laboratory for three sessions per day spaced over a four week period in July-August 2012. The main elements of the experiment were initially explained to participants in a conference room. Participants were told by the lab facilitator (and author HJL), using a detailed script, there would be an initial set of questionnaires, an opportunity to taste some beef products and some follow-on questions. Participants were not informed about recycled water in the meatballs prior to entering the sensory booth, nor were they informed about when or how tasting would occur. Once signed consent forms were collected, participants were escorted to tasting booths with privacy shields. Participants were asked to not engage in any discussion with other participants: however, l66ab facilitator answered participants' queries, reminded participants to fill out questionnaires completely, collect questionnaires, etc.

In the individual sensory booth, participants were given their first eight choice sets in a single stapled booklet. Definitions of all the attributes in the choice experiment were provided on a separate summary sheet. In group 1, after eight choice sets, the questionnaire was collected. This is the intervention point where the participants were told they would be asked to taste four meatballs randomly selected from their next eight choices and the second questionnaire was distributed. Upon completion, the questionnaire was passed through a one-way slot to the kitchen where a generic choice set form was filled out corresponding to choices 13, 14, 15 and 16. A cooked meatball[2] was placed in a sample cup along with the filled out form, placed on a tray and passed back through the one-way slot back to the participant (double-blind) and repeated for the last three choice sets. In group 2, after eight choice sets, the questionnaire was collected and the participants were asked to taste four meatballs with the different levels of recycled water (tap water, recycled water used to clean the floors, recycled water used to clean the equipment, recycled water used as an ingredient). All meatballs used in the experiment were from the same recipe and cooked in exactly the same way.

The sample

The final sample consisted of 203 participants, of whom 102 were allocated to the first intervention group and 101 to the second. Socio-demographic characteristics of the final sample are provided in Table 1. The entire sample is similar to the gender proportions, household size of the Greater Adelaide metropolitan area but is statistically different at the five percent level on education and income (Australian Bureau of Statistic 2008).

[2] To avoid confounding effects in the experiment, the same meatballs were served each and every time to participants based on insights from two pre-tests. In the first pre-test, three fat levels in the meatballs were used. Participants were able to accurately identify the extra lean meatballs due to problems with the meatball drying out as it reached safe temperatures. Participants could not differentiate regular and lean fat levels. In pre-testing, 49 participants recruited randomly on a university campus were asked to taste one meatball (with different purported recycled water content) and asked to fill out a hedonic scale of emotional responses and then taste another meatball. While these meatballs were all identical, participants reported differences in taste that were not there. To avoid confounding effects in the main experiment, the same meatballs were served each and every time to participants.

Table 1: Socio-demographic characteristics of the sample

	Entire Sample (*n* = 203) *N* (%)	Group 1 (*n* = 102) *N* (%)	Group 2 (*n* = 101) *N* (%)
Gender			
Female	51.2%	61%	42.6%
Mean Age (years)			
Age	42.25	42.58	41.91
Education			
Did not complete high school	8.4%	6.9%	9.9%
Completed high school	21.7%	19.6%	23.8%
Certificate or Diploma	38.4%	36.3%	40.6%
Bachelor degree or higher	31.5%	37.3%	25.7%
Household income			
$1 – 33,799	18.7%	14.7%	22.8%
$33,800 – 62,399	29.1%	30.4%	27.7%
$62,400 – 103,999	28.1%	28.4%	27.7%
$104,000 – 155,999	17.3%	18.6%	15.8%
$156,000 +	5.9%	6.9%	5.0%
Number of children under 16 in household			
0 children	66.5%	72.5%	60.4%
1 child	12.8%	11.8%	13.9%
2 or more children	20.7%	15.7%	25.7%

The stated choice experiment

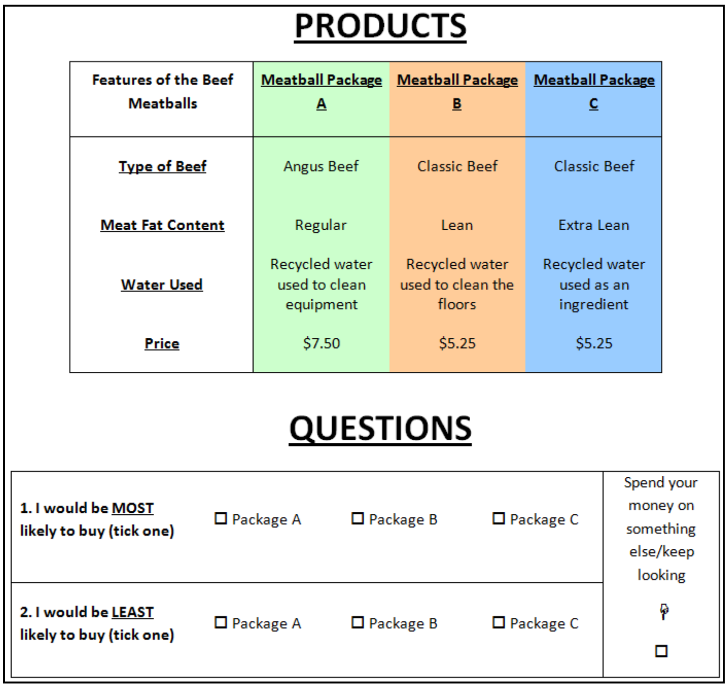
A paper-based survey was employed where participants were asked to review three hypothetical meatball alternatives as part of a Stated Choice (SC) experiment. The alternatives in each survey task were described by four attributes: type of beef, meat fat content, water type used in preparing the meatball, and price. Each of the four attributes was then further described by two or more attribute levels, the values of which are detailed in Table 2.

Table 2: Attributes and attribute levels

400 g packet of meatballs		
Type of Beef	2 levels	Classic Beef [1] or Angus Beef
Meat Fat Content	3 levels	Regular (15% fat) [2], Lean (10%) or Extra Lean (5%)
Water used	4 levels	Tap Water Used throughout, Recycled water used to clean the floors, Recycled water used to clean equipment, or Recycled water used as an ingredient
Price	8 levels	$3.00, $3.75, $4.50, $5.25, $6.00, $6.75, $7.50, $8.25

[1] Classic beef is all other breeds of cattle and not a premium product – it is not a brand as such. Angus beef is a breed of cow generally associated with higher quality.

The response mechanism used was a best-worst case three scenario (see Louviere et al. *forthcoming*) with the addition of a no choice alternative. That is, based on the attribute levels of the alternatives, participants were asked to select the packages they liked the most and least from the three presented to them, or select a no choice alternative. Given that the survey instrument was a paper and pencil questionnaire, the order that the best is chosen or the worst is chosen was not restricted. An example choice set is shown in Figure 1.



PRODUCTS

Features of the Beef Meatballs	Meatball Package A	Meatball Package B	Meatball Package C
Type of Beef	Angus Beef	Classic Beef	Classic Beef
Meat Fat Content	Regular	Lean	Extra Lean
Water Used	Recycled water used to clean equipment	Recycled water used to clean the floors	Recycled water used as an ingredient
Price	$7.50	$5.25	$5.25

QUESTIONS

1. I would be MOST likely to buy (tick one)	☐ Package A	☐ Package B	☐ Package C	Spend your money on something else/keep looking ☐
2. I would be LEAST likely to buy (tick one)	☐ Package A	☐ Package B	☐ Package C	

Figure 1: An example of a choice set

The design of the overall food experiment is presented in Figure 2. As such, participants were observed to make up to 32 choices depending on how many times they selected the no choice option across the 16 choice sets.

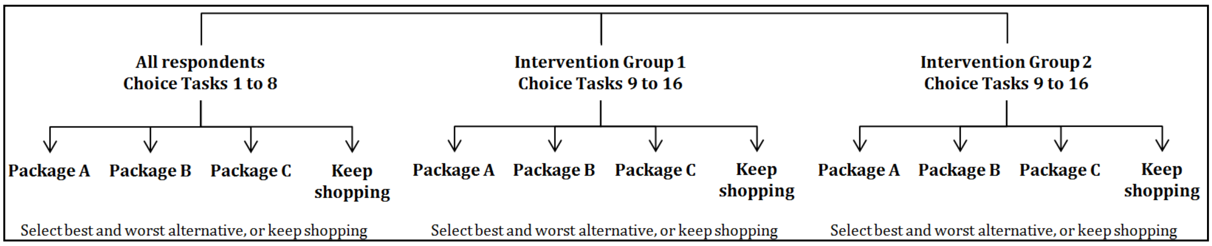


Figure 2: Experimental design structure

The stated choice experimental design

The experimental design underlying an SC experiment may play an important role in determining the final results of the study. Exactly how analysts distribute the levels of the design attributes over the course of an experiment, as determined by the underlying experimental design, may play a big part in whether or not an independent assessment of each attributes contribution to the choices observed to have been made by sampled participants can be determined. Further, the allocation of the attribute levels within the experimental design may also impact upon the statistical power of the experiment insofar as its ability to detect statistical relationships that may exist within the data. Given a set of attributes and attribute levels, the problem for the analyst is thus how best to allocate those levels over the course of the experiment.

For the present study, an *efficient* design was generated and used. Given a set of attributes and attribute levels, *efficient* designs are constructed such that the levels are allocated to the design in such a way that the elements (or subsets thereof) of the variance-covariance (VC) matrix are expected to be minimised once data is collected. Rather than work with the elements in the VC matrix directly, the literature suggests working with different measures that summarise the values that populate the VC matrix. In order to calculate the VC matrix for a design, the analyst must first assume a set of prior parameter estimates. If these are not known with certainty (as would typically be expected), the analyst may use prior parameter estimates drawn from Bayesian distributions and calculate the Bayesian D-error statistic, D_b-error, which is represented as

$$E_\beta\left[\det\left(I(\beta)^{-1}\right)^{\frac{1}{k}}\right] = \int_{\mathbb{R}^k} \det\left(I(\beta)^{-1}\right)^{\frac{1}{k}}. \qquad (1)$$

which is the expected value of the determinant of the inverse of the Fisher Information matrix, *I*, calculated for a design given a particular econometric model form and certain parameter estimates, scaled by one over the number of parameters, *k*.

To generate a D-*efficient* design, whether Bayesian parameter priors are assumed or not, different attribute level allocations are tested, with attribute level combinations that produce lower D-error values, representing more statistically *efficient* designs. Such designs are expected to produce data that will maximise the *t*-ratios for the design parameters (for further discussion on the generation of such designs see Scarpa and Rose 2008).

For the present study, a single *Bayesian efficient* design was generated consisting of 16 choice tasks blocked into two blocks of eight questions. Parameter priors were obtained from a survey of the stated preference literature on meat choice. The design was optimised assuming participants answered the best and then worst choice task in that order using a rank explosion procedure (see e.g., Vermeulen et al. 2011), assuming an MNL model specification. Constraints were placed on the attribute level combinations throughout the design so that prices greater than $5.25 were associated only with Angus premium beef. During the survey, participants were randomly allocated to one of the two blocks and completed all eight choice tasks in that block. After the intervention, participants then completed the eight choice tasks from the second block. All participants completed the 16 choice tasks from the same design; however participants were randomly assigned to one of four different versions of the paper and pencil questionnaire in which the order of the choice sets differed.

Data and data set-up

The final sample consisted of 203 participants, provided a total of 6,496 choice observations (203×16 = 3,248 most preferred and 203×16 = 3,248 least preferred choices). Of the 6,496 choice observations, 140 had no recorded choice (in all cases, missing observations were for an entire best/worst task rather than participants not answering their least preferred when they provided their most preferred alternative, or vice versa), leaving 6,356 observed choices from which to model. Of the remaining 6,356 choice observations, the no choice option was chosen only 70 times, which in setting up the data, was applied to both the most preferred and least preferred choice tasks.

The final choice data set-up assumed that all four alternatives (three hypothetical plus the no choice) were available in the least preferred task (Marley and Louviere 2005). This is in contrast to other previous studies whereby the most preferred alternative is removed from the least preferred choice task (see e.g., Collins and Rose 2011 or Scarpa et al. 2011). The interested reader is referred to Marley and Islam (2012) for an in-depth discussion of the properties of these two different data set-up assumptions.

4. Model Specification

Model formulation

The structure of the experiment conducted for the present study brings about a number of unique modelling challenges. Firstly, the experimental conditions require that the data be treated as three separate datasets; the first dataset consists of the first eight choice tasks completed by the entire sample of participants. Given that the two groups of participants were unaware of any of the specific intervention conditions during the first eight choice tasks, it is assumed that no preference or scale differences exist between the two groups for the first eight choices (at least as a result of the intervention) and hence the data for the first eight choice tasks obtained can be naively pooled for purposes of modelling. No such assumption can be made however with regards to potential preference or scale differences post-intervention given that each group was exposed to a different intervention condition. As such, the two other datasets consist of choice observations 9-16 for groups 1 and 2. The assumption that the data should be treated as three different datasets suggests that direct comparison of most model outputs obtained separately from models estimated independently from each treatment condition will not generally be possible given possible differences in scale. Likewise, simple comparisons of the log-likelihood functions and other model fit statistics are not possible given the non-nested nature of the three datasets. Given the above, it is necessary to estimate models for each treatment condition simultaneously whilst allowing for tests of possible scale and preference differences. The most common approach to modelling multiple datasets is to use a nested logit (NL) model.

Secondly, unlike most data, SC data typically involves the collation of multiple observations from each participant, albeit during a single session. Failure to properly account for the pseudo-panel nature of the data in the econometric modelling will at best affect only the standard errors of the model (and hence tests of parameter statistical significance) and at worst the parameter estimates themselves (see Hess and Rose 2009). Unfortunately, the NL model fails to account for this aspect of SC data. Hensher et al. (2008) proposed using a panel version of the error component (EC) model to approximate the nesting structure of the NL model, whilst at the same time also accounting for the pseudo-panel nature of the data. The EC model however assumes heteroskedastic error terms across subsets of alternatives within a dataset resulting from the need to normalise at least one EC for one alternative to be zero. This restriction regrettably requires that at least one alternative be treated in a separate nest to other alternatives within a dataset for purposes of model identification. For example, in the current context, for a given treatment condition, the analyst may assume a specification with an EC associated with the three hypothetical alternatives whilst the *keep shopping* or no choice alternative has no associated error component. In such a case, the model structure suggests that any differences in error variance are between the hypothetical alternatives and the no choice alternative. Some normalisation is also required within the model specification of the other treatment conditions when combining datasets. Assuming the no choice alternative is chosen for this normalisation, then the overarching model structure is one in which the error variances for the no choice alternatives for each datasets are constrained to be equal to zero, and empirically different to the error variances of the hypothetical alternatives. As such, the model will account for differences between datasets in terms of the error variances for the hypothetical alternatives whilst constraining the error variances of the no choice alternatives to be the same. This is different to assuming that error variance differences are dataset specific.

As noted by Scarpa et al. (2005) in the presence of a no choice or status quo alternative however, there exists the possibility that participants may treat that alternative systematically different to other alternatives present within the choice task. Systematic differences may arise as a result of (i) status quo or no choice alternative being more familiar to them as real world option relative to other hypothetical alternatives present within a choice task, and (ii) the fact that a no choice or status quo alternative is typically held constant across choice tasks, and hence unchanging, whereas the remaining alternatives are vary by way of the underlying experimental design. To further complicate matters, it is theoretically possible for the hypothetical alternatives of a SC experiments to be more highly correlated with each other than with a no choice alternative. This correlation has traditionally been captured via the inclusion of EC which are shared across the non-no choice alternatives in the utility specification of the model, but is absent from the utility function of the no choice alternative.

As such, the challenge is to allow for dataset specific scale differences, whilst at the same time allowing for heteroskedastic error between the hypothetical and no choice alternatives within each dataset (see Figure 3). As shown in Figure 3, the no choice alternatives remain within the dataset specific nests for purposes of testing, however they are allowed to have different error variances than the remaining alternatives in each choice task.

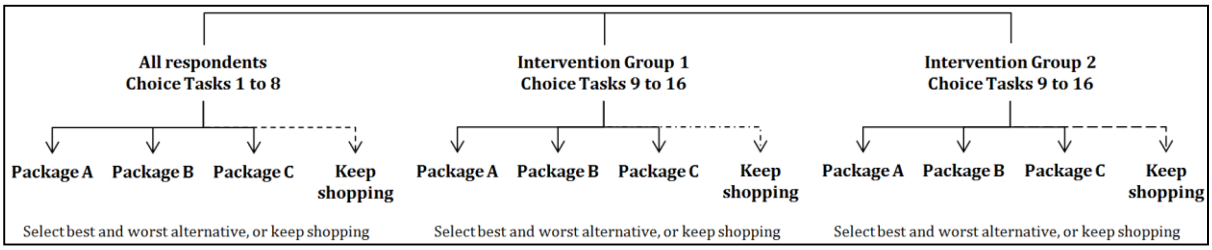


Figure 3: Final model nesting structure

Model Estimation

In order to understand the model better, let $U_{nsj|d.}$ denote the utility of alternative *j* obtained by participant *n* in choice situation *s*, in dataset *d*. =1, 2 or 3, where d_1 represents the first eight choice tasks for all participants, and d_2 and d_3 are the utility specifications for the group 1 and group 2 for choice tasks nine to 16 respectively. In order to be able to identify potential scale differences, it is necessary to constrain one or more preference parameters to be generic across all datasets. As is common practice, utility is assumed to be described by a linear relationship of observed attribute levels of each alternative, $x_{nsj|d.}$ and $z_{nsj|d.}$, and their corresponding weights (parameters), $\beta_{d.}$ and θ. Under this specification, θ represents a vector of parameters which are treated as being generic across each nest within the overall model structure, whilst $\beta_{d.}$ represent a vector of dataset specific parameters. Alternative specific constants (ASCs), $\alpha_{j|d.}$, are estimated for all no choice alternatives and are allowed to vary across the three datasets. In order to account for potential heteroskedastic error between the hypothetical and no choice alternatives, dataset specific EC, $\eta_{n|d.}$ are estimated for the three non-no choice alternatives. As is common practice, the e error components, $\eta_{n|d.}$ are treated as Normally distributed random parameters with means fixed at zero. Finally, the unobserved component of each utility, $\varepsilon_{nsj|d.}$, is assumed to be independently and identically extreme value type 1 (EV1) distributed. The model specification used for the current paper is shown in Equations 2(a) to 2(c).

$$U_{nsj|d_1} = \begin{cases} \alpha_{j|d_1} + \beta_{d_1} x_{nsj|d_1} + \theta z_{nsj|d_1} + \eta_{n|d_1} + \varepsilon_{nsj|d_1}, \forall j \neq no\ choice \\ \varepsilon_{ns,no\ choice|d_1} \qquad , j = no\ choice \end{cases} \tag{2a}$$

$$U_{njs|d_2} = \begin{cases} \exp\left(\lambda_{d_2}\right)\left(\alpha_{j|d_2} + \beta_{d_2} x_{nsj|d_2} + \theta z_{nsj|d_2} + \eta_{d_2} + \varepsilon_{nsj|d_2}\right), \forall j \neq no\ choice \\ \exp\left(\lambda_{d_2}\right)\left(\varepsilon_{ns,no\ choice|d_2}\right) \qquad , j = no\ choice \end{cases} \tag{2b}$$

$$U_{njs|d_3} = \begin{cases} \exp\left(\lambda_{d_3}\right)\left(\alpha_{j|d_3} + \beta_{d_3} x_{nsj|d_3} + \theta z_{nsj|d_3} + \eta_{d_3} + \varepsilon_{nsj|d_3}\right), \forall j \neq no\ choice \\ \exp\left(\lambda_{d_3}\right)\left(\varepsilon_{ns,no\ choice|d_3}\right) \qquad , j = no\ choice \end{cases} \tag{2c}$$

In order to account for potential error variance differences, dataset specific scale parameters $\lambda_{d.}$ are estimated for datasets *d*. = 2 and 3. By taking the exponentials of the scale parameters in model estimation, scale is ensured to be positive and hence consistent with random utility theory. By not estimating a scale parameter for *d*. = 1, the remaining scale parameters are estimated relative to the dataset obtained from the first eight choice tasks for all participants.

Within the model, only the EC are assumed to be randomly distributed. Unlike other models which assume random scale (e.g., the scaled MNL model; see Breffle and Morey 2000 or Fiebig et al. 2010) we treat scale using fixed parameters, according to Hess and Rose (2012). In order to avoid issues of preference and scale confoundment, we also treat the remaining preference parameters as fixed parameters (Hess and Rose 2012).

Assuming that unobserved components of utility are EV1 IID, the probability, $P_{nsj|d.}$, that participant n chooses alternative j in choice situation s is

$$P_{nsj|d.} = \int_{-\infty}^{\infty} \frac{\exp\left(\exp\left(\lambda_{d.}\right)V_{nsj|d.}\right)}{\sum_{i \in J_{ns|d.}} \exp\left(\exp\left(\lambda_{d.}\right)V_{nsi|d.}\right)} \phi\left(\sigma^2_{\eta|d.}\right) d\eta_{d.} \qquad (3)$$

where $V_{nsj|d.}$ is the modelled component of utility consisting of $\alpha_{j|d.}$, $\beta_{d.}$, θ, $x_{nsj|d.}$, $z_{nsj|d.}$ and $\eta_{n|d.} \sim N\left(0, \sigma^2_{\eta|d.}\right)$.

Let $y_{nsj|d.}$ equal one if alternative j is the chosen alternative in choice situation s shown to participant n, and zero otherwise. The joint probability for participant n making a sequence of choices is

$$P^*_{n|d.} = \prod_{s=1}^{S} \prod_{j=1}^{J} \left[P_{nsj|d.}\right]^{y_{nsj|d.}}. \qquad (4)$$

Unlike Equation (3) which represents the choice set specific probability, Equation (4) represents the probability that a particular sequence of alternatives will be observed for each participant n.

The parameters $\alpha_{j|d.}$, $\beta_{d.}$, θ, and $\lambda_{d.}$ are unknown and require estimation. Unfortunately, the integral in Equation (3) is mathematically intractable, and hence in order to estimate these parameters simulated maximum likelihood (SML) techniques are used. In this instance, SML utilises random draws to simulate the EC distributions to calculate the expected value of Equation (3) given $\alpha_{j|d.}$, $\beta_{d.}$, θ, $x_{nsj|d.}$, $z_{nsj|d.}$ and the distributional parameters of $\eta_{n|d.}$. The parameters can be estimated by maximizing the likelihood function

$$\log E(L) = \sum_{n=1}^{N} \log E\left(P^*_{n|d.}\right). \qquad (5)$$

5. Model results

Two models: a MNL model and an EC model assuming a panel specification were estimated using Python Biogeme (Bierlaire 2008). The EC model was estimated using 500 MLHS quasi Monte Carlo draws (Hess et al. 2005). In order to estimate the model, it is necessary to force at least one parameter to be generic across nests for purposes of identification. After extensive testing it was found that the best parameter for this was the price parameter. As such, only a single price parameter is reported for both the MNL and EC models in Table 3. For reasons of brevity, we limit our discussion to the EC model.

Beginning with an examination of the scale parameters, the scale parameter for group 1 is not statistically significantly different from zero suggesting that the error variance for participants operating under the assumption that their choices are binding is not different to participants who believe that their choices are not binding. The scale parameter for group 2 is statistically significant and positive (the scale parameter is estimated as an exponential and hence can be negative in this instance) suggesting that the scale for this group is greater than that for group 1 as well as for the combined choices of tasks 1-8. As such, participants who ate the meatballs prior to answering the second set of questions were found to have a lower error variance relative to the pre-intervention group as well as to participants who were told that they would be asked to consume their chosen alternative in a randomly selected future choice task.

Table 3: model results

Attribute	M1: MNL Par.	M1: MNL (Rob. *t*-rat.)	M2: Error component Par.	M2: Error component (Rob. *t*-rat.)
Homogenous attribute choice tasks 1 to 16 (Groups 1 and 2 combined)				
Price (S1-S8)	-0.139	(-14.18)	-0.147	(-8.06)
Choice tasks 1 to 8 (Groups 1 and 2 combined)				
ASC1 (S1-S8)	1.460	(17.66)	3.070	(13.57)
ASC2 (S1-S8)	1.580	(19.56)	3.190	(14.01)
ASC3 (S1-S8)	1.460	(17.92)	3.070	(13.41)
Beef (classic beef)# (S1-S8)	0.490	(10.84)	0.493	(8.35)
Fat (15% fat content)* (S1-S8)	-0.346	(-12.30)	-0.351	(-7.69)
Fat (10% fat content)* (S1-S8)	0.202	(5.60)	0.196	(5.06)
Water (tap water)Γ (S1-S8)	0.211	(6.35)	0.218	(4.45)
Water (recycled water/floor)Γ (S1-S8)	-0.175	(-5.05)	-0.175	(-2.87)
Water (recycled water/equip)Γ (S1-S8)	0.190	(4.99)	0.185	(3.65)
Error component (S1-S8)	-	-	2.110	(9.39)
Choice tasks 9 to 16 (Group 1)				
ASC1 (G1:S9-S16)	0.697	(6.77)	-0.919	(-3.71)
ASC2 (G1:S9-S16)	0.773	(7.61)	-0.841	(-3.41)
ASC3 (G1:S9-S16)	0.741	(7.28)	-0.876	(-3.53)
Beef (classic beef)# (G1:S9-S16)	0.471	(8.51)	0.469	(6.84)
Fat (15% fat content)* (G1:S9-S16)	-0.587	(-10.65)	-0.605	(-9.68)
Fat (10% fat content)* (G1:S9-S16)	0.295	(6.93)	0.289	(6.50)
Water (tap water)Γ (G1:S9-S16)	0.766	(10.95)	0.794	(8.84)
Water (recycled water/floor)Γ (G1:S9-S16)	-0.366	(-6.45)	-0.383	(-3.36)
Water (recycled water/equip)Γ (G1:S9-S16)	-0.079	(-1.62)	-0.092	(-1.66)
Scale (G1:S9-S16)	-0.030	(-0.49)	-0.036	(-0.74)
Error component (G1:S9-S16)	-	-	4.290	(10.94)
Choice tasks 9 to 16 (Group 2)				
ASC1 (G2:S9-S16)	0.621	(7.00)	-0.570	(-1.72)
ASC2 (G2:S9-S16)	0.741	(8.21)	-0.449	(-1.36)
ASC3 (G2:S9-S16)	0.592	(6.63)	-0.597	(-1.80)
Beef (classic beef)# (G2:S9-S16)	0.670	(13.42)	0.680	(8.73)
Fat (15% fat content)* (G2:S9-S16)	-0.396	(-9.37)	-0.397	(-5.89)
Fat (10% fat content)* (G2:S9-S16)	0.274	(7.66)	0.274	(7.69)
Water (tap water)Γ (G2:S9-S16)	0.411	(7.98)	0.405	(6.36)
Water (recycled water/floor)Γ (G2:S9-S16)	-0.044	(-0.99)	-0.035	(-0.48)
Water (recycled water/equip)Γ (G2:S9-S16)	0.083	(2.12)	0.086	(2.14)
Scale (G2:S9-S16)	0.128	(2.28)	0.134	(2.34)
Error component (G2:S9-S16)	-	-	4.110	(7.63)
Model fit				
LL(0)	-14635.231		-14635.231	
LL(β)	-13484.673		-11702.885	
ρ^2	0.079		0.200	
Adjusted ρ^2	0.074		0.196	
AIC (normalised)	4.253		3.693	
BIC (normalised)	4.284		3.728	
Number of parameters	30		33	
Sample information				
Number of participants	402		402	
Number of observations	6356		6356	

relative to premium; * relative to extra lean (fat content 5%); and Γ relative to recycled water used as an ingredient

Turning to the EC, for all three datasets the EC are statistically significant supporting the hypothesis expounded within the existing literature that there should exist a greater level of error variance for the hypothetical alternatives of a SC experiment than for a status quo alternative. A statistically significant EC also suggests that there exists a higher degree of substitution between the alternatives to which the

EC belongs, indicating that participants, irrespective of which data segment they belong to, are more likely to trade among the three hypothetical meatballs than between one of the meatballs and the status quo alternative. Table 4 presents the results of *t*-tests of statistical differences between the parameter estimates. As can be seen from this table, the EC for the three datasets are statistically different from one another after controlling for overall data-specific scale differences. This suggests that participants assigned to group 1 have the greatest degree of error variance, followed by those assigned to group 2, with the pre-assigned choices having the lowest error variance related to the hypothetical alternatives. Alternatively, these results can also be viewed as suggesting that for the combined data representing the first eight choice tasks of all participants, participants are more likely to trade-between the hypothetical choice alternatives than the no-choice alternative than they are when later assigned to either groups 1 or 2.

For choice tasks one to eight for all participants, the ASCs for the three hypothetical alternatives are all positive and statistically significant. This suggests that all else being equal, participants under the impression that their choices were not binding were more likely on average to select one of the hypothetical meatballs, than they were to select the no-choice alternative. Interestingly, the ASCs for choice tasks 9-16 for group 1 were all statistically significant but negative suggesting that, ceteris paribus, they were less inclined on average to select one of the hypothetical meatballs and choose the no-choice alternative when they believed that they would be asked to consume a meatball they had chosen from a randomly selected choice task. For the second intervention group who were asked to consume the meatballs before answering choice tasks 9-16, the ASCs were not statistically significant suggesting that on average, after controlling other factors, participants were equally inclined on average to choose one of the hypothetical meatball alternatives as they were in selecting the no-choice alternative.

Examining the design attributes, the model suggests that participants prefer classic beef as an ingredient over the use of premium beef, independent of which experimental condition they were assigned. Whilst this result appears to be somewhat spurious at first glance, it is likely to be an artefact of the experimental design as constraints were imposed on the design such that meatballs made of Angus beef were always associated with higher prices. As such, the preference for classic beef may also reflect in part a preference for lower prices. Whilst it would be tempting to incorporate a price-beef type interaction effect to counter this effect, such an interaction effect would reflect the premium beef-higher price confoundment as the price levels did not overlap between the two beef types. Examining the *t*-tests of statistical differences between the parameter estimates presented in Table 4, it can be seen that the marginal utility for the classic beef attribute level was not statistically different between experimental conditions 1 and 2, but is different between experimental condition 3 and experimental conditions 1 and 2. This suggests that those who were made to taste the meatballs prior to making their choices had a higher marginal utility for classic beef than those who had not tasted the meatballs, or those who were told that they would have to taste them based on their future choices.

For the fat content attribute, participants appear to have a preference for 10 percent fat content over 15 or five percent fat content, irrespective of the experimental condition they belong to. One possible explanation for this is that fatty foods may taste better but too much fat may be considered as being unhealthy. Examining the tests of statistical difference, participants were observed to become less predisposed towards meatballs with 15 percent fat content when their choices were thought to be binding. Participants who were asked to taste the meatballs prior to making their last eight choices experienced no change in preferences when compared to the choices they made prior to eating the meatballs. Both intervention groups experienced an increase in their marginal utility for meatballs with 10 percent fat content post-intervention, with the post-intervention preferences not differing between the two groups.

Table 4: Parameter differences

Coefficient1	Coefficient2	M1: MNL (Rob. *t*-rat.)	M1: MNL *p*-value	M2: Error component (Rob. *t*-rat.)	M2: Error component *p*-value
ASC1 (S1-S8)	ASC1 (G1:S9-S16)	(9.93)	0.00	(12.43)	0.00
ASC1 (S1-S8)	ASC1 (G2:S9-S16)	(12.73)	0.00	(8.29)	0.00
ASC2 (S1-S8)	ASC2 (G1:S9-S16)	(10.93)	0.00	(12.53)	0.00
ASC2 (S1-S8)	ASC2 (G2:S9-S16)	(12.73)	0.00	(8.26)	0.00
ASC3 (S1-S8)	ASC3 (G1:S9-S16)	(9.26)	0.00	(12.18)	0.00
ASC3 (S1-S8)	ASC3 (G2:S9-S16)	(12.68)	0.00	(8.30)	0.00
ASC1 (G1:S9-S16)	ASC1 (G2:S9-S16)	(0.93)	0.35	(-1.24)	0.22
ASC2 (G1:S9-S16)	ASC2 (G2:S9-S16)	(0.41)	0.68	(-1.40)	0.16
ASC3 (G1:S9-S16)	ASC3 (G2:S9-S16)	(1.80)	0.07	(-0.98)	0.33
Beef (classic beef)$^{\#}$ (S1-S8)	Beef (classic beef)$^{\#}$ (G1:S9-S16)	(0.30)	0.77	(0.33)	0.74
Beef (classic beef)$^{\#}$ (S1-S8)	Beef (classic beef)$^{\#}$ (G2:S9-S16)	(-3.45)	0.00	(-2.69)	0.01
Beef (classic beef)$^{\#}$ (G1:S9-S16)	Beef (classic beef)$^{\#}$ (G2:S9-S16)	(-3.00)	0.00	(-2.13)	0.03
Fat (15% fat content)* (S1-S8)	Fat (15% fat content)* (G1:S9-S16)	(3.94)	0.00	(3.89)	0.00
Fat (15% fat content)* (S1-S8)	Fat (15% fat content)* (G2:S9-S16)	(1.02)	0.31	(0.79)	0.43
Fat (15% fat content)* (G1:S9-S16)	Fat (15% fat content)* (G2:S9-S16)	(-3.02)	0.00	(-2.53)	0.01
Fat (10% fat content)* (S1-S8)	Fat (10% fat content)* (G1:S9-S16)	(-2.08)	0.04	(-2.07)	0.04
Fat (10% fat content)* (S1-S8)	Fat (10% fat content)* (G2:S9-S16)	(-1.79)	0.07	(-2.06)	0.04
Fat (10% fat content)* (G1:S9-S16)	Fat (10% fat content)* (G2:S9-S16)	(0.43)	0.66	(0.28)	0.78
Water (tap water)$^{\Gamma}$ (S1-S8)	Water (tap water)$^{\Gamma}$ (G1:S9-S16)	(-7.10)	0.00	(-7.00)	0.00
Water (tap water)$^{\Gamma}$ (S1-S8)	Water (tap water)$^{\Gamma}$ (G2:S9-S16)	(-3.28)	0.00	(-2.64)	0.01
Water (tap water)$^{\Gamma}$ (G1:S9-S16)	Water (tap water)$^{\Gamma}$ (G2:S9-S16)	(4.38)	0.00	(3.53)	0.00
Water (recycled water/floor)$^{\Gamma}$ (S1-S8)	Water (recycled water/floor)$^{\Gamma}$ (G1:S9-S16)	(2.98)	0.00	(2.07)	0.04
Water (recycled water/floor)$^{\Gamma}$ (S1-S8)	Water (recycled water/floor)$^{\Gamma}$ (G2:S9-S16)	(-2.37)	0.02	(-1.59)	0.11
Water (recycled water/floor)$^{\Gamma}$ (G1:S9-S16)	Water (recycled water/floor)$^{\Gamma}$ (G2:S9-S16)	(-6.06)	0.00	(-2.63)	0.01
Water (recycled water/equip)$^{\Gamma}$ (S1-S8)	Water (recycled water/equip)$^{\Gamma}$ (G1:S9-S16)	(-4.56)	0.00	(4.16)	0.00
Water (recycled water/equip)$^{\Gamma}$ (S1-S8)	Water (recycled water/equip)$^{\Gamma}$ (G2:S9-S16)	(2.05)	0.04	(1.61)	0.11
Water (recycled water/equip)$^{\Gamma}$ (G1:S9-S16)	Water (recycled water/equip)$^{\Gamma}$ (G2:S9-S16)	(-2.65)	0.01	(-2.68)	0.01
Scale (G1:S9-S16)	Scale (G2:S9-S16)	(-2.41)	0.02	(-2.60)	0.01
Error component (S1-S8)	Error component (G1:S9-S16)	-	-	(11.50)	0.00
Error component (S1-S8)	Error component (G2:S9-S16)	-	-	(-5.50)	0.00
Error component (G1:S9-S16)	Error component (G2:S9-S16)	-	-	(-10.09)	0.00

$^{\#}$ relative to premium; * relative to extra lean (fat content 5%); and $^{\Gamma}$ relative to recycled water used as an ingredient

Participants were observed to prefer tap water to be used throughout the entire cooking and cleaning process relative to the use of recycled water in any one part of the process. Relative to choice tasks 1-8, this effect was observed to become particularly strong with group 1 participants across choice tasks 9-16, but was also present, though to a lesser degree, for group 2 in choice tasks 9-16. Using recycled water to clean the floors was perceived negatively relative to using recycled water as an ingredient prior to either intervention occurring. This effect became more pronounced for participants group 1 choice tasks 9-16, but disappeared for participants in group 2. There also existed a statistically significant preference for using recycled water to clean kitchen equipment over using recycled water as an ingredient for all participants in choice tasks 1-8 (prior to intervention). For participants in group 1 in choice tasks 9-16, this difference was no longer observed. For participants assigned to group 2, participants were found once more to have a statistically significant preference for using recycled water to clean kitchen equipment relative to using recycled water as an ingredient, however the marginal utility for doing so was found to be statistically less post-intervention.

For completeness, we compute and report the marginal willingness to pay (MWTP) estimates for both models with the 95 percent confidence intervals calculated using the Delta method (Hole 2007). Purely for purposes of comparison, we do not take the absolute values of the MWTP estimates, instead letting them be either positive or negative. Examination of the results shows that for both the MNL and EC models, the MWTPs are mostly not different statistically across the three datasets. Only three differences are observed for both models. These are the MWTP for tap water for groups 1 and 2 relative to the pre-intervention data segment, and the MWTP for recycled water used to clean kitchen equipment between group 1 and the pre-intervention data segment.

Table 5: Marginal Willingness to Pay Estimates and Confidence Intervals

M1: MNL model

	Choice tasks 1 to 8 (Groups 1 and 2 combined)				*Choice tasks 9 to 16 (Group 1)*				*Choice tasks 9 to 16 (Group 2)*			
Attribute	MWTP	(rob. *t*-rat.)	Lower 95%	Upper 95%	MWTP	(rob. *t*-rat.)	Lower 95%	Upper 95%	MWTP	(rob. *t*-rat.)	Lower 95%	Upper 95%
Beef (classic beef)[#]	-\$3.53	(-11.27)	-\$4.14	-\$2.91	-\$3.39	(-8.01)	-\$4.22	-\$2.56	-\$4.82	(-12.27)	-\$5.59	-\$4.05
Fat (25% fat content)[*]	\$2.49	(9.38)	\$1.97	\$3.01	\$4.22	(2.47)	\$0.87	\$7.58	\$2.85	(7.62)	\$2.12	\$3.58
Fat (15% fat content)[*]	-\$1.45	(-5.49)	-\$1.97	-\$0.93	-\$2.12	(-2.33)	-\$3.91	-\$0.34	-\$1.97	(-7.12)	-\$2.51	-\$1.43
Water (tap water)[Γ]	-\$1.52	(-5.74)	-\$2.04	-\$1.00	-\$5.51	(-3.18)	-\$8.91	-\$2.11	-\$2.96	(-8.02)	-\$3.68	-\$2.23
Water (recycled water/floor)[Γ]	\$1.26	(4.90)	\$0.76	\$1.76	\$2.63	(1.90)	-\$0.08	\$5.34	\$0.32	(0.98)	-\$0.31	\$0.95
Water (recycled water/equip)[Γ]	-\$1.37	(-4.89)	-\$1.91	-\$0.82	\$0.57	(1.37)	-\$0.25	\$1.38	-\$0.60	(-2.08)	-\$1.16	-\$0.04
M2: EC model												
Attribute	MWTP	(rob. *t*-rat.)	Lower 95%	Upper 95%	MWTP	(rob. *t*-rat.)	Lower 95%	Upper 95%	MWTP	(rob. *t*-rat.)	Lower 95%	Upper 95%
Beef (classic beef)[#]	-\$3.35	(-6.76)	-\$4.33	-\$2.38	-\$3.19	(-5.43)	-\$4.34	-\$2.04	-\$4.63	(-6.27)	-\$6.07	-\$3.18
Fat (25% fat content)[*]	\$2.39	(5.72)	\$1.57	\$3.21	\$4.12	(7.62)	\$3.06	\$5.17	\$2.70	(4.51)	\$1.53	\$3.88
Fat (15% fat content)[*]	-\$1.33	(-5.76)	-\$1.79	-\$0.88	-\$1.97	(-4.20)	-\$2.88	-\$1.05	-\$1.86	(-6.67)	-\$2.41	-\$1.32
Water (tap water)[Γ]	-\$1.48	(-3.45)	-\$2.33	-\$0.64	-\$5.40	(-6.24)	-\$7.10	-\$3.71	-\$2.76	(-4.99)	-\$3.84	-\$1.67
Water (recycled water/floor)[Γ]	\$1.19	(2.83)	\$0.37	\$2.02	\$2.61	(3.06)	\$0.93	\$4.28	\$0.24	(0.48)	-\$0.73	\$1.21
Water (recycled water/equip)[Γ]	-\$1.26	(-3.95)	-\$1.88	-\$0.63	\$0.63	(1.60)	-\$0.14	\$1.40	-\$0.59	(-2.04)	-\$1.15	-\$0.02

[#] relative to premium; [*] relative to extra lean (fat content 5%); and [Γ] relative to recycled water used as an ingredient

6. Discussion and Conclusions

This paper adds to the growing research examining the issue of incentive compatibility and its effects on SC results. Unlike other studies in this area where participants are typically informed prior to commencing the study that once their choices are made, one choice will be selected at random to be binding, we developed an experiment that allows for a within-subject test of whether making choices more consequential changes underlying choice behaviour. Further, we add another dimension to the literature, by asking some participants to actually experience one or more of the alternatives prior answering a second set of SC questions.

Our results suggest that making choices binding in SC experiments can lead to changes in the observed choice behaviour of participants. The most telling change was observed in terms of the model ASCs which suggested that participants were much more open to hypothetically choosing to eat or purchase a meatball over not purchasing or eating one when the choices were considered not to be binding, but were much more likely to opt for the no-choice alternative when the choices where made binding, all else being equal. This suggests that participants may become more conservative in their preferences when confronted with the choice of a potentially distasteful product that they may have to consume compared to when they make choices that offer no real consequences.

Our findings also suggest that having participants experience or consume a product prior to taking part in a SC exercise may result in both scale and preference differences compared to choices made in hypothetical markets where the actual product has yet to be experienced. Further, our findings indicate that having participants who are asked to consume a product prior to partaking in a SC experiment may exhibit different choice behaviour to those who are assigned to the more traditional method of dealing with incentive compatibility, that being, make one or more of the choice tasks binding. In some respects, this finding is somewhat troubling, although it points to a potential future line of research enquiry. What this study suggests is that participants may be more likely to overstate the fact that they will choose a potentially distasteful product in a hypothetical market when their choices have no real cost to them relative to when their choices matter. However once a potentially unpleasant product is experienced and found not to be to objectionable, their preferences shift again. If such a finding is found to hold in general, then there exist consequences in terms of using SC experiments for the purposes of forecasting. It might be that making choices binding will better approximate markets in which products are new, or in which consumers have little experience, however in more mature markets, ensuring that participants experience the actual product, or at least some form of it, prior to undertaking the SC experiment might better reflect the choices they are likely to make in real markets. Unfortunately, in the current study, no real preference data was collected in terms of real market meatball choice involving meatballs prepared using different types of water; hence we are not able to state categorically whether either intervention condition is more or less likely to reduce potential hypothetical bias, or reflect different aspects of a market in terms of the level of product maturity. Independent of the above however, it is somewhat reassuring that we found relative few differences in the MWTP results in this current study.

Two further limitations to the current study are worth mentioning. Firstly, despite allowing for scale and preference differences between the different experimental conditions, we have retained the assumption that both preferences and scale are homogenous within experimental conditions, although we have allowed for heterogeneous error between the hypothetical and no-choice options. It is interesting to note how history tends to repeat itself, as the assumption of homogeneity of preferences was also made by Rousseas and Hart (1951) which was heavily criticised by MacCrimmon and Toda (1969). Whilst it would be possible to estimate random parameters to capture preference heterogeneity, we leave this to future research. A second limitation worth noting is that the CSIRO is a recognized and trusted scientific institution in Australia. The CSIRO is well known for Total Well-being dietary books (Noakes and Clifton, 2005) and as such, there may be a trust factor that facilitated the group 2 intervention that would not extend to a food manufacturer.

Acknowledgements: The project was funded by the Australian Water Recycling Centre of Excellence under the Commonwealth's Water for the Future Program and the CSIRO Water for a Healthy Country

Research Flagship. The collegial reviews by Rosalind Bark and Sorada Tapsuwan are also acknowledged. The assistance of Julie Syrette, Eva Calvaresi, Myrtille Rouer and Chloe Devauze is gratefully acknowledged. We wish to thank the 49 meatball testers at Waite campus and the three little meatball recipe testers who endured eight recipes over two weeks.

References

Alfnes, F., Guttormsen, A., Steine, G., Kolstad, K. (2006). Consumers' willingness to pay for the color of salmon: a choice experiment with real economic incentives. American Journal of Agricultural Economics, 88(4): 1050-1061.

Australian Bureau of Statistics (2008). Adelaide: A Social Atlas. ABS Catalogue No. 2030.4

Bierlaire, M. (2008). An introduction to BIOGEME Version 1.6. biogeme.epfl.ch.

Bateman, I.J., Carson, RT., Day, B., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R., Wang, P. (2008). Choice set awareness and ordering effects in discrete choice experiments. CSERGE Working Paper EDM 08-01.

Breffle, W.S., Morey, E.R. (2000). Investigating preference heterogeneity in a repeated discrete-choice recreation demand model of atlantic salmon fishing. Marine Resource Economics 15(1): 1-20.

Camerer, C.F., Hogarth, R.M. (1999). The effects of financial incentives in experiments: a review and capital labor-production framework. Journal of Risk Uncertainty 19(1–3):7–42

Campbell, D., Hensher, D.A., Scarpa, R. (2011). Non-attendance to attributes in environmental choice analysis: a latent class specification. Journal of Environmental Planning and Management 54(8): 1061-1076.

Cantollo, V., De D. Ortuzar, J. (2006). Implications of Thresholds in Discrete Choice Modelling. Transport Reviews Vol. 26(6): 667-691.

Carlsson, F., Daruvala, D., Jaldell, H. (2010). Do you do what you say or do you do what you say others do?. Journal of Choice Modelling 3(2): 113-133.

Carlsson, F., Frykblom, P., Lagerkvist, C.-J. (2005). Using cheap-talk as a test of validity in choice experiments, Economics Letters 89(2): 147-152.

Carson, RT., Groves, T. (2007). Incentive and informational properties of preference questions, Environmental and Resource Economics 37(1): 181-210.

Champ, P. A., Moore, R. Bishop, R.C. (2004). Hypothetical bias: The mitigating effects of certainty questions and cheap talk, Proceedings of Agricultural and Applied Economics Association, 2004 Annual meeting, Denver, CO, 4t^{h} August.

Collins, A.C., Rose, J.M. (2011). Estimation of stochastic scale with best-worst data, International Choice Modelling Conference, July 4-6, Leeds, United Kingdom.

Curtis, V., Aunger, R., Rabie, T. (2004). Evidence that disgust evolved to protect from risk of disease 10.1098/rsbl.2003.0144 Proceedings of the Royal Society, B, Biological Sciences 271(S4): S131-S133.

Diamond, P.A., Hausman, J.A. (1994). Contingent valuation: is some number better than no number. Journal of Economic Perspective 8(4): 45–64

Ding, M. (2007). An incentive aligned mechanism for conjoint analysis. Journal of Marketing Research 44(2): 214–223

Dosman, D. M., Adamowicz, W. L., Hrudey, S. E. (2001), Socioeconomic Determinants of Health- and Food Safety-Related Risk Perceptions. Risk Analysis 21: 307–318

Farquharson R. (1969). Theory of voting. Yale University Press, New Haven

Fiebig, D.G., Keane, M., Louviere, J.J., Wasi, N. (2010). The generalized multinomial logit: accounting for scale and coefficient heterogeneity. Marketing Science 29(3): 393-421

Garcia, D., Hoyos, D., Riera, P. (2008). Allowing for uncertainty in choice experiments, Proceedings of III AERNA Conference Spanish-Portuguese Association of Natural Resources and Environmental Economics, Palma, Mallorca, 6th June.

Hensher, D.A., Rose, J.M. Greene, W.H. (2008). Combining RP and SP data: biases in using the nested logit 'trick' – contrasts with flexible mixed logit incorporating panel and scale effects. Journal of Transport Geography 16(2): 126-133.

Hensher, D.A., Rose, J.M., Greene, W.H. (2005). The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes. Transportation 32(3): 203-222.

Herriges, J., Kling, C., Liu, C.-C., Tobias, J. (2010). What are the consequences of consequentiality?, Journal of Environmental Economics and Management 59(1): 67-81.

Hess, S., Train, K.E., Polak, J.W. (2005). On the use of a Modified Latin Hypercube Sampling (MLHS) approach in the estimation of a Mixed Logit model for vehicle choice. Transportation Research part B 40(2): 147-163.

Hess, S., Rose, J.M. (2009). Allowing for intra-respondent variations in coefficients estimated on stated preference data. Transportation Research Part B 43(6): 708-719.

Hess, S., Rose, J.M. (2012). Can scale and coefficient heterogeneity be separated in random coefficients models?, Transportation, 39(6): 1225-1239.

Hole, A. (2007). A comparison of approaches to estimating confidence intervals for willingness to pay measures, Health Economics 16(8): 827-40.

Jacquemet, N., Joule, R.-V., Luchinic, S., Shogren, J., (2013). Preference elicitation under oath. Journal of Environmental Economics and Management 65(1): 110–132.

Ladenburg, J., Bonnichsen, O. and Dahlgaard, J.O. (2011). Testing the effect of a short cheap talk script in choice experiments. Nationaloekonomisk Tidsskrift 149: 25-54.

List, J.A. (2001). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. American Economic Review 91(5): 1498–1507.

List, J., Sinha, P., Taylor, M. (2006). Using choice experiments to value non-market goods and services: evidence from field experiments. Advances in Economic Analysis and Policy 6(2): 1132-1132.

Louviere, J.L., Flynn, T.N., Marley, A.A.J. (forthcoming). Best-Worst Scaling: Theory and Methods, Handbook of Choice Modelling.

Lusk, J., Schroeder, T. (2004). Are choice experiments incentive compatible? a test with quality differentiated beef steaks. American Journal of Agricultural Economics 86(2): 467-482.

Lusk, J., Fields, D., Prevatt, W., (2008). An incentive compatible conjoint ranking mechanism. American Journal of Agricultural Economics 90(2): 487-498.

MacCrimmon, K.R. Toda, M. (1969). The experimental determination of indifference curves. Review of Economic Studies, 36(4), 433-451.

Marley, A.A.J., Louviere, J.J. (2005). Some probabilistic models of best, worst, and best–worst choices, Journal of Mathematical Psychology, 49(6), 464–480.

Marley, A.A.J., Islam, T. (2012). Conceptual relations between expanded rank data and models of the unexpanded rank data, Journal of Choice Modelling, 5(1), 38-80.

Marks, J.S., Martin, B., Zadoronznyi, M. (2006). Acceptance of Water recycling in Australia: national baseline data. Water 33(2): 151-157.

McNair, B., Bennett, J., Hensher, D.A. (2011). A comparison of responses to single and repeated discrete choice questions, Resource and Energy Economics, 33(3), 544-571.

Moser, R., Notaro, S., Raffaelli, R. (2010). Using your own money makes the difference: testing the hypothetical bias with a Real Choice Experiment, Proceedings of 2010 World Congress of Environmental and Resource Economics, Montreal, Canada, 2nd July.

Nancarrow B.E., Leviston, Z., Po M., Porter, N.B., Tucker, D.I. (2008). What drives communities' decisions and behaviours in the reuse of wastewater. Water Science and Technology 57(4): 485-91

Noakes, M., Clifton, P. (2005). The CSIRO Total Wellbeing Diet. Penguin, Australia

Norwood, F.B. (2005). Can calibration reconcile stated and observed preferences?, Journal of Agricultural and Applied Economics 37(1): 237-248

Ready, R.C., Champ, P.A., Lawton, J.L. (2010). Using respondent uncertainty to mitigate hypothetical bias in a stated choice experiment. Land Economics 86(2), 363-381.

Rose, J.M., Bliemer, M.C.J., Hensher, D.A., Collins, A.C. (2008). Designing efficient stated choice experiments involving respondent based reference alternatives. Transportation Research Part B 42(4), 395-406.

Rose, J.M., Hensher, D.A., Greene, W.H. (2005). Recovering costs through price and service differentiation: accounting for exogenous information on attribute processing strategies in airline choice. Journal of Air Transport Management 11(6): 400-407

Rousseas, S.W., Hart, A. G. (1951). Experimental verification of a composite indifference map, Journal of Political Economy 59(4): 288-318

Scarpa, R., Ferrini, S. Willis, K.G. (2005). Performance of error component models for status-quo effects in choice experiments, in Scarpa, R. and Alberini, A.(ed.) Applications of Simulation Methods in Environmental and Resource Economics, Dordrecht, the Netherlands: Springer 247–74.

Scarpa, R., Notaro, S., Raffelli, R., Pihlens D, Louviere, J.J. (2011). Exploring scale effects of best/worst rank ordered choice data to estimate visitor's benefits from alpine transhumance. American Journal of Agricultural Economics 93: 809-824.

Scarpa, R. and Rose, J.M. (2008). Designs efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. Australian Journal of Agricultural and Resource Economics 52(3): 253-282.

Sparks, P. and Shepherd, R. (1994). Public Perceptions of the Potential Hazards Associated with Food Production and Food Consumption: An Empirical Study. Risk Analysis 14: 799–806.

Swait, J. (2001). A non-compensatory choice model incorporating attribute cut-offs: Transportation Research Part B 35(10): 903–928.

Train, K. Wilson, W.W. (2008). Estimation on stated-preference experiments constructed from revealed preference choices. Transportation Research Part B 42(3): 191–203.

Thurstone, L. (1931). The indifference function. Journal of Social Psychology 2(2): 139-167.

Urban, G.L., Hauser, J.R., Qualls, W.J., Weinberg, B.D., Bohlmann, J.D., Chicos R.A. (1997). Information acceleration: validation and lessons from the field. Journal of Marketing Research 34(1): 143-53.

Vermeulen, B., Goos, P., Vandeboek, M. (2011). Rank-order choice-based conjoint experiments: Efficiency and design. Journal of Statistical Planning and Inference 141(8): 2519–2531.

Vossler, C.A. Evans, M.F. (2009). Bridging the gap between the field and the lab: Environmental goods, policy maker input, and consequentiality. Journal of Environmental Economics and Management 58(3): 338-345.

Wallis, W.A., Friedman, M. (1942). The empirical derivation of indifference functions, in Lange, O., McIntyre, F. and Yntema, T.O. (eds.) Studies in Mathematical Economics and Econometrics in memory of Henry Schultz, University of Chicago Press, Chicago, IL, 175-189.