DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

Can a Repeated Opt-Out Reminder remove hypothetical bias in discrete choice experiments?

An application to consumer valuation of novel food products

Mohammed H. Alemu
Søren B. Olsen

2017 / 05

IFRO Working Paper 2017 / 05

Can a Repeated Opt-Out Reminder remove hypothetical bias in discrete choice experiments? An application to consumer valuation of novel food products

Authors: Mohammed H. Alemu, Søren B. Olsen

JEL-classification: C12, C13, C83, C93, D12, 013, Q01, Q11, Q13, Q18

Published: May 2017

See the full series IFRO Working Paper here:
www.ifro.ku.dk/english/publications/foi_series/working_papers/

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk/english/

Can a Repeated Opt-Out Reminder remove hypothetical bias in discrete choice experiments? An application to consumer valuation of novel food products

Mohammed H. Alemu and Søren B. Olsen
Department of Food and Resource Economics, University of Copenhagen, Rolighedsvej 25, 1958, Frederiksberg C Denmark

Corresponding author: Mohammed H. Alemu, Phone: +45 35 33 14 92, E-mail: mha@ifro.ku.dk

Abstract

Recent papers have suggested that use of a so-called Repeated Opt-Out Reminder (ROOR) might mitigate hypothetical bias in stated Discrete Choice Experiments (DCE), but evidence so far has only been circumstantial. We provide the first comprehensive test of whether a ROOR can actually mitigate hypothetical bias in stated DCE. The data originates from a field experiment concerning consumer preferences for a novel food product made from cricket flour. Utilizing a between-subject design with three treatments, we find significantly higher marginal willingness to pay values in hypothetical than in nonhypothetical settings, confirming the usual presence of hypothetical bias. Comparing this to a hypothetical setting where the ROOR is introduced, we find that the ROOR effectively eliminates hypothetical bias for one attribute and significantly reduces it for the rest of the attributes. Our results further suggest that these reductions of hypothetical bias are brought about by a decrease in the tendency to ignore the price attribute.

Key words: Hypothetical bias, novel food, repeated opt-out reminder, willingness to pay

Introduction

Since its advent in the early 1980's in the marketing research, the stated discrete choice experiment (DCE) method has been used extensively in the environmental, agricultural, food, transport and marketing sectors to elicit the value of goods and services. However, critics have continually questioned the validity and reliability of the results obtained from hypothetical experiments including stated DCE. If results from hypothetical experiments are to be useful for testing theories or to guide policy decisions, such approaches should yield reliable results. Nevertheless, more often than not this seems to not be the case (List and Gallet 2001; Lusk and Schroeder 2004; Murphy et al. 2005; Ding, Grewal, and Liechty 2005), and one of the recurring issues tied to the stated DCE and stated preference methods in general is the issue of hypothetical bias. Due to the hypothetical nature of the choice questions asked, respondents may for various reasons overstate their Willingness-To-Pay (WTP) since there is no incentive compelling them to reveal their true value (e.g. Fox et al. 1996; Lusk and Schroeder 2004; Harrison 2006; Murphy et al. 2005; Vossler, Doyon, and Rondeau 2012; Vossler and Watson 2013).

There is a rich body of literature investigating hypothetical bias within the areas of environmental valuations employing the Contingent Valuation (CV) method to elicit the value of public goods (e.g. Carson et al. 1996; List and Shogren 1998; Cummings and Taylor 1999; Carlsson and Martinson 2001; List 2001; List and Gallet 2001; Little and Berrens 2004; Murphy et al. 2005; Blumenschein et al. 2008). Somewhat opposed to this, hypothetical bias has not been investigated in-depth in DCE (Hensher 2010) mainly due to the belief that stated DCEs are less likely to suffer from hypothetical bias than CV applications (Lusk and Schroeder 2004). Given that people may behave differently depending on the type of the good and the experimental setup (Carson et al. 1996; Carson and Groves 2007; Chang, Lusk, and Norwood 2009), the conclusions from external validity tests on CV studies based on public goods cannot be transferred directly to the DCE context nor to a private goods context. Hence, similar tests should be applied in DCE applications to thoroughly examine their validity. This is of particular importance as DCEs are extensively being used to assess demand for new products to guide policy development. In this regard, recent studies (e.g. Lusk and Schroeder 2004; Ding, Grewal, and Liechty 2005; Chang, Lusk, and Norwood 2009; Loomis et al. 2009; Chowdhury et al. 2011; De-Magistris, Gracia, and Nayga 2013; Grebitus, Lusk, and Nayga 2013; Moser, Raffaelli, and Notaro 2013) have shown that DCEs are indeed susceptible to hypothetical bias even when they are used to value marketable private goods.

As a result, researchers are focusing on increasing the realism of stated DCE studies by either employing hypothetical bias mitigation strategies (HBMS) (see Loomis 2011; and Loomis 2014 for a review) or by shifting to nonhypothetical choice experiments in order to reduce or completely eliminate hypothetical bias. Employing the latter, however, can be difficult since in addition to being costly and time consuming, the products of interest may not be available (Fox et al. 1996; De-Magistris, Gracia, and Nayga 2013). Therefore, HBMS may serve as important tools to remove hypothetical bias, and one such strategy, which has been often used in the literature, is the Cheap Talk (CT) script (Cummings and Taylor 1999). However, the existing evidence with respect to the effectiveness of the CT script is mixed. While some have found that the CT script effectively mitigates hypothetical bias (Lusk 2003; Carlsson, Frykblom, and Lagerkvist 2005; Murphy, Stevens, and Weatherhead 2005; List, Sinha, and Taylor 2006; Chowdhury et al. 2011; Tonsor and Shupp 2011), others have found it to be effective only for certain sub-groups of respondents (List 2001; Aadland and Caplan 2003, 2006; Lusk 2003; Barrage and Lee 2010; Ami et al. 2011). Furthermore, the findings of Brown, Ajzen, and Hrubes (2003) suggest that the CT is bid range sensitive, and Nayga, Woodward, and Aiew (2006); Brummet, Nayga, and Wu (2007); and Blumenschein et al. (2008) all found that the CT did not have any influence on the WTP values.

Relatively few studies (List, Sinha, and Taylor 2006; Özdemir, Johnson and Hauber 2009; Carlsson, Garcia, and Löfgren 2010; Bosworth and Taylor, 2012 and Moser, Raffaelli, and Notaro 2013) have tested the credibility of the CT script in DCE settings and their results generally show that the CT is ineffective in removing the hypothetical bias. This might be explained by the fact that the CT script was originally developed for CV studies, and applying it to the DCEs may not be appropriate as there are contextual and structural differences between the two valuation approaches (Ladenburg and Olsen 2014), thus, responses may differ markedly (List and Shogren 1998; Carson and Groves 2007). Given these inconsistencies surrounding the effectiveness of the CT script, researchers have proposed other HBMS. For instance, Jacquemet et al. (2013) implemented a *Solemn Oath* script, and De-Magistris, Gracia, and Nayga (2013) tested an *Honesty priming* approach. Both approaches were found to be effective in removing hypothetical bias. Other researchers have instead relied on calibrating hypothetical values using respondents' stated certainty in choice (Blumenschein et al. 2008).

More recently, along the same lines Ladenburg and Olsen (2014) suggested augmenting the CT script with an *Opt-Out Reminder* (OOR), which reminds respondents that if the prices of the experimentally designed alternatives are greater than what their household will pay, they should choose the opt-out alternative. In a split sample setup, they found that when the CT was applied with the OOR added, WTP estimates were significantly reduced compared to using the CT without the OOR. Varela et al. (2014) also tested the impact of presenting the CT with the OOR and contrary to Ladenburg and Olsen (2014), the OOR was not found to influence WTP. A possible explanation might be that Ladenburg and Olsen (2014) repeated the OOR before each single choice set whereas Varela et al. (2014) only presented it once in the scenario description. This seems to support Ladenburg and Olsen (2014) who speculate that, given the repeated choice nature of DCE, it may be of particular importance to repeat the reminder since respondents might otherwise forget about the reminder as they progress through the choice tasks.

This article contributes to these developments in the literature in terms of extending the work by Ladenburg and Olsen (2014). The aim is to properly test the effectiveness of including a Repeated OOR (ROOR) in stated DCEs in terms of testing its ability to mitigate hypothetical bias. The major limitation of both Ladenburg and Olsen (2014) and Varela et al. (2014) is that they tested the OOR in a purely hypothetical setup. They did not include a nonhypothetical treatment which is necessary for establishing the extent to which hypothetical bias is present in the first place. Hence, they could not conclude whether the OOR actually reduced hypothetical bias. Therefore, in this article we include an incentivized discrete choice experiment ("REAL") treatment in addition to a fairly standard hypothetical DCE ("HYPO") to investigate whether hypothetical bias exists at the outset. A third treatment presents the exact same hypothetical DCE except that the ROOR is added ("ROOR"), in order to examine the impact of the ROOR. Both Ladenburg and Olsen (2014) and Varela et al. (2014) presented the OOR script in combination with a CT script, thus confounding the effects of both scripts. To isolate the effect of the OOR in terms of mitigating hypothetical bias, we do not use a CT script but simply present respondents with the ROOR alone.

With this article we contribute to the literature in several aspects. First and foremost, to our knowledge, this is the first valid test of the effectiveness of the ROOR in terms of mitigating hypothetical bias. Second, the majority of studies investigating hypothetical bias have done so in a developed country context whereas only a couple of papers have considered it in a developing country context (Ehmke, Lusk, and List 2008; Chowdhury et al. 2011). We conduct our study in a developing country, thus adding to the sparse DCE literature in these settings. This also represents a novel addition to the growing literature on nonhypothetical DCE (Carlsson and Martinsson 2001; Lusk and Schroeder 2004, 2006; Ding, Grewal, and Liechty 2005; Alfnes et al. 2006; Johansson-Stenman and Svedsäter 2008; Lusk, Fields, and Prevatt 2008; Chang, Lusk and Norwood 2009; Loomis et al. 2009; Volinsky et al. 2009; Yeu and Tong 2009; Grebitus, Lusk, and Nayga 2013; Michaud, LIerena, and Joly 2013; Gracia 2014). Finally, whilst the aim of the article is mainly methodological, it also makes a novel contribution to the literature on consumers' food preferences by estimating Kenyan consumers' WTP for novel food products (NFPs) made with edible insects.

Background: Edible insects as a sustainable source of food

The world population is projected to increase to 9 billion in 2050 (UN 2015). Faced with climate change and environmental sustainability concerns, the current food production systems may not be able to meet the growing demand for food which is estimated to rise by 60% in 2050 (FAO 2009, 2013a). In this respect, it is evident that the number of malnourished people in developing countries is (still) alarmingly high although some nutritional improvements have been observed in recent years (Braun 2007). Specifically, despite a globally positive change in the supply of food energy, total protein and animal protein in the past few decades, a decreasing trend has been observed in many developing regions (Pellet and Ghosh 2004), especially in several sub-Saharan African (SSA) countries where food security has been and still is one of the biggest societal challenges. Increasing livestock production is one option to tackle the shortage of animal protein but it contributes significantly to the emission of greenhouse gasses (Steinfeld et al. 2006) while also requiring substantial inputs of water and crop-based feed. Both these required production inputs are expected to decrease in availability in SSA with future global warming. Increasing food security thus requires that either the productivity of the current agricultural sector increases e.g. through new technological developments (FAO 2013b) or alternative sources of food that increase food security in a sustainable manner must be identified through research and development (Boland et al. 2013; van Huis 2013).

In line with this, the concept of producing edible insects for food has received substantially increasing attention recently (FAO 2013a). Insects can be a good source of important nutrients for human consumption (Bukkens 1997; Christensen et al. 2006). As shown by Oonincx and Boer (2012) and Oonincx et al. (2010), edible insects can be considered environmentally friendly as they emit significantly lower amounts of greenhouse gasses than livestock production. Despite insects being traditionally consumed as food in large parts of Africa, Asia and Latin America, and historically all over the world, the many different aspects of eating insects, which has earlier been labelled with the collective term 'entomophagy' (Evans et al. 2015), has received rather limited attention in research (van Huis 2013). Recently, however, the literature investigating e.g. food safety, nutritional value, consumer attitude and legislative aspects of producing insects for food has begun to increase very rapidly(e.g. Finke 2013; van Huis 2013; Looy, Dunkel, and Wood 2014; Ruby, Rozin, and Chan 2015; Alemu et al. 2017). Most of the recent research focuses on such supply side issues. However, the demand side, i.e. consumers' preferences for insects as food, will ultimately be equally important for a successful introduction of edible insects as a new and substantial component of the future food production sector. Therefore, gathering information about consumers' preferences is necessary since decisions concerning establishment of new insect production sectors as well as regulatory, standardization and quality control schemes are likely to be suboptimal if knowledge about potential demand in the market is not accounted for.

Conceptual framework and econometric specification

The random utility theory (RUT) (McFadden 1986) serves as a theoretical framework to specify utility expressions in DCEs. In what follows, we follow Train and Weeks (2005). According to RUT, the utility of individual *n* choosing alternative *k* among *J* given alternatives in choice situation *t* can be expressed as:

(1) $$U_{nkt} = \alpha_n m_{nkt} + \boldsymbol{\beta}_{\boldsymbol{n}}' \boldsymbol{x}_{\boldsymbol{nkt}} + \epsilon_{nkt}$$

The utility is expressed as a component of price, m, and non-price attributes, $\boldsymbol{x}$. The parameters α_n and vector $\boldsymbol{\beta}_{\boldsymbol{n}}'$ are the coefficients of the price and non-price attributes. For logit models, the variance of the error term ϵ_n can be written as $s_n^2 \left(\frac{\pi^2}{6}\right)$, where $s_n$is the scale parameter which differs across decision makers. Without loss of generality, equation (1) can be divided by s_n as:

(2) $$U_{nkt} = \left(\frac{\alpha_n}{s_n}\right) m_{nkt} + \left(\frac{\boldsymbol{\beta_n}}{\boldsymbol{s_n}}\right)' \boldsymbol{x}_{\boldsymbol{nkt}} + \varepsilon_{nkt}$$

Letting $\lambda_n = \left(\frac{\alpha_n}{s_n}\right)$, and $c_n = \left(\frac{\boldsymbol{\beta_n}}{\boldsymbol{s_n}}\right)$, equation (2) becomes:

$$(3) \qquad U_{nkt} = \lambda_n m_{nkt} + \boldsymbol{c_n'} \boldsymbol{x_{nkt}} + \varepsilon_{nkt}$$

where the error term, ε_n, is distributed IID extreme value.

Equation (3) describes utility in *preference space*. As suggested by Train and Weeks (2005), an alternative option is to employ a *WTP space* approach. The utility in preference space in equation (3) may be reformulated so that the estimated coefficients refer directly to the WTP values. This approach has been applied by Train and Weeks (2005); Sonnier, Ainslie, and Otter (2007); Scarpa, Thiene, and Train (2008); Balcombe, Chalak, and Fraser (2009); Thiene and Scarpa (2009); Hensher and Greene (2011); and Hole and Kolstad (2012). Most of the evidences so far suggest that models based on the WTP space approach yield more plausible behavioral explanations. Specifically, the distributions of WTP values are found to be within reasonable margins than values based on the preference space formulation. Therefore, there is an increasing support for implementation of the WTP space specification when the goal is to obtain appropriate WTP estimates which would be used for policy guidance. As mentioned above, the WTP space is expressed by reformulating the preference space utility expression. Reparametrizing equation (3) gives:

$$(4) \qquad U_{nkt} = \lambda_n [m_{nkt} + w_n' \boldsymbol{x_{nkt}}] + \varepsilon_{nkt}$$

Equation (4) is called utility in *WTP space.* Train and Weeks (2005) mentioned that equation (3) and (4) are behaviorally equivalent. Nevertheless, one can see, and as mentioned before, WTP calculation based on the preference space specification is directly impacted by the distribution of λ_n and c_n, and in some cases like normal distributions, the moments of the WTP distribution may be undefined and furthermore the WTP distribution may suffer from a fat tail problem causing identification problems.

Heterogeneity in preferences can be accounted for by specifying mixed logit models, which can be derived in a number of different ways (Train 2003). The most common specification is a random parameter logit (RPL) model in which attribute parameters are treated as random parameters. The random parameters may in principle follow any relevant distribution, but often a normal distribution is used for quality attribute parameters and a lognormal is used for the negative of the price attribute parameter (Train 2003; Train and Sonnier 2005). We employ the RPL model as focus group interviews had indicated preference heterogeneity across participants. In line with Scarpa, Thiene, and Train (2008), we assume that $\lambda_n = -exp(\upsilon_n)$, where υ_n is the latent random factor of the price coefficient. In addition, let $\boldsymbol{\beta}$ represent all the random parameters estimated in the WTP space model, then equation (4) can be rewritten as:

$$(5) \qquad U_{nkt} = V_{nkt}(\beta_n, x_{nkt}) + \varepsilon_{nkt}$$

where V_{nk} *is* the indirect utility function, which is a function of the attributes of the alternatives, i.e. $V_{nk} = \beta_n' X_{nkt}$. The (negative of the) coefficient on price (λ_n) is assumed to follow a lognormal distribution in order to ensure a strictly positive impact of income on utility. The focus group discussions and pilot tests indicated that some people would have negative preferences for the non-price attributes of the products while others would have positive. As a result, the distributions of the coefficients for these attributes are assumed to follow a normal distribution. We allow the scale, which is accommodated through λ_n, to vary across decision makers by specifying all random parameters to be correlated with each other, i.e. no restrictions are placed on the variance-covariance matrix. As the estimated parameters, β', vary over individuals according to the specified distributions with density $f(\beta)$ (Train 2003), the probability that an individual *n* chooses alternative *k* among *J* given alternatives in choice situation *t* can be derived as:

$$(6) \qquad P_{nk} = \int \left(\frac{e^{\beta_n' X_{nkt}}}{\sum_j^J e^{\beta' x_{ntj}}} \right) f(\beta) d\beta$$

In line with, e.g. Thiene and Scarpa (2009) and Hole and Kolstad (2012), Equation (6), which is based on the WTP space specification, is estimated using the MSL procedure by maximizing the full simulated log-likelihood over the sequence of choices and over the sample. We estimate the models using the BIOGEME 2.2 software (Bierlaire 2003), maximizing the simulated likelihood function with the CFSQP algorithm (Lawrence, Zhou, and Tits 1997) to avoid the local optima problem (Scarpa, Thiene, and Train 2008). For the simulation procedure, 3000 Modified Latin Hypercube Sampling draws (Hess, Train, and Polak 2006) were used since this amount of draws was found to produce stable parameter estimates.

Experimental design and procedure

We test the effectiveness of the ROOR in a DCE field experiment in Kenya, where there is increasing interest in introducing edible insects into the food production sector. Specifically, we focus on consumers' preferences for crickets as food since this type of insect is considered particularly promising in terms of developing viable production systems for household consumption as well as for large scale commercialization in Kenya.

The product contexts and DCE design

Rather than presenting respondents with whole crickets on a plate, traditional buns were chosen as the product context for the DCE since it was considered a more realistic market entry strategy to introduce crickets by adding cricket flour (CF) to buns which are generally considered affordable and available in most food outlets throughout Kenya. Buns are one of the bread types that form the major staple food in Kenya according to focus group interviews and key informant discussions. Furthermore, the practical manageability of product development and handling guided the choice of buns for this experiment. CF was mixed into wheat flour to bake the buns. The amounts of CF used to produce the buns were determined based on the recipe specification developed by Kinyuru et al. (2009). Serving as attribute levels in the DCE, the amounts of CF were varied at 0%, 5%, and 10% of the total flour. The product development process, the recipe formulation and the characteristics of the final bun products used for the field experiment are presented in Appendix A.

Non-hypothetical DCEs involving new products require development of all the product combinations with all the relevant attributes. This is, however, typically constrained by budget and other logistic complications. For this study, we identified three attributes that would enable us to actually develop and produce all possible product combinations for the DCE. The first attribute represents the insect component in terms of the amount of CF in the buns *(CF)*. The second attribute describes whether a portion of the wheat flour is fortified or not *(Fortified)*, while the third attribute is the price of a bag of buns *(Price)*. Insects like crickets may be consumed in their whole form or grinded into flour. Besides representing the more realistic market entry strategy, we focus on the flour form because low protein foods such as maize and sorghum, which are popular in east Africa (Stevens and Winter-Nelson 2008), can be easily enhanced with CF to increase their nutritional value. In addition, CF is already being produced on a rather large scale elsewhere in the world, hence, a similar production would seem achievable in Kenya.

The *Fortified* attribute was included to assess Kenyan consumers' preferences and WTP for fortified foods. The level for the fortified wheat flour was arbitrarily set to be 40% of the total flour used to bake the different buns. Consumers in Kenya are familiar with food fortification because the Kenyan government through the Kenya National Food Fortification Alliance (KNFFA) has introduced a food fortification program to reduce micronutrient deficiency. This program covers foods including maize and wheat (KNFFA 2011). A secondary purpose with including the fortified flour attribute was an attempt to (at least to some extent) separate out the pure effect of nutritional improvements. In other words, we wanted to give participants the option to choose buns that were nutritionally better than the standard alternative without necessarily forcing them to choose cricket flour. This way, if participants strongly rejected the CF, they might still exhibit preferences for better nutrition, thus avoiding confounding preferences for nutrition, which we expected to be positive, with preferences for other aspects of eating insects which might be negative, e.g. if respondents would find the idea of eating insects disgusting (Looy, Dunkel, and Wood 2014).

The quality attribute levels are presented in table A.1 in Appendix A. The amount of CF has three levels: 0g (*Standard with no CF*), 6.25g (*Medium CF*), and 12.5g (*High CF*). This is equivalent to 0%, 5% and 10%, respectively, of the total wheat flour used per bag of buns. The *Fortified* attribute has two levels at 0g and 50g. The total amount of flour is kept constant at 125g per bag. Thus, when CF and/or fortified wheat flour is added, the amount of wheat flour is reduced accordingly. Besides that quality attributes, a *Price* attribute is included with six levels. An initial investigation of actual market prices of buns showed that a single bun would typically cost around 10 Kenyan Shillings[1] (KShs), rarely less than KShs 5. Since focus group discussions indicated that respondents would not respond negatively to the other attributes, it was thus decided to set the minimum price for a bag with three buns at KShs 20. Focus group discussions further suggested that a suitable choke price for the price range would be KShs 90, and the six price levels were subsequently set at KShs 20, 25, 30, 40, 60, and 90.

[1] 1 US Dollar ~ 90.50 KShs at the time of the experiment.

Once the attributes and their levels had been determined, the DCE design was produced. It was subjected to design procedures and multiple rounds of testing according to theories on experimental design (e.g. Louviere, Hensher, and Swait 2000). Based on a pilot study with 42 respondents, priors were obtained to inform and update a D-efficient fractional factorial design generated using the Ngene software (ChoiceMetrics 2012). The design produced 12 choice sets which were presented to participants.

"Insert Figure 1 about here"

Figure 1 shows an example of one of the choice sets used. As is evident from the example, we use a forced choice type of DCE by including an alternative with a standard bag of buns which is constant across all choice sets. While this may be considered problematic from a welfare theoretical point of view and especially in cases concerning public goods, it is quite common when assessing preferences for private goods (Carlsson and Martinsson 2001; Carlsson, Frykblom, and Lagerkvist 2007a, b; Harrison 2006). The decision to include a no-choice option or not in an experimental design depends on the objective of the research at hand (Dhar and Simonson 2003; Parker and Schrift 2010). Specifically, when the experiment involves new products, the inclusion of a no-choice option may be less desirable as consumers may use it as a simplifying strategy without making the required trade-offs between attributes. While one may argue that such situations exist in a real market, it may not represent the aims and motivations of the specific experiment (Dhar and Simonson 2003).

The experimental treatments and hypothesis testing

Three experimental treatments were designed to investigate whether hypothetical bias persists and whether it can be mitigated by a ROOR. In the first treatment (HYPO) participants make purely hypothetical choices without having to pay for the bun products. In the second treatment (REAL) participants' choices are no longer hypothetical as they are subjected to an incentivization mechanism similar to that of Alfnes et al. (2006) and Gracia, Loureiro, and Nayga (2011). The inclusion of this treatment enables testing the extent to which hypothetical bias is present. The lack of such a treatment was the key limitation of Ladenburg and Olsen (2014) and Varela et al. (2014) who could only speculate about the impact of the OOR on hypothetical bias. In the third treatment (ROOR), participants make hypothetical choices but now framed with a ROOR in which subjects are provided with the following opt-out reminder:

"*If the prices of Bag 1 and Bag 2 are higher than what you think your household will pay, you should choose Bag 3*".

This treatment is specifically designed to investigate if hypothetical bias can be mitigated by the ROOR. As the name suggests the ROOR is placed repeatedly before each choice set so as to take the repeated nature of the choices into account and help subjects to stay on the "true" preference path (Ladenburg and Olsen 2014). In this manner, information flow and consistent trade-offs across choice sets would be ensured. In addition, the ROOR enables participants to attend to both the designed and the opt-out alternatives appropriately, in which case participants should make the required trade-offs among the attributes to reach a decision that ideally mirrors their "true" preferences. Moreover, as discussed in Ladenburg and Olsen (2014), unlike the CT script, the ROOR is not affected by conformity issues as it does not convey any implicit cues as to what others do and how they behave in relation to value revelations.

The experiment is undertaken based on a between-subject design to control for behavioral shift and confounding effects which might result if a within-subject design was adopted (Lusk and Schroeder 2004; Charness, Gneezy, and Kuhn 2012). In total, 334 participants took part in the survey. Participants were randomized into the three treatments, resulting in 109 participants in each of the HYPO and ROOR treatments, and 116 participants in the REAL treatment. The participants are either household heads or spouses as they are the decision makers in dietary selection for their household members.

With this experimental setup we are able to test the following four research hypotheses. First, based on our expectation that the ROOR addresses the tendency of forgetting the opt-out alternatives in DCEs, we hypothesize that the hypothetical setting will result in fewer opt-out choices than the incentivized setting, but when adding the ROOR the amount of opt-out choices will be similar to the incentivized setting:

$$\text{H1: Opt-out frequency}_{\text{HYPO}} < \text{Opt-out frequency}_{\text{REAL}} = \text{Opt-out frequency}_{\text{ROOR}}$$

Second, we hypothesize that the hypothetical setting will be subject to hypothetical bias in terms of WTP estimates being higher than in the incentivized setting:

$$H2: WTP_{HYPO} > WTP_{REAL}$$

Conditional on confirmation of H2, we hypothesize that by adding the ROOR we can reduce hypothetical bias in terms of achieving WTP estimates that are lower than those of the standard hypothetical setting:

$$H3: WTP_{HYPO} > WTP_{ROOR}$$

Finally, also conditional on H2 being confirmed, but representing a stronger test than H3, we hypothesize that adding the ROOR will not only reduce hypothetical bias, but even completely eliminate it. In other words, the ROOR treatment will generate WTP estimates equivalent to the incentivized setting:

$$H4: WTP_{ROOR} = WTP_{REAL}$$

Experimental procedure

Survey interviews were conducted face-to-face using trained interviewers. The personal interviews took place in three different parts of Kenya that were selected to ensure inclusion of respondents from both rural and urban areas, of various different ethnicities, and where some still maintained the tradition of eating insects while others had left it. Respondents were recruited using a stratified random sampling approach to further ensure variation across key socio-demographic variables. For further descriptions of the sampling procedure, see XXX (REMOVED TO PRESERVE ANONYMITY IN REVIEW PROCESS). The actual interview procedure followed four steps. First, participants were welcomed and seated and one trained interviewer was assigned to each for conducting the personal interview. Second, they were given KShs 30 in cash for their participation in the experiment. Participants were then randomly assigned to one of the three experimental treatments described above. The participants in the REAL treatment received an additional remuneration of KShs 90 which they could use fully or partly to buy the chosen products in the DCE. The provision of the monetary remuneration before the start of the

DCE process was intended to induce a sense of ownership and avoid the house money effect that may be associated with windfall earnings in experiments (Thaler and Johnson 1990; Cárdenas et al. 2014). Third, the samples of the three bun products were provided to participants so that they could view, smell and taste them. The products were shown to respondents in random orders to avoid product ordering effects on their sensory evaluations. They were then asked to fill in a 9-point hedonic scale concerning the taste and overall liking of the products (Chowdhury et al. 2011 and Holmquist, McCluskey, and Roos 2012). Fourth, participants were then presented with the DCE questionnaire. In addition to the choice sets, it contained information[2] about crickets as food as well as a description of the attributes. Additionally, participants in the REAL treatment received instructions regarding the incentivization mechanism (See Appendix B). In Lusk and Schroeder (2004), Chang, Lusk, and Norwood (2009) and Mørkbak, Olsen, and Campbell (2014) a single random number was drawn for a whole group of participants to identify the binding choice set for all participants in that group. As opposed to this, but in line with Alfnes et al. (2006) and Gracia, Loureiro, and Nayga (2011), the identification of the binding choice set in this study was implemented by asking each individual participant to draw a numbered ball from 1 to 12 from a hat after they had answered the 12 choice tasks. Participants were then given the bag of buns they had chosen in the binding choice set and they had to pay the price of that bag.

Results

The summary of sample socio-demographics presented in table 1 shows no differences across treatments. Chi-square tests failed to reject that any of the sociodemographic distributions differ across treatments, suggesting that participants in the three treatments are sufficiently similar in their characteristics to rule this potential factor out as cause of any preference differences found.

"Insert table 1 about here"

[2] This information was provided before consumers were subject to the DCE to mirror real market situations in that consumers often process a range of information before they make purchase decisions. However, we have not tested the effects of information provision on consumers' responses as it falls beyond the scope of this paper.

As a first step in investigating possible preference differences across sample splits, table 2 provides a comparison of the choice distributions. Interestingly, while the *Standard* buns were chosen in 14% of the choices in the HYPO treatment, they were chosen in 27% and 31% of the choices in the ROOR and REAL treatments, respectively. These results are in line with Lusk and Schroeder (2004) who also found participants to choose the opt-out option more frequently in a real setting than in a hypothetical setting, though they used a "none of these" opt-out. The chi-square tests confirm that the choice frequencies differ significantly for HYPO versus ROOR and HYPO versus REAL. However, there seems to be no significant difference when comparing ROOR and REAL. In other words, we can confirm hypothesis H1. These results provide a first indication that the ROOR may indeed serve its purpose as it seems to generate a better correspondence between hypothetical and non-hypothetical choices.

"Insert table 2 about here"

The estimation results are presented in table 3. As noted before, the estimated parameters directly represent the implied WTP distributions and the coefficients of the attributes are interpreted as marginal WTP estimates. Preference equality across the three treatments is tested using a likelihood ratio test (Swait and Louviere 1993; Louviere, Hensher, and Swait 2000; Lusk and Schroeder, 2004; Chowdhury et al. 2011; Moser, Raffaelli, and Notaro 2013). The null hypothesis is rejected ($X^2 = 152.8, p - \text{value} < 0.01$) suggesting that the overall preference structure is not identical across the three treatments. Furthermore, all the estimated standard deviations of the random parameters[3] are significant, indicating substantial preference heterogeneity across participants in all three treatments.

The model results presented in table 3 show that consumers are willing to pay a premium for buns containing CF. More interestingly, the WTP estimates differ across the treatments. Generally, the WTP values obtained from the HYPO sample are three to six times as large as the WTP values obtained from the REAL sample, suggesting that the HYPO WTPs suffer noticeably from hypothetical bias. Specifically, the WTP estimate for *High CF* in the HYPO sample is almost six times that of the REAL sample. In addition, the WTP estimate for the *Medium CF* is more than three times as large in the HYPO

[3] The estimates of the cholesky matrices, which were obtained from the finite difference hessian, are reported in Appendix C.

as in the REAL sample. Turning to the WTP for *fortified* relative to *non-fortified wheat flour*, the estimate obtained in the HYPO sample is more than four times that obtained from the REAL sample. In accordance with the significantly fewer choices of bag 3, i.e. the opt-out, in the HYPO sample (see table 2), the ASC estimate for bag 3 is much more negative in the HYPO compared to the REAL, confirming that participants in the HYPO react much more negatively to the opt-out alternative, irrespective of the attribute levels of the offered alternatives.

"Insert table 3 about here"

In line with e.g. Lusk and Schroeder (2004), Tonsor and Shupp (2011), de-Magistris, Gracia, and Nayga (2013), and Moser, Raffaelli, and Notaro (2013) we use the complete combinatorial method (Poe, Giraud, and Loomis 2005) to test whether the differences between WTP estimates obtained for a given attribute in the HYPO and REAL treatments are significantly different from zero. For this purpose, 1,000 WTP estimates are bootstrapped for each attribute in each sample based on the Krinsky and Robb (1986) method using 1,000 draws from multivariate normal distributions of the coefficient estimates and the variance-covariance matrix of the random parameters. According to the test results in table 3, we can reject the null hypothesis of equal WTP estimates across the HYPO and REAL treatments for all attributes. Given that the attribute WTPs obtained in the HYPO sample are all higher than those of the REAL treatment, we thus confirm hypothesis H2, i.e. hypothetical bias is present in the HYPO sample. Furthermore, the negative opt-out effect observed for the ASC estimate in the HYPO treatment, i.e. participants disproportionally often choosing the non-opt-out alternatives regardless of their attributes, is not present in the REAL treatment. This serves as another indication of hypothetical bias, corresponding to the findings in table 2.

Turning to the question of particular interest in this article namely the effect of the ROOR, the complete combinatorial test results show that for all attributes the ROOR treatment results in significantly lower WTP estimates than the HYPO treatment. Hence, we can also confirm hypothesis H3. For the *High CF* attribute the WTP obtained in the ROOR treatment is though still significantly higher than in the REAL treatment, while for the *Medium CF* the difference is not significant, and for the *Fortified* attribute the difference is only borderline significant. The ASC estimates are not significantly different, but the fact that the ASC estimate is still significantly lower than zero in the ROOR indicates that the negative status quo effect is not completely removed. Hence, the slightly stronger hypothesis H4 can only be partly confirmed since hypothetical bias, despite being significantly reduced, is not completely removed for all attributes. Notwithstanding this, presenting the ROOR before each choice set appears to be highly effective, removing 87-98% of the hypothetical bias in absolute terms across the range of attributes as well as for the negative opt-out effect.

As indicated in the introductory section of this article, some authors found that the CT script was effective only for certain sub-groups of respondents. For instance, List (2001); Lusk (2003); and Barrage and Lee (2010) found that the CT script did not reduce hypothetical bias for consumers who are experienced and knowledgeable about the product in question. Aadland and Caplan (2003) concluded that the CT script was effective in reducing stated WTP for respondents who are young, educated women and members of environmental organizations. To examine whether hypothetical bias and the impacts of the ROOR differ across sub-groups of participants, we estimate separate models for low income participants and middle-to-high income participants[4]. Results in tables 4 and 5 show that WTP estimates of the middle-to-high income participants are, as expected, markedly higher than those of the low income participants across all treatments. While the ROOR reduces the hypothetical bias for both groups, its impact appears to be more pronounced for the low income participants as it effectively eliminates the hypothetical bias for two out of the three attributes as well as the negative opt-out effect in this group. In the middle-to-high income group the ROOR is unable to completely eliminating hypothetical bias for any of the attributes nor the negative opt-out effect.

"Insert table 4 about here"
"Insert table 5 about here"

The above results confirm that people tend to overstate their WTP when they are presented with simple hypothetical scenarios without any financial incentives or without any hypothetical bias mitigation strategy. Participants in DCE often ignore one or several attributes when making their choices

---

[4] Participants are categorized into three groups based on their monthly household income as shown in table 1. The low income participants include those with monthly income of less than 15000 KShs whereas the medium-to-high income participants include those with monthly income above 15000 KShs. This choice of a demographic to investigate the impact of the ROOR across sub-groups was somewhat arbitrary but mainly motivated by the fact that we would expect hypothetical bias to depend on income.

(e.g. Campbell, Hutchinson, and Scarpa 2008; Scarpa et al. 2009; Campbell, Hensher, and Scarpa 2011; Alemu et al. 2013; Mørkbak, Olsen, and Campbell 2014) and non-attendance to price may explain hypothetical bias. Attribute non-attendance (AN-A) can be obtained using either stated non-attendance (SNA) or inferred non-attendance (INA) approaches. INA can be inferred from equality constrained latent class (ECLC) models (see Scarpa et al. 2009, 2013) or by calculating the coefficient of variation of individual specific posterior means and variances (see Hess and Hensher 2010). Here, we use the INA approach based on the ECLC model to infer AN-A and we refer the reader to Scarpa et al (2009, 2013), Hensher and Greene (2010), and Campbell, Hensher, and Scarpa (2011) for further details of this approach. In line with previous literature on AN-A, we are interested in attribute processing strategies rather than in preference heterogeneity in the ECLC models. Therefore, while the attributes, which are assumed to be non-attended, take zero values, the others are constrained to take the same value across classes. Different combinations of AN-A can be assumed in the ECLC model. For completeness, we specify a model with all possible combinations of AN-A (Hensher, Rose, and Greene 2012). The AN-A combinations are: 1) all attributes attended, 2) all attributes ignored, 3) one attribute ignored, 4) two attributes ignored, 5) three attributes ignored. The results reported in table 6 show that the overall probability of ignoring the *Price* attribute in the HYPO sample is 70% which is much higher than the 40% in the REAL sample. This offers an explanation of the hypothetical bias in the HYPO sample. Table 6 also shows that this might be one of the key aspects making the ROOR effective. By specifically directing attention to the tradeoff between quality attributes and the price attribute, the ROOR treatment obtains a 20 percentage points decrease in price non-attendance. While price non-attendance is still some 10 percentage points higher than in the REAL treatment, this does seem to provide a likely the behavioral explanation why hypothesis H3 is confirmed and hypothesis H4 is only partly confirmed.

"Insert table 6 about here"

Conclusions

Eliciting the value of goods based on approaches that reflect the real behavior of decision makers is a central element of valuation studies be it for non-market or market applications. Stated DCE approaches are widely used to elicit the value of goods and their attributes, e.g. in order to predict demand for new products. How to reduce or eliminate hypothetical bias remains an important challenge for DCE researchers. Nonhypothetical DCEs are obviously the preferred option; however, in many cases product unavailability and the associated high costs of implementation impair researchers from relying on such approaches. As a result, hypothetical DCEs are still used extensively. A range of different strategies, e.g. consequential survey questions, honesty priming, inferred valuation, and solemn oath scripts, have been suggested as amendments to hypothetical DCEs, aiming to reduce or eliminate hypothetical bias (Carson and Groves 2007, Lusk and Norwood 2009, Jacquemet et al. 2013, and De-Magistris, Gracia, and Nayga 2013).

Inspired by previous studies by Ladenburg and Olsen (2014) and Varela et al. (2014), we extend this line of research by investigating the effectiveness of a ROOR in terms of reducing or eliminating hypothetical bias. This study is based on a field experiment concerned with consumer valuation of novel food products made with cricket flour in Kenya. Three treatments – one hypothetical, one hypothetical with ROOR, and one non-hypothetical incentivized DCE – are tested in a between-subject design aimed at determining, firstly, the magnitude of hypothetical bias and, secondly, the ability of the ROOR to remove it.

The results confirm that the mean marginal WTP values based on the simple hypothetical DCE suffer severely from hypothetical bias with estimates being up to several percentages higher than the WTP estimates obtained in the non-hypothetical treatment. This is in contrast with Carlsson and Martinsson (2001), and Lusk and Schroeder (2004) where the hypothesis of equal marginal WTP values in real and hypothetical settings could not be rejected. It should be noted that direct comparison of our results with these studies may not be appropriate due to reasons associated with the nature of the goods being valued, the sampling design and the nature of the experimental settings (Harrison and List 2004; Harrison 2006). More interestingly, the results indicate that the ROOR considerably reduces or even eliminates the hypothetical bias while also being capable of reducing or removing negative opt-out effects. The results based on the ROOR are considerably closer to the results based on the nonhypothetical sample than the results based on the hypothetical sample without the ROOR. The ROOR eliminates hypothetical bias for one attribute while strongly reducing, though not completely eliminating, it for others, highlighting the promising potential of introducing it in DCEs. Furthermore, although the ROOR appears to have a stronger impact for participants from low income groups than from high income groups, it reduces up to 85% of hypothetical bias for the later groups. This is encouraging since, unlike the CT which is found to be effective only for certain groups of respondents (List 2001; Aadland and Caplan 2003, 2006; Lusk 2003; Barrage and Lee 2010; and Ami et al. 2011), the ROOR seems to be quite effective for all income groups even though it is more effective for some than for others.

Our results differ from Ladenburg and Olsen (2014) and Varela et al. (2014) in four important ways. First, their studies considered preferences for public goods whereas the current study considers a private good context in terms of food choice. Second, these studies did not establish the extent to which hypothetical bias existed, which may be seen as a major limitation since it essentially precluded the authors from making any conclusions concerning the effectiveness of the ROOR in terms of mitigating hypothetical bias. Our experimental setup allows us to conclude that hypothetical bias is indeed present and of a non-negligible magnitude. Third, while the ROOR did not influence responses in the case of Varela et al. (2014), it significantly reduced the marginal WTP estimates of only some of the attributes in Ladenburg and Olsen (2014). Contrary to this, the results of the current study reveal that the ROOR significantly reduces the hypothetical bias for all attributes and even completely eliminates it for one attribute. Fourth, they implemented the ROOR together with a CT script, which confounded the effects of the ROOR and the CT. Thus, compared to these, our study provides a more strict experimental setup allowing a much more rigorous test of the ROOR in isolation.

The cause of hypothetical bias in hypothetical experiments is largely tied to the fact that people ignore price or they give less emphasis to price in part because they will not be asked to pay in real circumstances (List and Gallet 2001; Ding, Grewal, and Liechty 2005; Murphy et al. 2005). We find that the ROOR increases attendance to the *Price* attribute in a hypothetical DCE. This has particular relevance to stated DCE since framing them with the ROOR does not pose further challenges in terms of monetary cost, product availability and or logistic burden that otherwise might be the case in nonhypothetical DCEs. In sum, the main results of this study show that not only financial incentives but also ROOR can lead to plausible behavioral manifestations. Therefore, in situations where it is not possible to employ nonhypothetical DCEs, framing hypothetical DCEs with a ROOR would seem to be a promising strategy. Our results, however, provide only first-hand evidence and are not conclusive by themselves. Therefore, future studies are important in order to replicate our results and provide further empirical evidence in relation to the ability of the ROOR to improve the external validity of stated DCEs. Future studies may consider conducting research in different experimental settings and with different types of goods so as to validate whether the effect of ROOR is context-specific or not. Moreover, the actual wording of the

ROOR is not tested here as it is beyond the scope of this article. Future studies may test other wordings of the ROOR, e.g. by focusing on the designed alternatives rather than on the opt-out alternative. Another area of further research could be testing whether the ROOR leads to the same results as the results from the current study when it is applied to experimental designs that include a no-choice alternative rather than a forced choice.

Acknowledgements

This research is part of the *GREEiNSECT* (13-06KU) project which was funded by the Danish International Development Agency (DANIDA), Ministry of Foreign Affairs of Denmark. We thank the management of the Department of Food Science and Technology at the Jomo Kenyatta University of Agriculture and Technology for giving us the opportunity to bake the bun products at the Food Processing Workshop Unit. Finally, we thank seminar participants at the university of Copenhagen and University of Stirling for comments and suggestions. The usual disclaimer applies for the remaining errors.

References

Aadland, D., and A. J. Caplan. 2003. Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias. *American Journal of Agricultural Economics* 85 (2):492-502.

———. 2006. Cheap talk reconsidered: New evidence from CVM. *Journal of Economic Behavior and Organization* 60 (4):562-578.

Alemu, M., M. Mørkbak, S. Olsen, and C. Jensen. 2013. Attending to the Reasons for Attribute Non-attendance in Choice Experiments. *Environmental and Resource Economics* 54 (3):333-359.

Alemu, M. H., S. B. Olsen, S. E. Vedel, K. O. Pambo, and V. O. Owino. 2017. Combining product attributes with recommendation and shopping location attributes to assess consumer preferences for insect-based food products. *Food Quality and Preference* 55:45-57.

Alfnes, F., A. G. Guttormsen, G. Steine, and K. Kolstad. 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.

Alpizar, F., F. Carlsson, and O. Johansson-Stenman. 2008. Anonymity, reciprocity, and conformity: Evidence from voluntary contributions to a national park in Costa Rica. *Journal of Public Economics* 92 (5–6):1047-1060.

Ami, D., F. Aprahamian, O. Chanel, and S. Luchini. 2011. A Test of Cheap Talk in Different Hypothetical Contexts: The Case of Air Pollution. *Environmental and Resource Economics* 50 (1):111-130.

Ayieko, M.A., Ndong'a, M.F.O., Tamale, A., 2010. Climate change and the abundance of edible insects in the Lake Victoria region. *Journal of Cell and Animal Biology* 4(7): 112-118.

Balcombe, K., A. Chalak, and I. Fraser. 2009. Model selection for the mixed logit with Bayesian estimation. *Journal of Environmental Economics and Management* 57 (2):226-237.

Barrage, L., and M. S. Lee. 2010. A penny for your thoughts: Inducing truth-telling in stated preference elicitation. *Economics Letters* 106 (2):140-142.

Bierlaire, M. 2003. BIOGEME: A Free Package for the Estimation of Discrete Choice Models. Paper presented at the 3rd Swiss Transport Research Conference Monte Verita, Ascona, March 2003.

Blumenschein, K., G. C. Blomquist, M. Johannesson, N. Horn, and P. Freeman. 2008. Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment. *The Economic Journal* 118 (525):114-137.

Boland, M. J., A. N. Rae, J. M. Vereijken, M. P. M. Meuwissen, A. R. H. Fischer, M. A. J. S. van Boekel, S. M. Rutherfurd, H. Gruppen, P. J. Moughan, and W. H. Hendriks. 2013. The future supply of animal-derived protein for human consumption. *Trends in Food Science and Technology* 29 (1):62-73.

Bosworth, R., and L. O. Taylor. 2012. Hypothetical Bias in Choice Experiments: Is Cheap Talk Effective at Eliminating Bias on the Intensive and Extensive Margins of Choice? *The B.E. Journal of Economic Analysis and Policy* 12(1): 1-26.

Braun, J., V. 2007. Focus on the world's poorest and hungry people: IFPRI 2006-2007 Annual Report Essay: International Food Policy Research Institute (IFPRI).

Brown, S.L., and K.M. Eisenhardt. 1995. PRODUCT DEVELOPMENT: PAST RESEARCH, PRESENT FINDINGS, AND FUTURE DIRECTIONS. *Academy of Management Review* 20:343-378.

Brown, T. C., I. Ajzen, and D. Hrubes. 2003. Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation. *Journal of Environmental Economics and Management* 46 (2):353-361.

Brummett, R. G., R. Nayga, and X. Wu. 2007. On the use of cheap talk in new product valuation. *Economics bulletin* 2 (1):1-9.

Bukkens, S. G. F. 1997. The nutritional value of edible insects. *Ecology of Food and Nutrition* 36 (2-4):287-319.

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

Campbell, D., W. G. Hutchinson, and R. Scarpa. 2008. Incorporating Discontinuous Preferences into the Analysis of Discrete Choice Experiments. *Environmental and Resource Economics* 41 (3):401-417.

Cárdenas, J. C., N. De Roux, C. R. Jaramillo, and L. R. Martinez. 2014. Is it my money or not? An experiment on risk aversion and the house-money effect. *Experimental Economics* 17 (1):47-60.

Carlsson, F., and P. Martinsson. 2001. Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?: Application to the Valuation of the Environment. *Journal of Environmental Economics and Management* 41 (2):179-192.

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

Carlsson, F., P. Frykblom, and C. J. Lagerkvist. 2007. Consumer Benefits of Labels and Bans on GM Foods—Choice Experiments with Swedish Consumers. *American Journal of Agricultural Economics* 89 (1):152-161.

Carlsson, F., J. García, and Å. Löfgren. 2010. Conformity and the Demand for Environmental Goods. *Environmental and Resource Economics* 47 (3):407-421.

Carson, R. T., N. E. Flores, K. M. Martin, and J. L. Wright. 1996. Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods. *Land Economics* 72 (1):80-99.

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

Chang, J. B., J. L. Lusk, and F. B. Norwood. 2009. How Closely Do Hypothetical Surveys and Laboratory Experiments Predict Field Behavior? *American Journal of Agricultural Economics* 91(2): 518-534.

Charness, G., U. Gneezy, and M. A. Kuhn. 2012. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior and Organization* 81 (1):1-8.

ChoiceMetrics, 2012. Ngene 1.1.1 User Manual and Reference Guide, Australia.

Chowdhury, S., J. V. Meenakshi, K. I. Tomlins, and C. Owori. 2011. Are Consumers in Developing Countries Willing to Pay More for Micronutrient-Dense Biofortified Foods? Evidence from a Field Experiment in Uganda. *American Journal of Agricultural Economics* 93 (1):83-97.

Christensen, D. L., F. O. Orech, M. N. Mungai, T. Larsen, H. Friis, and J. Aagaard-Hansen. 2006. Entomophagy among the Luo of Kenya: a potential mineral source? *International Journal of Food Sciences and Nutrition* 57 (3-4):198-203.

Cummings, R. G., and L. O. Taylor. 1999. Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. *The American Economic Review* 89 (3):649-665.

de-Magistris, T., A. Gracia, and R. M. Nayga. 2013. On the Use of Honesty Priming Tasks to Mitigate Hypothetical Bias in Choice Experiments. *American Journal of Agricultural Economics* 95 (5):1136-1154.

Dhar, R., and I. Simonson. 2003. The Effect of Forced Choice on Choice. *Journal of Marketing Research* 40 (2):146-160.

Ding, M., Grewal, R., and Liechty, J. 2005. Incentive-Aligned Conjoint Analysis. *Journal of Marketing Research* 42(1): 67-82.

Ehmke, M. D., J. L. Lusk, and J. A. List. 2008. Is Hypothetical Bias a Universal Phenomenon? A Multinational Investigation. *Land Economics* 84 (3):489-500.

Evans, J., M. H. Alemu, R. Flore, M. B. Frøst, A. Halloran, A. B. Jensen, G. Maciel-Vergara, V. B. Meyer-Rochow, C. Münke-Svendsen, S. B. Olsen, C. Payne, N. Roos, P. Rozin, H. S. G. Tan, A. v. Huis, P. Vantomme, and J. Eilenberg. 2015. ‘Entomophagy’: an evolving terminology in need of review. *Journal of Insects as Food and Feed* 1 (4):293-305.

Finke, M. D. 2013. Complete Nutrient Content of Four Species of Feeder Insects. *Zoo Biology* 32 (1):27-36.

Fox, J. A., D. J. Hayes, J. F. Shogren, and J. B. Kliebenstein. 1996. Experimental Methods in Consumer Preference Studies. *Journal of Food Distribution Research* 27 (2): 1-7.

Gracia, A., M. L. Loureiro, and R. M. Nayga. 2011. Are Valuations from Nonhypothetical Choice Experiments Different from Those of Experimental Auctions? *American Journal of Agricultural Economics* 93 (5):1358-1373.

Gracia, A. 2014. Consumers' preferences for a local food product: a real choice experiment. *Empirical Economics* 47 (1):111-128.

Grebitus, C., J. L. Lusk, and R. M. Nayga Jr. 2013. Explaining differences in real and hypothetical experimental auctions and choice experiments with personality. *Journal of Economic Psychology* 36:11-26.

Harrison, G. W., and J. A. List. 2004. Field Experiments. *Journal of Economic Literature* 42 (4):1009-1055.

Harrison, G. 2006. Experimental Evidence on Alternative Environmental Valuation Methods. *Environmental and Resource Economics* 34 (1):125-162.

Hensher, D. A. 2010. Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological* 44 (6):735-752.

Hensher, D. A., and W. H. Greene. 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. *Empirical Economics* 39 (2):413-426.

Hensher, D. A., and W. H. Greene. 2011. Valuation of Travel Time Savings in WTP and Preference Space in the Presence of Taste and Scale Heterogeneity. *Journal of Transport Economics and Policy* 45 (3):505-525.

Hensher, D.A., J.M. Rose, and W.H. Greene. 2012. Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design. *Transportation* 39:235-245.

Hess, S., K. E. Train, and J. W. Polak. 2006. On the use of a Modified Latin Hypercube Sampling (MLHS) method in the estimation of a Mixed Logit Model for vehicle choice. *Transportation Research Part B: Methodological* 40 (2):147-163.

Hess, S., and D. A. Hensher. 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B: Methodological* 44 (6):781-790.

Hole, A., and J. Kolstad. 2012. Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. *Empirical Economics* 42 (2):445-469.

Holmquist, C., J. McCluskey, and C. Ross. 2012. Consumer Preferences and Willingness to Pay for Oak Attributes in Washington Chardonnays. *American Journal of Agricultural Economics* 94:556-561.

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

Johansson-Stenman, O., and H. Svedsäter. 2008. Measuring Hypothetical Bias in Choice Experiments: The Importance of Cognitive Consistency. *The B.E. Journal of Economic Analysis and Policy* 8(1): (Article 41).

Kinyuru, J., G. Kenji, and M. Njoroge. 2009. Process development, nutrition and sensory qualities of wheat buns enriched with edible termites (Macrotermes subhylanus) from Lake Victoria region, Kenya. *African Journal of Food, Agriculture, Nutrition and Development* 9: 1739-1750.

Krinsky, I., and A. L. Robb. 1986. On Approximating the Statistical Properties of Elasticities. *The Review of Economics and Statistics* 68 (4):715-719.

Ladenburg, J., and S. B. Olsen. 2014. Augmenting short Cheap Talk scripts with a repeated Opt-Out Reminder in Choice Experiment surveys. *Resource and Energy Economics* 37:39-63.

Lawrence, C., and J. L. Zhou. 1997. User's Guide for CFSQP Version 2.5: A C Code for Solving (Large Scale) Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality Constraints. Technical Report TR-94-16r1, Institute for Systems Research, University of Maryland, College Park, MD 20742, 1997.

Layton, D. F., and G. Brown. 2000. Heterogeneous Preferences Regarding Global Climate Change. *Review of Economics and Statistics* 82 (4):616-624.

Levitt, S. D., and J. A. List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *The journal of economic perspectives 21*:153-174.

List, J. A., and J. F. Shogren. 1998. Calibration of the difference between actual and hypothetical valuations in a field experiment. *Journal of Economic Behavior & Organization* 37 (2):193-205.

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., and C. Gallet. 2001. What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? *Environmental and Resource Economics* 20(3): 241-254.

List, J. A., P. Sinha, and M. H. Taylor. 2006. Using Choice Experiments to Value Non-Market Goods and Services: Evidence from Field Experiments. *Advances in Economic Analysis and Policy* 6(2): 1-37 (Article 2)

Little, J., and R. Berrens. 2004. Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. *Economics bulletin* 3 (6):1-13.

Loomis, J., P. Bell, H. Cooney, and C. Asmus. 2009. A Comparison of Actual and Hypothetical Willingness to Pay of Parents and Non-Parents for Protecting Infant Health: The Case of Nitrates in Drinking Water. *Journal of Agricultural and Applied Economics* 41 (03):697-712.

Loomis, J. 2011. What's to know about hypothetical bias in stated preference valuation studies? *Journal of Economic Surveys* 25 (2):363-370.

Loomis, J. B. 2014. Strategies for overcoming hypothetical bias in stated preference surveys. *Journal of Agricultural and Resource Economics* 39 (1):34-46.

Looy, H., F. Dunkel, and J. Wood. 2014. How then shall we eat? Insect-eating attitudes and sustainable foodways. *Agriculture and Human Values* 31 (1):131-141.

Louviere, J. J., D. A. Hensher, and J. D. Swait. 2000. *Stated choice methods: analysis and applications*: Cambridge University Press.

Lusk, J. L. 2003. Effects of Cheap Talk on Consumer Willingness-to-Pay for Golden Rice. *American Journal of Agricultural Economics* 85 (4):840-856.

Lusk, J. L., J. Roosen, and J. A. Fox. 2003. 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(1): 16-29.

Lusk, J.L., and T.C. Schroeder. 2004. "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks." *American Journal of Agricultural Economics* 86:467-482.

Lusk, J. L., D. Fields, and W. Prevatt. 2008. An Incentive Compatible Conjoint Ranking Mechanism. *American Journal of Agricultural Economics* 90 (2):487-498.

McFadden, D. 1986. The Choice Theory Approach to Market Research. *Marketing Science* 5 (4):275-297.

Meijer, E., and J. Rouwendal. 2006. Measuring welfare effects in models with random coefficients. *Journal of Applied Econometrics* 21 (2):227-244.

Michaud, C., D. Llerena, and I. Joly. 2013. Willingness to pay for environmental attributes of non-food agricultural products: a real choice experiment. *European Review of Agricultural Economics* 40 (2):313-329.

Moser, R., R. Raffaelli, and S. Notaro. 2014. Testing hypothetical bias with a real choice experiment using respondents' own money. *European Review of Agricultural Economics* 41 (1):25-46.

Mørkbak, M. R., S. B. Olsen, and D. Campbell. 2014. Behavioral implications of providing real incentives in stated choice experiments. *Journal of Economic Psychology* 45:102-116.

Murphy, J., P. G. Allen, T. Stevens, and D. Weatherhead. 2005. A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics* 30(3): 313-325.

Murphy, J., T. Stevens, and D. Weatherhead. 2005. Is Cheap Talk Effective at Eliminating Hypothetical Bias in a Provision Point Mechanism? *Environmental and Resource Economics* 30 (3):327-343.

Nayga, R. M., R. Woodward, and W. Aiew. 2006. Willingness to Pay for Reduced Risk of Foodborne Illness: A Nonhypothetical Field Experiment. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 54 (4):461-475.

Oonincx, D. G. A. B., and I. J. M. de Boer. 2012. Environmental Impact of the Production of Mealworms as a Protein Source for Humans – A Life Cycle Assessment. *PLoS ONE* 7 (12):e51145.

Oonincx, D. G. A. B., J. v. Itterbeeck, M. J. W. Heetkamp, H. van den Brand, J. J. A. v. Loon, and A. v. Huis. 2010. An Exploration on Greenhouse Gas and Ammonia Production by Insect Species Suitable for Animal or Human Consumption. *PLoS ONE* 5 (12):1-7.

Özdemir, S., F. R. Johnson, and A. B. Hauber. 2009. Hypothetical bias, cheap talk, and stated willingness to pay for health care. *Journal of Health Economics* 28 (4):894-901.

Parker, J. R. and R. Y. Schrift. 2011. Rejectable Choice Sets: How Seemingly Irrelevant No-Choice Options Affect Consumer Decision Processes. *Journal of Marketing Research* 48 (5):840-854.

Pellett, P.L., and S. Ghosh. 2004. Lysine fortification: past, present, and future. *Food Nutr Bull* 25:107-113.

Poe, G. L., K. L. Giraud, and J. B. Loomis. 2005. Computational Methods for Measuring the Difference of Empirical Distributions. *American Journal of Agricultural Economics* 87 (2):353-365.

Revelt, D., and K. Train. 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. *Review of Economics and Statistics* 80 (4):647-657.

Ruby, M. B., P. Rozin, and C. Chan. 2015. Determinants of willingness to eat insects in the USA and India. *Journal of Insects as Food and Feed* 1 (3):215-225.

Scarpa, R., F. Spalatro, and M. Canavari. 2007. Investigating Preferences for Environment Friendly Production Practices. In *Organic Food: Consumers' Choice and Farmers' Opportunities*, edited by M. Canavari and K. Olson: Springer New York, 115-124.

Scarpa, R., M. Thiene, and K. Train. 2008. Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps. *American Journal of Agricultural Economics* 90 (4):994-1010.

Scarpa, R., T. J. Gilbride, D. Campbell, and D. A. Hensher. 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics* 36 (2):151-174.

Scarpa, R., R. Zanoli, V. Bruschi, and S. Naspetti. 2013. Inferred and Stated Attribute Non-attendance in Food Choice Experiments. *American Journal of Agricultural Economics* 95 (1):165-180.

Sonnier, G., A. Ainslie, and T. Otter. 2007. Heterogeneity distributions of willingness-to-pay in choice models. *Quantitative Marketing and Economics* 5 (3):313-331.

Steinfeld, H., P. Gerber, T. Wassenaar, V. Castel, M. Rosales, and C. De Haan. 2006. *Livestock's long shadow*: FAO Rome.

Stern, S. 1997. Simulation-Based Estimation. *Journal of Economic Literature* 35 (4):2006-2039.

Stevens, R., and A. Winter-Nelson. 2008. Consumer acceptance of provitamin A-biofortified maize in Maputo, Mozambique. *Food Policy* 33 (4):341-351.

Swait, J., and J. Louviere. 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research* 30 (3):305-314.

Thaler, R. H., and E. J. Johnson. 1990. Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice. *Management Science* 36 (6):643-660.

Thiene, M., and R. Scarpa. 2009. Deriving and Testing Efficient Estimates of WTP Distributions in Destination Choice Models. *Environmental and Resource Economics* 44 (3):379-395.

Tonsor, G. T., and R. S. Shupp. 2011. Cheap Talk Scripts and Online Choice Experiments: Looking Beyond the Mean. *American Journal of Agricultural Economics* 93 (4):1015-1031.

Train, K. E. 1998. Recreation Demand Models with Taste Differences over People. *Land Economics* 74 (2):230-239.

Train, K. E. 2003. *Discrete choice methods with simulation*: Cambridge university press.

Train, K., and G. Sonnier. 2005. Mixed Logit with Bounded Distributions of Correlated Partworths. In *Applications of Simulation Methods in Environmental and Resource Economics*, edited by R. Scarpa and A. Alberini. Dordrecht: Springer Netherlands, 117-134.

Train, K., and M. Weeks. 2005. Discrete Choice Models in Preference Space and Willingness-to-Pay Space. In R. Scarpa and A. Alberini, eds. *Applications of Simulation Methods in Environmental and Resource Economics*. Dordrecht, Springer Netherlands, 1-16.

United Nations, Food and Agricultural Organizations. 2009. *The state of food and agriculture: Livestock in the balance.* Rome.

United Nations, Food and Agricultural Organizations. 2013a. *Edible Insects: future prospects for food and feed security*. Rome.

United Nations, Food and Agricultural Organizations. 2013b. *The state of food insecurity in the world: The multiple dimension of food security.* Rome.

United Nations. 2015. *World Population Prospects: The 2015 Revision, Key Findings and Advance Tables*. New York.

van Huis, A. 2013. Potential of Insects as Food and Feed in Assuring Food Security. *Annual Review of Entomology,* 58: 563-583.

van Kleef, E., H. C. M. van Trijp, and P. Luning. 2005. Consumer research in the early stages of new product development: a critical review of methods and techniques. *Food Quality and Preference* 16 (3):181-201.

Varela, E., P.-A. Mahieu, M. Giergiczny, P. Riera, and M. Soliño. 2014. Testing the single opt-out reminder in choice experiments: An application to fuel break management in Spain. *Journal of Forest Economics* 20 (3):212-222.

Volinskiy, D., W. L. Adamowicz, M. Veeman, and L. Srivastava. 2009. Does Choice Context Affect the Results from Incentive-Compatible Experiments? The Case of Non-GM and Country-of-Origin Premia in Canola Oil. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 57 (2):205-221.

Vossler, C. A., M. Doyon, and D. Rondeau. 2012. Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. *American Economic Journal: Microeconomics* 4 (4):145-171.

Vossler, C. A., and S. B. Watson. 2013. Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior and Organization* 86:137-147.

Yue, C., and C. Tong. 2009. Organic or Local? Investigating Consumer Preference for Fresh Produce Using a Choice Experiment with Real Economic Incentives. *HortScience* 44 (2):366-371.

Tables and Figures

Table 1. Summary Statistics of Participants' Characteristics by Treatment (Percentages)

Variable	HYPO	ROOR	REAL
Age			
18 - 34 years	38.5	45.9	40.5
35 – 54 years	55.0	41.3	47.4
55 – 64 years	5.5	8.3	9.5
Above 64 years	1	4.5	2.6
Household size			
1 – 4 persons	56.9	48.6	53.4
More than 4 persons	43.1	51.4	46.6
Gender			
Female	53.2	52.3	50.1
Male	46.8	47.7	49.9
Region			
Rural	60.1	60.1	64
Urban	39.9	39.9	36
Monthly income in			
KShs	66	63	58
< 15,000	25	30	37
15,000 – 50,000	9	7	5
> 50,000			

Education			
Primary school	31.2	43.1	27.6
Secondary school	36.7	31.2	28.5
Tertiary level	11	14.7	15.5
University education	6.4	2.8	3.5
Other[a]	14.7	8.2	24.6
Number of participants	109	109	116

[a] no education, and drop outs from primary and secondary schools.

Table 2. Choice Frequencies

Alternative	HYPO(1)	ROOR(2)	REAL(3)	x^2-test, *p*-value		
	%	%	%	(1) vs (2)	(1) vs (3)	(2) vs (3)
Bag1	43.1	34.5	33.9			
Bag2	42.7	38.2	35.5	<0.001	<0.001	0.136
Bag3 (Standard buns)	14.2	27.3	30.6			

Note: The chi-square tests have been carried out based on the actual number of choices rather than based on percentages.

Table 3. Model Results and Comparison of WTP Estimates

Parameters	HYPO (1)	ROOR (2)	REAL (3)	(1) vs (3) *p-value*	(1) vs (2) *p*-value	(2) vs (3) *p*-value
Mean parameter estimates						
ASC (Standard bun)	-32.1 (5.42)*	-5.24 (2.24)*	-3.40 (2.29)	0.000*	0.000*	0.287
High CF	143 (13.8)*	40.8 (3.76)*	25.7 (3.67)*	0.000*	0.000*	0.000*
Medium CF	138 (11.5)*	43.4 (3.33)*	41.7 (4.28)*	0.000*	0.000*	0.407
Fortified	64.1 (6.79)*	20.4 (2.95)*	14.4 (2.03)*	0.000*	0.000*	0.049*
Price	-3.2 (0.185)*	-2.17 (0.177)*	-2.29 (0.131)*			
Standard deviation estimates						
ASC (Standard bun)	37.8 (5.04)*	12.5 (2.50)*	7.72 (2.58)*	0.000*	0.000*	0.104
High CF	118 (9.3)*	86.1 (8.51)*	53.5 (5.67)*	0.000*	0.002*	0.004*
Medium CF	45.8 (3.34)*	28.9 (2.56)*	11.1 (1.75)*	0.000*	0.000*	0.000*
Fortified	14.0 (2.21)*	4.34 (1.31)*	3.62 (1.36)*	0.000*	0.000*	0.320
Price	0.770 (0.251)*	0.742 (0.239)*	0.752(0.247)*			
No. of observations	1308	1308	1392			
Null log-likelihood	-1437.0	-1437.0	-1529.3			
Final Log-likelihood	-835.7	-825.5	-993.3			
Adjusted ρ^2	0.405	0.412	0.337			

Note: '*' indicates significance at 5% level or lower. Figures in parenthesis are standard errors. *p*-values obtained using the complete combinatorial method (Poe, Giraud, and Loomis 2005) based on 1,000 bootstrapped WTP estimates derived using Krinsky and Robb (1986) method. *p*-values represent results of the one sided test that the differences between the equivalent WTP estimates of two treatments are positive.

Table 4. Model Results and Comparison of WTP Estimates for Low Income Participants

Parameters	HYPO (1)	ROOR (2)	REAL (3)	(1) vs (3) *p*-value	(1) vs (2) *p-value*	(2) vs (3) *p-value*
Low income participants						
Mean parameter estimates						
ASC (Standard bun)	-10.6 (2.42)*	-1.47 (1.76)	-3.32 (2.34)	0.034*	0.000*	0.663
High CF	112 (14.1)*	31.4 (1.39)*	18.8 (3.25)*	0.000*	0.000*	0.001*
Medium CF	117 (9.70)*	33.4 (3.21)*	32.1 (4.60)*	0.000*	0.000*	0.430
Fortified	55.9 (6.80)*	15.9 (1.68)*	13.5 (2.47)*	0.000*	0.000*	0.249
Price	-3.05 (0.213)*	-1.65 (0.355)*	-2.30 (0.174)*			
Standard deviation estimates						
ASC (Standard bun)	12.9 (3.95)*	17.5 (3.49)*	17.3 (8.13)*	0.577	0.836	0.494
High CF	116 (12.5)*	63.8 (2.98)*	45.8 (5.73)*	0.000*	0.000*	0.003*
Medium CF	38.8 (3.26)*	23.0 (2.08)*	8.42 (2.32)*	0.000*	0.000*	0.002*
Fortified	20.7 (2.34)*	11.9 (0.943)*	3.23 (1.72)*	0.000*	0.000*	0.012*
Price	0.608 (0.305)*	1.32 (0.338)*	0.873 (0.272)*			
No. of observations	852	828	804			
Null log-likelihood	-936.0	-909.7	-883.3			
Final Log-likelihood	-574.3	-512.7	-607.6			
Adjusted Rho-square	0.365	0.414	0.289			

Note: '*' indicates significance at 5% level or lower. Figures in parenthesis are standard errors. *p*-values obtained using the complete combinatorial method (Poe, Giraud, and Loomis 2005) based on 1,000 bootstrapped WTP estimates derived using Krinsky and Robb (1986) method. *p*-values represent results of the one sided test that the differences between the WTP estimates of two treatments are positive.

Table 5. Model Results and Comparison of WTP Estimates for Middle-to-High Income Participants

Parameters	HYPO (1)	ROOR (2)	REAL (3)	(1) vs (3) *p*-value	(1) vs (2) *p-value*	(2) vs (3) *p-value*
Middle-high income participants						
Mean parameter estimates						
ASC (Standard bun)	-81.1 (12.8)*	-10.1 (1.90)*	-4.23 (3.23)	0.000*	0.000*	0.023*
High CF	239 (52.6)*	77.4 (6.49)*	44.9 (5.75)*	0.000*	0.000*	0.000*
Medium CF	209 (42.2)*	83.2 (5.37)*	61.3 (4.82)*	0.000*	0.000*	0.000*
Fortified	80 (16.8)*	42.1 (4.24)*	19.7 (3.31)*	0.000*	0.000*	0.000*
Price	-3.53 (0.30)*	-2.35 (0.336)*	-2.13 (0.202)*			
Standard deviation estimates						
ASC (Standard bun)	68.5 (9.25)*	25.2 (4.20)*	9.34 (2.61)*	0.000*	0.000*	0.000*
High CF	170 (36.6)*	96.8 (11.5)*	77.4 (10.6)*	0.002*	0.009*	0.054*
Medium CF	54 (12.2)*	40.6 (5.94)*	24.5 (2.72)*	0.001*	0.125	0.015*
Fortified	19.6 (6.51)*	15.2 (2.82)*	7.98 (1.69)*	0.007*	0.216	0.027*
Price	0.897 (0.315)*	1.23 (0.362)*	0.744 (0.241)*			
No. of observations	456	480	588			
Null log-likelihood	-501.0	-527.3	-646.0			
Log-likelihood	-245.1	-297.3	-367.1			
Adjusted Rho-square	0.471	0.398	0.401			

Note: '*' indicates significance at 5% level or lower. Figures in parenthesis are standard errors. *p*-values obtained using the complete combinatorial method (Poe, Giraud, and Loomis 2005) based on 1,000 bootstrapped WTP estimates derived using Krinsky and Robb (1986) method. *p*-values represent results of the one sided test that the differences between the WTP estimates of two treatments are positive.

Table 6. Estimated Membership Probabilities of AN-A from the ECLC Model

Class	Attribute non-attendance	Class membership probabilities		
		HYPO	ROOR	REAL
1	All attributes attended	0.1356 (3.73)	0.1262 (2.93)	0.1161 (3.39)
2	All attributes ignored	0.1528 (3.84)	0.0496 (1.76)	0.1227 (3.12)
3	Medium CF ignored	0	0	0.0011(0.26)
4	High CF ignored	0.0296 (1.44)	0.1577 (3.44)	0.2928 (5.65)
5	Fortified ignored	0.0102 (0.86)	0.0293 (1.19)	0.0188 (1.04)
6	Price ignored	0.2594 (5.29)	0.1481 (3.91)	0.0123 (0.90)
7	High CF and Medium CF ignored	0.0314 (1.66)	0.0314 (1.18)	0.0548 (1.69)
8	High CF and Fortified ignored	0.0131 (1.97)	0.0011(0.24)	0.0024 (0.27)
9	High CF and Price ignored	0	0.0024(0.29)	0.0156(1.02)
10	Medium CF and Fortified ignored	0	0	0
11	Medium CF and Price ignored	0.0179 (1.04)	0	0
12	Fortified and Price ignored	0.2247 (4.55)	0.1953 (3.99)	0.2492 (5.89)
13	High CF, Medium CF and Fortified ignored	0.0637 (2.52)	0.1385 (3.44)	0.0999 (3.19)
14	High CF, Medium CF and Price ignored	0	0.0034 (0.34)	0.0012(0.27)
15	High CF, Fortified and Price ignored	0.0493 (2.11)	0.0199 (1.21)	0.0016(0.25)
16	Medium CF, Fortified and Price ignored	0.0014 (0.25)	0.0945 (2.15)	0.0104(0.8)
	Accumulated probability of ignoring price	0.71	0.51	0.41

Note: Figures in parentheses are z-values. Probability values less than 0.001 are rounded to zero.

	Bag 1	Bag 2	Bag 3
Amount of cricket flour	0g	12.5g	I would purchase the standard scones at a price of 20 KShs
Amount of Fortified wheat flour	50g	50g	
Price	25 KShs	30 KShs	
I would choose (check √ one)			

Figure 1. An example of the choice sets

Appendices

Appendix A

The buns were baked at the Food Processing Workshop Unit at the Jomo Kenyatta University of Agriculture and Technology in Kenya by professional technicians. Since Kinyuru et al (2009) developed their recipe for buns from termite flour, we chose to validate this by first producing trial buns from CF. At this point, the product characteristics were assessed in cooperation with a nutrition scientist and several technicians. Following this, ten participants that included five males and five females were recruited to evaluate the products for taste, texture, color and smell. The recipe formulation that was used for the main baking is presented in table A.1. And the final bun products which were used for the field experiment are characterized as shown in table A.2. One can see that as the amount of the CF increases, the bun products become heavy, soft and brown. The increase in weight of the buns can be linked to the fact that the increase in the amount of CF lead to an increase in the amount of fat in the buns that can reduce water transpiration.

Table A.1. Ingredient Composition of the Buns

Buns	Amount of wheat flour (g)	Amount of cricket flour (g)	Amount of fortified wheat flour (g)
Standard bun	125	0	0
Fortified standard bun	75	0	50
Medium CF bun	118.75	6.25	0
Fortified Medium CF bun	68.75	6.25	50
High CF bun	112.5	12.5	0
Fortified High CF bun	62.5	12.5	50

Source: XXX (REMOVED TO PRESERVE ANONYMITY IN REVIEW PROCESS).

Note: 'g' refers to grams, and 'ml' refers to milliliters. Baking fat (7.5g), salt (1.25g), sugar (5g), yeast (2.5g), and acetic acid (0.125ml) were added to each bun.

Table A.2. Texture and Visual Appearance of the Buns

Buns	Final weight (g)	Texture	Appearance
Standard bun	158	Firm	White
Fortified standard bun	158	Firm	White
Medium CF bun	159	Medium firm	Medium brown
Fortified Medium CF bun	159	Medium firm	Medium brown
High CF bun	160	Soft	Brown
Fortified High CF bun	160	Soft	Brown

Source: XXX (REMOVED TO PRESERVE ANONYMITY IN REVIEW PROCESS).

Appendix B

Subject instruciton for the REAL treatment

"You will be provided with twelve different choice scenarios within which three bags (Bag 1, Bag 2, and Bag 3) of buns are included. Bag 1 and Bag 2 may contain buns made from wheat flour mixed with cricket flour. Some portion of the wheat flour can be fortified. Bag 3 contains only buns made from the wheat flour which was not fortified. In each scenario, you should choose ONE of the bags you would like to purchase (Bag 1 or Bag 2) or you can choose Bag 3 if you would not like to purchase Bag1 or Bag2. After you complete all 12 shopping scenarios, we will ask you to draw a number (1 to 12) from an envelope to determine which shopping scenario will be binding. In the envelope are numbers 1 through 12. If the number 1 is drawn then the first shopping scenario will be binding, and so on. For the binding scenario, we will look at the product you have chosen, give you your chosen product, and you will pay the listed price in that scenario. You should use the 90 KShs for the purchase. The most expensive alternatives cost 90 KShs. Although only one of the 12 shopping scenarios will be binding there is an equal chance of any shopping scenario being selected as binding, so think about each answer carefully."

Subject instruction for the HYPO and ROOR treatments

"You will be provided with twelve different choice scenarios within which three bags (Bag 1, Bag 2, and Bag 3) of buns are included. Bag 1 and Bag 2 may contain buns made from wheat flour mixed with cricket flour. Some portion of the wheat flour can be fortified. Bag 3 contains only buns made from the wheat flour which was not fortified. In each scenario, you should choose ONE of the bags you would like to purchase (Bag 1 or Bag 2) or you can choose Bag 3 if you would not like to purchase Bag1 or Bag2. For each choice scenario, assume that you have the opportunity to, here and now, to purchase ONE and ONLY ONE of the bags at the listed prices. While you will not actually buy any products today or pay the posted prices, please respond to each choice scenario as if it were a real one and you would have to give up real money were one of the 12 scenarios to be selected as binding."

Appendix C

Table C.1. HDCE Cholesky Matrix from WTP Space Estimates

Parameters	ASC	High CF	Medium CF	Fortified	$\ln(\lambda)$
ASC	37900 (99.5)				
High CF	28791 (125)	35700 (14.5)			
Medium CF	10125 (56.1)	18114 (10.6)	13700 (8.8)		
Fortified	6364 (36.2)	8256 (10.4)	5347 (7.64)	2700 (6.9)	
$\ln(\lambda)$	-41.5 (-3.51)	-139 (-4.9)	-31 (-1.16)	-12 (-0.899)	2.6 (3.8)

Note: Z-values in parenthesis.

Table C.2. HDCEROOR Cholesky Matrix from WTP Space Estimates

Parameters	ASC	High CF	Medium CF	Fortified	$\ln(\lambda)$
ASC	156 (2.5)				
High CF	-83 (-2.8)	7460 (5.1)			
Medium CF	-7 (-0.19)	6068 (5.9)	5800 (6.0)		
Fortified	187 (2.9)	2788 (5.8)	2923 (6.3)	1760 (5.9)	
$\ln(\lambda)$	0.8 (0.3)	-26 (-1.4)	-12 (-0.73)	-6 (-0.69)	1.46 (2.94)

Note: Z-values in parenthesis.

Table C.3. NDCE Cholesky Matrix from WTP Space Estimates

Parameters	ASC	High CF	Medium CF	Fortified	$\ln(\lambda)$

ASC	3770 (94.5)				
High CF	1962 (71.0)	4850 (7.3)			
Medium CF	688 (20.2)	3121 (5.9)	2680 (5.6)		
Fortified	575 (25.3)	1059 (6.9)	677 (4.9)	261 (4.5)	
$\ln(\lambda)$	9 (5.6)	-31 (-3.0)	-19 (-2.0)	-5 (-1.64)	1.14 (2.81)

Note: Z-values in parenthesis.

Table C.4. Estimates of the ECLC Model by Treatment

Parameters	HDCE	HDCEROOR	NDCE
	Estimate	Estimate	Estimate
High CF	4.47 (14.42)	3.88 (14.79)	3.88 (14.69)
Medium CF	3.84 (14.39)	2.99 (12.15)	3.08 (13.05)
Fortified	2.69 (12.04)	2.19 (11.54)	2.34 (12.78)
Price	-0.074 (-12.14)	-0.082 (-13.32)	-0.079 (-16.06)
Number of parameters	19	19	19
Final log-likelihood	-876.2	-897.0	-1078.2
Rho-square	0.545	0.569	0.474
AIC	1790.4	1832.0	2194.4
BIC	1841.6	1883.1	2246.8
AIC3	1809.4	1851.0	2213.4
CAIC	1860.6	1902.1	2265.8

Note: Z-values in parenthesis.