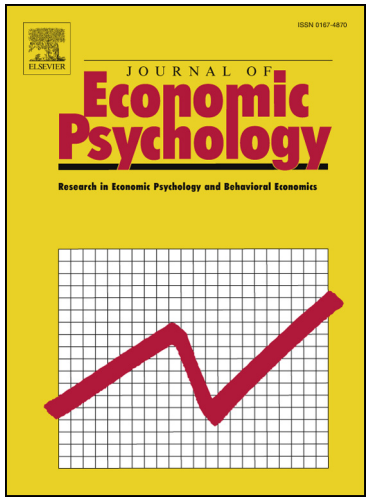

Behavioral implications of providing real incentives in stated choice experiments

Morten Raun Mørkbak, Søren Bøye Olsen, Danny Campbell

PII: S0167-4870(14)00059-2
DOI: http://dx.doi.org/10.1016/j.joep.2014.07.004
Reference: JOEP 1768

To appear in: *Journal of Economic Psychology*

Received Date: 7 August 2013
Revised Date: 11 April 2014
Accepted Date: 29 July 2014

Please cite this article as: Mørkbak, M.R., Olsen, S.B., Campbell, D., Behavioral implications of providing real incentives in stated choice experiments, *Journal of Economic Psychology* (2014), doi: http://dx.doi.org/10.1016/j.joep.2014.07.004

Title:

Behavioral implications of providing real incentives in stated choice experiments

Author names and affiliations:

Morten Raun Mørkbak[1,*], Søren Bøye Olsen[2], and Danny Campbell[3]

[1)] Department of Business and Economics, COHERE, University of Southern Denmark, Odense, Denmark.

[2)] Department of Food and Resource Economics, Faculty of Science, University of Copenhagen, Denmark

[3)] Economics Division, Stirling Management School, University of Stirling, Scotland

Corresponding author:

* Morten Raun Mørkbak
Department of Business and Economics, COHERE,
Campusvej 55,
University of Southern Denmark,
500 Odense,
Denmark
E-mail: mrm@sam.sdu.dk
Telephone: 0045-21949337

Behavioral implications of providing real incentives in stated choice experiments

Abstract

This paper explores the potential impacts of introducing real economic incentives in Choice Experiments (CE). While many others have investigated such impacts before, the majority of the literature has focused solely on mitigation of hypothetical bias. We contribute to this literature by widening the scope of research to other behavioural aspects where consumers in CE are often found to deviate from *homo economicus*. We develop a theoretical model where not only Willingness to pay (WTP) measures but also decision processing can be affected by the introduction of an economic incentive. Specifically, our model allows for differential impacts on attribute processing, depending on the character of the attribute as well as self-image effects. In an empirical CE survey, we find some, though not unequivocal, support of our model. Even though we find no impact on WTP from introducing an economic incentive, we find marked benefits in relation to a number of behavioural aspects that together would favour the use of an economic incentive regardless of hypothetical bias being present or not.

Key words: choice experiments, willingness to pay, food choice, real incentives, non-attendance, self-image

JEL codes: C25, C92, D63, Q13.

1. Introduction

The credibility of stated preference surveys has been questioned substantively in the literature for many years. At the same time, research aiming to investigate and increase the credibility of stated preference methods has emerged parallel to the criticism. One central criticism concerns the hypothetical nature of the methods which has been claimed to instigate respondents to overstate their 'true' willingness to pay (WTP) (see e.g. List et al. 2006; Lusk and Schroeder 2004; Murphy et al. 2005). Behavioural explanations offered for such overstatements relate to the fact that since respondents are not asked to actually pay the stated amounts out of their pockets, they have no incentives to answer truthfully – in accordance with their real preferences that is. Hence, respondents may exert free-riding or strategic behaviour in their responses or they may be susceptible to yea-saying or warm glow effects (Carson et al. 2001; Mitchell and Carson 1989). Covering these different types of behaviour, the term 'hypothetical bias' has been coined and is often referred to as an overarching, all-encompassing label. The issue of hypothetical bias has been investigated and treated in a number of ways in the literature, for instance by introducing budget reminders and cheap talk scripts or by seeking to decrease interviewer effects or by increasing scenario realism (e.g. Cummings and Taylor 1999; Carlsson et al. 2005). The latter has been accomplished by introducing valuation techniques where real economic incentives are incorporated. As a result, a number of choice experiment (CE) applications have incorporated real economic incentives in the experimental setup (e.g. Alfnes et al. 2006; Carlsson and Martinsson 2001; Chang et al. 2009; Gracia et al. 2011; Grebitus et al. 2013; Johansson-Stenman and Svedsäter 2012; Johansson-Stenman and Svedsäter 2008; List and Shogren 1998; Loomis et al. 2009; Lusk and Schroeder 2004; Mentzakis and Zhang 2012; Neill et al. 1994; Sinden 1988; Volinsky et al. 2009; Yeu and Tong 2009).

Because of this increased interest in and the broader acceptance of valuation methods with real economic incentives, there is a need for investigation of potential differences between methods with real incentives and the more traditional purely hypothetical stated choice experiments. Up until now, such differences have mainly been concentrated on hypothetical bias in terms of WTP differences (an exception is found in Scarpa et al. 2013, who examines attribute non-attendance), where findings have been ambiguous. For example, Carlsson and Martinsson (2001) find no differences in WTP[1], whereas Grebitus et al., (2013), Johansson-Stenman and Svedsäter (2008), Taylor et al., (2010); Broadbent et al., (2010) and Ready et al., (2010) all find differences between hypothetical and incentivized WTP estimates. At the same time Cameron et al., (2002) and List et al., (2006) only find differences to some extent. For a more thorough overview of studies see e.g. Ready et al. (2010). Based on meta-analysis of 28 studies of which the majority are CVM studies,

[1] It should be noted, though, that Carlsson and Martinsson (2001) is not a multi-attribute study, and this could possibly influence their results.

Murphy et al. (2005) find that hypothetical bias tends to be higher for public goods than for private goods. This is partly supported by Mjelde et al. (2012) who conclude that familiarity with the good being valued seems to decrease hypothetical bias and that bias is lower among older and higher educated respondents. Johansson-Stenman and Svedsäter (2012) further explain this difference by relating it to different impacts of self-image concerns, while Grebitus et al., (2013) show that respondents' personality traits have different impacts on choice behavior depending on whether it is a hypothetical or a non-hypothetical choice experiment. Of possible equal importance are other potential differences in the underlying behavioural decision mechanisms between the two survey methods, such as for instance the use of different decision making processing rules and the incidence of protesting and strategic behavior. While it is implicitly assumed in hypothetical CEs—as well as CEs with real incentives—that respondents are able to process all information in a fully rational manner, it is by now widely accepted that there are limits to how much information respondents can actually process. While many different attribute processing strategies have been identified in the literature of hypothetical choice behaviour (see e.g. Hensher (2007) and Hensher (2010) for overviews), one that has received increasing attention within the recent hypothetical choice behavior literature is attribute non-attendance, i.e. where respondents ignore one or more attributes of the alternatives when making their choices (e.g. Alemu et al. (2013), Balcombe et al. (2011), Hole et al. (2012), Hensher et al. (2005), Campbell et al. (2008), Hensher and Rose (2009) and Hensher (2010)). As examples of the extension of this behavior, Balcombe et al. (2011) find full attribute attendance by just above 70% of the sample while Hensher and Rose (2009) report that only 55% of the sample attended all attributes. This implies that individuals do not make the assumed trade-offs between attributes and attribute levels (Rosenberger et al. 2003; and Gowdy and Mayumi 2001) which may result in misleading WTP estimates. Hensher and Rose (2009) thereby conclude that WTP for travel time saving is higher when attribute non-attendance is implemented in the model specification. With respect to choices with real incentives, to the authors' knowledge, this behavioral aspect has only been explored partly by Scarpa et al. (2013). They employ a split sample design using two different goods—a beef product (without real incentives) and a chicken product (with real incentives). They find a higher degree of non-attendance in the survey with no real incentives, but unfortunately the results are confounded with the type of meat, since only the chicken sample received such real incentives. In the present paper we address this specific issue by examining the effect of providing real incentives holding the good in question constant. We examine potential differences in preferences as well as potential differences in respondents' decision making strategies including differences in protest behaviour as well as different degrees of non-attendance strategies both with respect to which attributes are ignored and the degree of non-attendance within each attribute, the latter is done by employing both a stated and an inferred approach, respectively. More specifically the paper explores potential differences in a choice experiment with real incentives and a hypothetical choice experiment with respect to general preference structure, error variance, willingness-to-pay, attribute non-attendance, using a case study on consumers' preferences for apple characteristics.

The paper is organised as follows: In section two we present the theoretical model as well as a range of hypotheses to be tested. This is followed by a presentation of the experimental setup and the data used. In section four a brief overview of the econometrics is provided, while section five reports the results and section six discusses and concludes.

2. Theoretical model

The standard utility model is:

$$U = v(X\beta') + \epsilon \quad (1)$$

where X and β' are vectors of the attributes of the good and their associated coefficients respectively. In the following we leave out the idiosyncratic error term ε for simplicity and focus on the observable indirect utility function v.

Similar to Johansson-Stenman and Svedsäter (2012), we follow Akerlof and Kranton (2000) and extend the standard economic utility model with the term s to capture self-image effects in the indirect utility function:

$$V = v(X\beta' + s) \quad (2)$$

An individual's utility of a good is then not only a function of the good but also of the impacts on self-image associated with buying or consuming the good. As noted by Johansson-Stenman and Svedsäter (2012), self-image may of course be influenced by a multitude of factors. They propose a model where self-image is a function of the extent to which individuals act in accordance with their ethical beliefs and the extent to which individuals are honest to themselves[2]. They test for self-image impacts by comparing two choice experiments; one considering a moral (public) good and one considering an amoral (private) good. The hypothesis is that the ethics and honesty self-image impacts will differ depending on whether a moral (public) or amoral (private) good is considered.

2.1 Hypotheses on WTP

[2] Any remaining self-image effects will be captured in the idiosyncratic error term, ε.

We continue along this line of thinking by proposing a similar model, though where the self-image impacts occur at the attribute level, x_i, rather than at the overall good level. Thus, adopting the notation used in Johansson-Stenman and Svedsäter (2012), we here define self-image as:

$$s = f\left(d_{x_i}^{ethics}, d_{x_i}^{honesty}\right) \quad (3)$$

where $d_{x_i}^{ethics} \equiv \left|MWTP_{x_i}^{hypothetical} - MWTP_{x_i}^{moral}\right|$ and $d_{x_i}^{honesty} \equiv \left|MWTP_{x_i}^{hypothetical} - MWTP_{x_i}^{true}\right|$. While $MWTP^{hypothetical}$ may be observed in surveys, it should be noted that $MWTP^{moral}$ and $MWTP^{true}$ are theoretical constructs that cannot be observed in practice, though $MWTP^{true}$ is often approximated by observing actual economic behavior in terms of $MWTP^{real}$. We also adopt assumptions from Johansson-Stenman and Svedsäter (2012) concerning first and second derivatives. This implies that self-image will be reduced in the ethical dimension if respondents state a hypothetical MWTP (or real MWTP) that is lower than their morally superior MWTP. It also implies that self-image will be increased in the honesty dimension if respondents state a hypothetical MWTP that is equal to their true (or real) MWTP. Thus, the ethics dimension would tend to drive hypothetical MWTP upwards while the honesty dimension would tend to drive it down towards the true MWTP. With only two attributes and two alternatives in their choice sets, Johansson-Stenman and Svedsäter (2012) as well as Grebitus et al. (2013) both use the simplest possible construction of a CE in their empirical investigations. However, CE applications in practice are typically more complex in that they have more attributes (commonly 4-6 attributes are used). We consider a more common CE setting with 5 attributes – some which may be considered moral attributes since they have some public good character (along with some private good character as well), and others which may be considered amoral due to their purely private good character. Moreover, we use a split sample setup similar to Johansson-Stenman and Svedsäter (2012) and Grebitus et al., (2013) where one split receives a purely hypothetical CE and the other sample receives an incentivized CE. While it is possible that both the ethical dimension as well as the honesty dimension $\left(d_{x_i}^{ethics}\right.$ and $\left.d_{x_i}^{honesty}\right)$ could affect behavior with regard to the public good attributes, extending on the findings in Johansson-Stenman and Svedsäter (2012), we hypothesize that the ethical dimension will not affect behavior for the private good attributes. In other words, we expect to find no differences between real and stated WTP for a private good attribute, whereas we expect that the ethical dimension could be causing stated MWTP to be higher than real WTP for the public good attributes. We thus put forward the following two hypotheses to be tested in the empirical analysis:

Hypothesis 1. For attributes X_i that have public good character: $MWTP_{X_i}^{real} < MWTP_{X_i}^{hypothetical}$

Hypothesis 2. For attributes X_j that have private good character: $MWTP_{X_j}^{real} = MWTP_j^{hypothetical}$

2.2 Hypotheses on decision processing

A simplifying decision strategy often observed in empirical CE surveys is when respondents resort to choosing the status quo alternative[3] disproportionately often – often referred to as status quo bias (Kahneman et al. 1991). In some cases, a respondent may simply choose the status quo in all choice tasks without even considering the experimentally designed alternatives. In these cases it is possible to ask clarifying follow-up questions which may be used to identify respondents exerting such behavior – commonly labeled protest behavior (e.g. Meyerhoff et al. 2013). It is conceivable that incentivizing the choice tasks and making them more consequential would reduce the propensity to resort to such a simplification strategy.

Hypothesis 3. The share of protesters will be lower in the real incentive sample than in the hypothetical sample.

Hypothesis 4. The share of occasional status quo choices will be lower in the real incentive sample than in the hypothetical sample.

Closely related to the use of simplifying heuristics, a proxy for the efforts respondents exert in choice tasks is the recorded response latency, *T*. As above, if the real incentive makes respondent consider their choices more closely, we would expect them to spend longer time making their decisions, and the decisions would exhibit less randomness.

Hypothesis 5. The real incentive will make respondents consider their choices more closely than in the hypothetical situation: $T^{real} > T^{hypothetical}$ and $\lambda^{real} > \lambda^{hypothetical}$, where λ is a scale parameter that is inversely proportional to the standard deviation of the error term (Swait and Louviere, 1993) (i.e. the higher the scale parameter, the less randomness in choices).

2.3 Hypotheses on attribute processing and self-image

While many different attribute processing strategies have been identified in the literature (see e.g. Hensher (2007) and Hensher (2010) for an overview), one that has received increasing attention in

[3] May also be referred to as the opt-out alternative, the none-of-these alternative or the buy nothing alternative.

the recent literature is attribute non-attendance, i.e. where respondents ignore one or more attributes of the alternatives when making their choices in CE (e.g. Hensher et al. (2005a), Campbell et al. (2008) and Hensher (2010)). This implies that individuals do not make the assumed trade-offs between attributes and attribute levels (Rosenberger et al. 2003; and Gowdy and Mayumi 2001). The indirect utility function can be elaborated to take attribute non-attendance into account:

$$V = v((1 - \gamma')X\beta' + s) \qquad (4)$$

Where γ' is a vector denoting whether each specific attribute is considered (γ=0) or ignored (γ=1) when processing the choice task. A number of reasons have been proposed why respondents might choose to ignore attributes (see e.g. Alemu et al. 2013). One often mentioned reason is that it serves as a decision heuristic in terms of simplifying choice tasks – an easy way to make fast choices. Making choices more consequential by introducing a real economic incentive would arguably reduce such behavior. Thus:

Hypothesis 6. In a sample of N respondents:

$$\sum_{n=1}^{N} \gamma_{ni}^{hypothetical} > \sum_{n=1}^{N} \gamma_{ni}^{real}$$

The vector of non-attendance can either be inferred from the respondents' choices, or respondents can upon completion of the choice tasks be asked to state whether they ignored attributes. While these two approaches should theoretically provide similar results, several empirical investigations have found that non-attendance is higher when stated than when inferred (see e.g. Hess and Hensher 2013). Alemu et al. (2013) suggest that one reason is that the way stated non-attendance questions are typically articulated, they are not equivalent to inferred non-attendance. We propose that another reason relates to self-image. Especially in personal interviews and group interviews, social desirability bias and interviewer-bias might kick in and make respondents over- or understate their degree of non-attendance to attributes. In particular, for public goods attributes we might expect people to understate their actual degree of non-attendance simply because they believe it to be socially desirable, thus observing differences between stated and inferred non-attendance. For private good attributes, we would not expect a similar social desirability effect. However, as opposed to our hypothesis 6 stated above, it has been argued in the literature, that when merely asked to state non-attendance there is no incentive to provide a truthful answer, regardless of the setting being hypothetical or real (e.g. Hess and Hensher, 2010; Alemu et al., 2013; Carlsson et al., 2010). Instead it is suggested to rely on non-attendance being inferred from the model. Following this, individuals in the hypothetical setting would be able to consider all attributes in their choices

free of any costs, whereas in the real setting – choices and considerations of certain attributes would have actual economic consequences. By segregating the attributes into public and private good attributes, we take this argument further, arguing that if we consider e.g. public good attributes, social desirability might make respondents actually consider all public good attributes in the hypothetical setting, i.e. inferred non-attendance would be low. However, in the real setting, acting social desirable comes at a real economic cost. Hence, we might expect inferred non-attendance to reveal that respondents are actually ignoring more of the public good attributes in the real setting. If instead considering private good attributes, we would not expect any social desirability effects, and, thus, no differences in inferred non-attendance between the hypothetical and the real setting. This leads to the following hypotheses:

Hypothesis 7. For attributes that have public good character: $\sum \gamma_n^{hypothetical} < \sum \gamma_n^{real}$

Hypothesis 8. For attributes that have private good character: $\sum \gamma_n^{hypothetical} = \sum \gamma_n^{real}$

3. Experimental design and data

Data for this study was collected using the CE method in a survey of consumers' preferences for apples[4] where data was collected from two split samples. While each sample completed identical choice experiments, one split was based on a hypothetical stated choice experiment whereas a real economic incentive was introduced for the second split. In both samples, respondents were given a show up fee of 200 DKK to participate in the survey. In the real incentive sample, an additional cash payment was determined depending on how they chose during the 12 shopping scenarios in the choice experiment. In particular, respondents in this sample were provided with a voucher worth 45 DKK, which they could spend partly or entirely on one of the chosen alternatives. Thus, this element of the participation payment was determined depending on how they chose during the choice experiment. Similar to the respondent instructions used in Chang et al. (2009), Johansson-Stenman and Svedsäter (2012) and Grebitus et al. (2013), the following description was provided to the respondents:

[4] Since hypothetical bias is generally found to a lesser degree in market than non-market good valuation surveys, the choice of a low-price, low-involvement market good may not be an obvious choice if one is aiming to assess the impact of the economic incentive on hypothetical bias. However, our main purpose goes beyond hypothetical bias, focusing much more on other behavioral issues that have not yet been treated extensively in the literature. For the majority of these issues, e.g. attribute processing heuristics in terms of non-attendance, there is no evidence or a priori expectation that they would not be present for the type of good used here.

'After everyone completes all 12 shopping scenarios, we will ask for a volunteer 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. If the number 2 is drawn, the second 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 will be given a value ticket of DKK 45, which you should use for the purchase. The most expensive alternatives cost DKK 45. If you choose a cheaper alternative you will be given the remaining money. 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.'

The procedure of drawing one binding choice set for all individuals follows the procedure of Lusk and Schroeder (2004) – they drew only one binding scenario for the entire group of participants. The characteristics and their levels are shown in table 1.

The characteristics were based on results from focus group interviews and pilot testing of the questionnaire. The experimental design used is a D-efficient fractional factorial design resulting in 12 different choice sets in total. The software Ngene was used to generate the alternatives and the choice sets (Rose et al. 2009). The 12 choice sets consist of two generic alternatives plus a status quo alternative, the later representing a base bag of apples (Conventional mixed colour of sour and mealy apples produced outside EU to a price of DKK 7). Both surveys were collected in March 2011 using face-to-face group interviews. The overall sample consists of 106 respondents, with 36 respondents in the hypothetical sub-sample and 70 respondents in the incentivized sub-sample. All respondents were recruited from a consumer panel of people living in Copenhagen[5]. Moreover, a cheap talk script as well as a budget reminder was used in the hypothetical survey to resemble the usual practice in the stated preference literature. The primary purpose of the budget reminder in the hypothetical sample was to equate to the text concerning the consequences of not using all the money from the value ticket in the real incentives sample. In all other respects, the questionnaires used in the two samples were identical.

Table 1 The characteristics and their levels in the CE.

Characteristics	**Levels**
Type of production	Conventional, Organic

[5] All procedures involving human subjects were performed in compliance with relevant laws and institutional guidelines. Informed consent was obtained for experimentation with human subjects, and respondents were ensured that they would remain anonymous to any third parties.

Origin	Locally produce (Danish), Danish produce, European produce (not Danish), Produced outside Europe
Colour	Red, Green, Yellow, Mix of colours
Taste and texture	Sweet and crunchy, Sweet and mealy, Sour and crunchy, Sour and mealy
Price (DKK per kg)	7, 15, 25, 45

Note: DKK 10 ~ EUR 1.34

4. Econometrics

The model applied in the parametric analysis of responses is a mixed logit model which can be derived in a number of different ways (see Train 2003; Hensher and Greene 2003). In the present case, an Error Component Logit model representation of mixed logit was found suitable[6].

Following Scarpa et al. (2005) an Alternative Specific Constant (ASC) is specified for the status quo alternative in order to capture the systematic component of a potential status quo effect. Furthermore, an error component additional to the usual Gumbel-distributed error term is incorporated in the model to capture any remaining status quo effects in the stochastic part of utility. The error component, μ, which is implemented as a zero-mean normally distributed random parameter, is assigned exclusively to the two non-status quo alternatives. By specifying a common error component across these two alternatives, correlation patterns in utility over these alternatives is induced. Thus, the error component captures any additional variance associated with the cognitive effort of evaluating experimentally designed hypothetical alternatives (Brownstone and Train 1999, Herriges and Phaneuf 2002, Scarpa et al. 2005, Scarpa et al. 2008). This results in the following general utility structure:

$$U_{ntj} = \begin{cases} V(x_{ntj}, \beta, \mu) + \varepsilon_{ntj}, & j = 1,2; \\ V(ASC, x_{ntj}, \beta) + \varepsilon_{ntj}, & j = 3(optout) \end{cases} \quad (5)$$

[6] Employing random parameter error component logit (ECL) models might be more informative if one aims for knowledge about heterogeneity in preferences (Greene and Hensher 2007; Scarpa et al. 2007; Scarpa et al. 2008). However, when testing these models on our data, we found only very limited preference heterogeneity and no overall improvement of model fit when adjusting for added parameters. Hence, we base our main results on the somewhat simpler ECL model.

where the indirect utility, V, is a function of the vector of explanatory variables, x_{njt}, and associated parameters, β. For the two experimentally designed policy alternatives, the common error component μ enters the indirect utility function, while it is restricted to zero for the status quo alternative. The unobserved error term ε_{ntj} is assumed iid extreme value distributed. The individuals are denoted by n, while j is the alternative and t is the choice set. The probability of individual n choosing alternative k out of j alternatives can be defined by the Conditional Logit model:

$$P_{ntk} = \frac{e^{\lambda\left(\beta' x_{ntk} + ASC_k + \mu_k\right)}}{\sum_j^J e^{\lambda\left(\beta' x_{ntj} + ASC_j + \mu_j\right)}} \qquad (6)$$

where β' is a vector of all betas, λ is the scale parameter which is typically normalised to unity. Following Train (2003) and Scarpa et al. (2005), the probabilities of the ECL mixed logit type model can be described as integrals of the standard conditional logit function evaluated at different μ's with a density function as the mixing distribution. Furthermore, this specification can be generalised to allow for repeated choices by the same respondent, i.e. a panel structure, by letting z be a sequence of alternatives, one for each choice occasion, $z = \{z_1, \ldots, z_T\}$. Thus, the error component coefficients may vary over people but are constant over the T choice occasions for each individual. The marginal choice probability then becomes:

$$P_{nkz} = \int \left(\prod_{t=1}^{T} \left[\frac{e^{\beta' x_{ntk} + ASC_k + \mu_k}}{\sum_j^J e^{\beta' x_{ntj} + ASC_j + \mu_j}} \right] \right) \phi\left(\mu \middle| 0, \sigma^2\right) d\mu \qquad (7)$$

where $\phi\left(\mu \middle| 0, \sigma^2\right)$ is the normal density distribution function for μ.

This model specification is used for the main part of the analysis. We have used the software package Biogeme (Bierlaire 2003) to estimate the econometric models. In all models we control for individual level heterogeneity through the use of a panel specification capturing the repeated choice nature of the data. The models are estimated with simulated maximum likelihood using Halton draws with 300 replications; see Train (2003) for details on simulated maximum likelihood and Halton draws.

Since one of the purposes of the paper is also to look at a possible impact of using a real incentive on the behavioral decision processes and in particular non-attendance, we elaborate on the above general model by incorporating the mutually excluding states based on whether or not the respondent has stated to have ignored or attended to a given attribute. Hence, we can describe the row vector β' in equation (7) more generally as

$$\beta' = (\beta_{nk}^{consider}, \beta_{nk}^{ignore}) \tag{8}$$

In case β^{ignore} is equal to zero, this specification essentially resembles that of equation 4 in the sense that an ignored attribute will have no impact on the indirect utility and thus the choice probabilities estimated for an individual respondent. Following Campbell and Lorimer (2009) and Hess and Hensher (2010), in order to assess the accuracy of the stated attribute non-attendance we thus estimate a model with two states for each attribute, depending on whether or not the attribute is considered or ignored. In other terms, we essentially estimate two beta coefficients for each attribute; one for respondents who said they considered the attribute and one for respondents who said they ignored it. For the latter, if respondents have really ignored the attribute as they say they have, this coefficient should not differ significantly from zero. If it is indeed different from zero it would suggest that the respondents have not completely ignored it, implying that the statements of attribute non-attendance are not perfectly valid. In that case it may be argued that it is better to infer non-attendance from analytical models rather than relying on respondent statements. To continue our investigation in this direction, we follow the approach to inferring non-attendance suggested by Hess and Hensher (2010) that was also recently applied in Scarpa et al. (2013). This approach to inferring attribute non-attendance is based on the coefficient of variation of individual-specific posterior means. Since this applies to random parameter models, for this specific purpose we elaborate on the general model specification in eq. (7) by allowing for taste heterogeneity across respondents. Assuming normal distribution for all attributes, the multivariate normal distribution of individual tastes across individuals can be written as $\beta_n' \sim N(\beta, \Omega)$. The variance-covariance matrix Ω is specified so as to allow for correlation across random parameters, i.e. diagonal as well as off-diagonal elements are estimated in the model (Train and Weeks 2005; Scarpa et al. 2008). Individual-specific conditional distributions of the random parameters can be derived from the estimated population parameters (Huber and Train 2001). Based on these conditional distributions which assume a normal distributed random parameter for each individual for each attribute, i.e. $\beta_{nk} \sim N(\beta_{nk}, \sigma_{nk}^2)$, the coefficient of variation can be obtained as the ratio $\kappa_{nk} = \frac{\sigma_{nk}}{\beta_{nk}}$. As noted by Hess and Hensher (2010) and Scarpa et al. (2012), a value of 2 for κ_{nk} would seem like a reasonable threshold value above which non-attendance observations can be expected since normal

distributions with spread-to-mean ratios above 2 are over-dispersed[7]. In other words, when $\kappa_{nk}>2$ the noise-to-signal ratio can be considered high, and it may thus be inferred that individual n has ignored attribute k. The share of individuals ignoring a given attribute can then be obtained by aggregating over the individuals.

5. Results

We initially identified protesters in the two splits based on responses to follow-up questions designed for this purpose. Respondents were classified as protesters if they had chosen bag 3 in all 12 choice tasks and moreover stated one of the following reasons for doing so: "Although I want a better bag of apples, I do not think that it is reasonable that it is me as a consumer who has to pay for it"; "I do not trust that I would actually get a better bag of apples even though I pay more"; "I do not think I got enough information to choose"; "I think it was too hard to choose". However, most likely due to the fact that we are dealing with a quite common and relatively low-priced market good, the numbers of protesters were low compared to what is often found in surveys considering e.g. nonmarket goods. Nevertheless, we find that the introduction of the real incentive reduces the occurrence of protest behaviour: while a protest rate of 11.1% (4 respondents) was found in the hypothetical sample, the equivalent number was significantly lower at only 1.4% (1 respondent) in the real sample (a χ^2_{Yates}-test comparing protest frequencies across the two samples confirms that this tendency is significant (p-value = 0.0170)[8] – implying that we can confirm hypothesis 3.

Another behavioural impact that we may expect from introducing a real economic incentive, is concerned with the frequency with which respondents opt out of choices by choosing the opt-out alternative. Considering hypothesis 4, our data confirms that the share of occasional status quo choices is lower in the real incentive sample than in the hypothetical sample. In the hypothetical sample, 12.5% of all choices made are opt-out choices whereas it is only 5.7% in the sample with real incentive. A χ^2_{Yates}-test comparing choice frequencies across the two samples confirms that this tendency is highly significant (p-value < 0.001).

Furthermore, the response time for each choice task is examined in the two data sets. In the real incentives sample, respondents spent on average 177 seconds reading the scenario description while

[7] Hess and Hensher (2010) and Scarpa et al. (2012) note that the choice of threshold value is a delicate and somewhat arbitrary choice which could warrant further investigation. This is however beyond the scope of this paper.
[8] This and the following chi-square tests have been calculated based on actual and expected numbers – not percentages.

respondents in the hypothetical sample spent 159. Though not a strong statistically significant difference, this is in accordance with hypothesis 5, and it is also a reassuring result since it suggests that the scenario descriptions were actually read. Somewhat surprisingly, and in contrast to our hypothesis 5, we find that the average response time for the 12 choice sets is slightly lower in the real incentive sample (273 seconds) than in the hypothetical sample (288 seconds) – though this difference is far from statistically significant. However, considering the results below, it may not be so surprising that they can answer slightly faster since they appear to be also ignoring more attributes.

Turning to stated attribute non-attendance, table 2 reports the presence of such behaviour in the data. The non-attendance question was asked at the end of the choice sequence- Respondents were asked to state whether they consistently had ignored some of the attributes in the 12 choices they had just made, and, in positive, which attributes they had ignored.[9] While non-attendance is clearly not a behavioural phenomenon confined only to stated preference surveys, we hypothesize that respondent to a lower degree resort to such behaviour when choices are more consequential (hypothesis 6). However, to our surprise, we cannot confirm this. In fact, when relying on stated non-attendance, we find the opposite – suggesting that the hypothesis should be rejected.

Table 2 Stated non-attendance

	Split 1 - Hypothetical		Split 2 – Real	
"Have you ignored any attributes?"	Number	Share	Number	Share
- Yes	18	60%	52	75%
- No	12	40%	17	25%
- Total	30		69	
Number of attributes ignored[a]				
- 0	12	40%	18	26%
- 1	10	33%	14	20%
- 2	5	17%	14	20%

[9] Non-attendance questions could also have been posed after each choice set, but this was discharged due to limitation of the length of the questionnaire.

- 3	1	3%	6	9%
- 4	2	7%	17	25%
Attributes ignored[a]				
- Type of production	6	20%	26	38%
- Origin	6	20%	24	35%
- Color	9	30%	45	65%
- Taste and texture	6	20%	8	12%
- Price	4	13%	25	36%

[a] One respondent in split 2 stated to have ignored attributes, but did not answer which attribute.

Rather than reducing non-attendance behavior, introducing the real incentive in our case actually increases the share of non-attenders from 60 to 75 percent (χ^2_{Yates}-test, p-value=0.0817). Moreover, it is evident that the tendency to ignore more than one attribute is more outspoken in the real incentive sample where as much as 25% of the respondents state to have ignored four attributes. A reason might be that when faced with a real economic incentive, more respondents would focus solely on the price attribute - this appears however not to be the case. In fact, 16 of the 17 respondents stating to have ignored four attributes in the real incentive sample claim that one of the ignored attributes was the price. The only attribute which all these 16 respondents state to have considered in their choices is the taste and texture attribute. In other words, their choices would seem to be governed by some elimination-by-aspects heuristic or a lexicographic decision rule. This behavior is much less present in the hypothetical sample. This somewhat surprising difference is further evident when looking at how often each of the attributes has been ignored. The table shows that price is ignored by 13% of the respondents in the hypothetical sample. It has often been argued that it is the hypothetical nature of CE surveys that leads to price-ignorance which has direct consequences for WTP and welfare measure estimations, which are strictly speaking invalid if respondents are not sensitive to price at all. However, our findings suggest that it is not the hypothetical nature of the survey method as such that instigates price insensitivity, because results show that in the real incentives sample as much as 36% of respondents state they have ignored price. While price insensitivity is of course not desirable, it is somewhat reassuring in the hypothetical sample that price is the least ignored of all attributes. This is not the case in the real incentive sample where the taste and texture attribute is clearly the least ignored. Compared to the hypothetical sample, all other attributes than taste and texture are ignored to a larger extent in the real incentives sample. Assuming that the real incentive does make respondents consider their choices more and that it does make them choose more in accordance with their actual preference structure, this would suggest that taste and texture is clearly the most important attribute of an apple (at least in the way it is framed in the study), but in hypothetical situations respondents tend to

underestimate this. This partly supports hypothesis 8, that attributes with private good characteristics are considered more frequently than their public good counterparts. This could go along with the explanation of interviewer bias, i.e. if respondents think that the interviewer expects them to take all attributes into account, since this comes at zero costs for the respondents due to the hypothetical setting. It is not necessarily a methodological problem that respondents tend to focus their attention on one particular attribute and ignore other quality attributes since this could simply reflect that they have genuine zero-preferences for the other attributes (Alemu et al. 2013). However, economic theory would preclude zero-preferences for the price attribute. With regard to this theoretically counterintuitive increase in price insensitivity when choices bear a real economic consequence, one obvious explanation could be that respondents are spending windfall money (see e.g. Carlsson et al. 2012). Furthermore, while the maximum price of 45 DKK is a very high price for a kilo of apples, the economic impact on the individual's total household budget is negligible, potentially to the extent where some respondents will completely ignore, especially if they have very strong preferences for another attribute. Looking at the attributes with more public good characteristics, like type of production and origin, the degree of non-attendance is larger in the real incentives sample than in the hypothetical sample.

5.1 Parametric results

Table 3 reports the results of the estimated models. Initial investigations concerning the demographic distributions in the two samples showed significant differences for gender, but not for age, household income, number of children and education level[10]. Hence, to ensure comparability, in all the following models for both split samples observations are weighted based on a propensity-based weighting procedure in order to fit the actual gender and age distributions in the target population, namely Copenhagen municipality.

Table 3 Error Component model results including LR test statistics.

	Split 1 - hypothetical		Split 2 – real incentive		Pooled without scale correction		Pooled with scale correction	
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value

[10] Results from chi-square tests can be obtained from the authors on request.

	(Std. Err.)		(Std. Err.)		(Std. Err.)		(Std. Err.)	
ASC SQ	-1.82		-2.41		-2.28		-0.797	
		1.90		3.90		4.35		3.74
	(0.957)		(0.618)		(0.524)		(0.213)	
Organic produce	0.599		0.685		0.656		0.23	
		6.67		8.34		10.61		7.40
	(0.0898)		(0.0821)		(0.0618)		(0.0311)	
Local produce within Denmark	1.19		1.37		1.31		0.459	
		6.30		8.20		10.40		7.61
	(0.189)		(0.167)		(0.126)		(0.0603)	
Danish produce	0.293		0.315		-0.126		-0.0472	
		1.24		1.96		0.96		1.02
	(0.237)		(0.161)		(0.131)		(0.0463)	
Green coloured apples	0.788		0.462		0.55		0.191	
		4.96		2.98		4.51		4.59
	(0.159)		(0.155)		(0.122)		(0.0416)	
Yellow coloured apples	1.84		1.66		1.7		0.594	
		9.15		7.00		9.55		7.80
	(0.201)		(0.237)		(0.178)		(0.0762)	
Red coloured apples	1.33		1.07		1.13		0.395	
		5.71		4.59		6.24		5.73
	(0.233)		(0.233)		(0.181)		(0.0689)	
Sweet and mealy apples	-0.0683		0.242		0.14		0.0504	
		0.23		1.28		0.88		0.92
	(0.293)		(0.189)		(0.159)		(0.0549)	
Price	-0.0532		-0.0359		-0.0412		-0.0143	
		4.55		5.53		7.24		0.26
	(0.0117)		(0.00649)		(0.00569)		(0.0549)	
Error Component	2.42		1.95		2.09		0.731	
		4.14		6.61		6.68		5.80
	(0.585)		(0.295)		(0.313)		(0.126)	
Scale (hypo. fixed to 1 – t-value relative to 1)							1.07	
								0.44
							(0.159)	
N	359		828		1187		1187	
LL at convergence	-249.1		-524.7		-783.6		-783.4	
Pseudo R^2	0.344		0.412		0.392		0.391	
LR-test statistics					-19.54		-19.17	

Note: As can be seen from comparing the attributes in table 1 and the estimation in table 3, for reasons of simplicity we have merged some of the attribute levels since they were not significantly different from each other. Moreover we have tested for taste heterogeneity, but found none, which potentially could be due to the relatively small samples.

It is evident from the table that the specified model describes the variation in both samples reasonably well. Due to the confounding scale parameter, parameter estimates cannot be compared directly across the two samples. However, looking at signs and significance levels there seem to be only minor differences. While different signs are found for the "Sweet and mealy apples"-attribute in the two samples, this parameter estimate is not significant in any of the samples. Another slight difference is the "Danish produce" parameter which as expected from previous surveys and previous analysis of data from the same survey is positive, but only significantly so in the real incentives sample.

The LR-test statistic not accounting for potential differences in error variance across samples suggests that the overall preference structures in the two samples do indeed differ significantly (critical value at 5% and 9 df. is 16.92), and adjusting for differences in scale, the LR-test result remains significant. Recall that the scale parameter is inversely proportional to the standard deviation of the error term in our specification (Swait and Louviere 1993). The estimated relative scale factor of 1.07 in Table 3 implies that the variance of the error term or "noise" in the model based on the real incentives sample is only 87% of the variance of the model based on the hypothetical sample[11]. Even though it is not statistically significant different from one, this reduction of the error variance is in accordance with hypothesis 5 concerning respondents making more consistent and well-considered choices when facing a real economic incentive, but taking the insignificance as face value, the result suggests that choices are as consistent and well-considered when real economic incentives are present as when choices are purely hypothetical. Potentially, this could be explained by the nature of the good being very familiar to the respondents, thereby making it easier to behave in a consistent manner, both cognitively and with respect to preferences, respectively.

Table 4 Unconditional WTP estimates based on models (i) and (ii).

	Split 1 – hypothetical		Split 2 – real		
	WTP	Std. Err.	WTP	Std. Err.	t-value
ASC SQ	-34.2	15.3	-67.1	21.7	1.24
Organic produce	11.3	3.7	19.1	5.2	1.22
Local produce within Denmark	22.4	7.0	38.2	7.4	1.54
Danish produce	5.5	4.3	8.8	5.7	0.46

[11] $(1/1.07^2) = 0.873$. Hence, there is a 12.7% difference in the overall unexplained variance between the two models.

Green coloured apples	14.8	4.8	12.9	4.7	0.29
Yellow coloured apples	34.6	10.0	46.2	9.2	0.86
Red coloured apples	25.0	8.8	29.8	8.1	0.40
Sweet and mealy apples	-1.3	6.4	6.7	6.2	0.90

Since the scale parameter cancels out in WTP calculations, the WTP estimates in table 4 are directly comparable across the two samples. Even though none of the differences in WTP are statistically significant, it is interesting to note that almost all quality attribute WTP estimates are higher in the real incentives sample suggesting an overall higher WTP in this sample. Moreover it is noteworthy that the attributes that have public good character, i.e. organic produce and local and Danish produce, seems larger in the real incentives sample than in the hypothetical sample, contrasting our hypothesis 1, though not significant. With respect to the private good characteristics – we cannot reject our hypothesis 2 of equal marginal WTP between the two samples. In general, this is in contrast to previous findings concerning hypothetical bias, but a potential explanation of this could be found in the setting of the experiment. Since the experiment were designed as face-to-face group interviews, respondents could have been influenced by the presence of the other respondents, thus not revealing their individual 'true' preferences, but rather exhibit some kind of warm glow behaviour (Andreoni 1989) or wanting to 'show-off' to the other respondents, exhibiting social desirability behavior (Levy 1981). However, this remains rather speculative in the case of WTP. The results of higher WTP in the real incentives sample is countered by a more negative WTP associated with the status quo alternative in the real incentive sample, indicating that respondents in this sample take a stronger dislike to this default bag of apples.

Table 5 Error component models separating considered and ignored attributes

	Split 1 - hypothetical				Split 2 – real incentive			
	Considered		Ignored		Considered		Ignored	
	Coefficient (Std. Err.)	t-value	Coefficient (Std. Err.)	t-value	Coefficient (Std. Err.)	t-value	Coefficient (Std. Err.)	t-value
ASC SQ	-2.26 1.00	-2.26			-2.45 0.642	-3.82		
Organic produce	0.553 0.0718	7.70	0.325 0.202	1.61	0.605 0.102	5.93	0.493 0.185	2.66

Local produce within Denmark	1.2	5.74	0.214	0.49	1.16	6.74	1.06	2.57
	0.209		0.436		0.172		0.412	
Danish produce	0.289	1.02	0.0642	0.12	-0.139	-0.69	-0.816	-2.44
	0.284		0.517		0.201		0.334	
Green coloured apples	0.63	2.99	0.75	2.28	0.497	1.85	0.0664	0.19
	0.211		0.329		0.269		0.356	
Yellow coloured apples	1.93	7.04	-0.001	< 0.001	1.5	6.02	0.502	1.59
	0.274		0.368		0.249		0.315	
Red coloured apples	1.23	4.33	0.639	1.35	0.99	4.34	0.36	1.02
	0.284		0.472		0.228		0.354	
Sweet and mealy apples	0.218	0.74	-1.69	-4.41	0.414	2.12	-2.11	-3.23
	0.293		0.383		0.195		0.654	
Price	-0.0565	-4.28	-0.00019	-0.01	-0.0392	-4.78	-0.00333	-0.42
	0.0132		0.025		0.0082		0.00784	
Error Component	2.55	4.02			1.95	6.25		
	0.635				0.312			
N	359				828			
LL	-240				-491			
Pseudo R^2	0.346				0.435			

Going back to behavioural issues, we also examine stated attribute non-attendance from a parametric point of view. We follow the line of Campbell and Lorimer (2009), Hess and Hensher (2010) and Alemu et al. (2013), who examine attribute non-attendance in hypothetical settings and estimate two separate coefficients for each attribute, depending on whether the attribute is stated to have been considered or ignored. The reason for doing so is to test whether the ignored attribute represents a genuine zero value for the respondent or whether they actually have taken the attribute into consideration despite their opposite statement. Ideally, if an attribute has been truly ignored, its impact on choices should be insignificant. The results are shown for the two sample in table 5.

As the results show, even though respondents have stated that they have ignored certain attributes, some of them have actually had a significant effect on the choice probability according to the model. In the hypothetical setting two out of the eight attributes/levels which the respondents have stated to have ignored come out significant, whereas this proportion doubles in the survey with the real incentives. This is in line with Campbell and Lorimer (2009), Hess and Hensher (2010) and Hess et al. (2012) who find that respondents who state that they have ignored an attribute might in fact have paid some attention to it in their choice behavior. Moreover, the results show that respondents stating to have ignored the price appear to have actually been telling the truth. As described earlier this is in contrast to economic theory, though it may be explained by the use of windfall money at least in the real incentives sample. If respondents do indeed not consider the windfall money as part of their own household budget, this could also explain why 36% of the respondent state to have ignored price in the real incentives sample – a much larger share than the 13% in the hypothetical sample. On the other hand, if we had seen respondents always choosing the cheapest alternative, this could suggest, that respondents realizing that the money (the vouchers) has value in the real-world market outside these lab settings would maximize utility by saving some of the 'voucher' money from the apples and spending them on some other preferred goods. The differences between the hypothetical and the real incentives setting could suggest some behavioral aspects concerning respondents' decision making strategies; In the hypothetical setting it seems that respondents stating to have ignored an attribute in fact have ignored it in the majority of cases (6 out of 8), implying a zero contribution to the choice probabilities. This is in contrast to the findings in the real incentives sample where half of the (stated) ignored attributes do in fact have an effect on respondents' choices, and, hence, have a significant impact on the choice probabilities. This is also partially in contrast to the results showing that respondents in the real incentives sample have ignored attributes more frequently than respondents in the hypothetical sample. This could suggest that respondents use these non-attendance statements in another way than originally intended and interpreted by the analyst. Originally, these statements were intended to help identify attributes which respondents have truly ignored when making their choices. While they might do so for a number of reasons more or less in agreement with the basic assumptions concerning *homo economicus* (Alemu et al. 2013), failing to account for nonattendance could lead to biased welfare measures. While Balcombe et al. (2011) find that respondents actually ignore attributes stated to be ignored, Campbell and Lorimer (2009), Hess and Hensher (2010) and Alemu et al. (2013), found that respondents do not completely ignore an attribute even though they say so. Contrary to our expectations, our results suggest that this effect is even more pronounced in surveys with real incentives[12]. Of course, the really important question here is whether or not the significant estimates obtained from attributes which respondents claim they have ignored, can be given a meaningful behavioral interpretation? Or perhaps even more interesting; do the estimates represent preferences? Hess et al. (2012), Alemu et al. (2013) and Carlsson et al. (2010) suggest that rather than ignoring attributes completely, respondents might simply put less weight on attributes they claim to have ignored. They base this on the fact that the most often ignored non-monetary attribute also receives the lowest preference ranking in their estimated utility model. However, this seems to not be the

[12] We note that the findings of a less pronounced effect of the non-zero value estimates for the ignored attributes in the hypothetical sample could be caused by the relative small sample size.

case in our survey. Finally, the contradicting results, that respondents state a higher degree of non-attendance in the real incentives sample (75% vs. 60%), while at the same time the parametric model shows that half of the ignored attributes do apparently contribute to the choice probabilities, could suggest that even though we have introduced real incentives, and thus should have removed hypothetical bias from the preferences, the non-attendance statements are prone to this.

Since the statements of non-attendance are maybe not perfectly valid, we further investigate the attribute non-attendance issue by inferring it from analytical models rather than relying on the self-reported statements. Here we follow Hess and Hensher (2010) and Scarpa et al. (2012) and infer non-attendance by estimating individual-specific coefficients of variation based on a random parameter logit model that allows for correlation across attributes[13]. Table 6 reports the main findings of interest in terms of the shares of respondents that are found to ignore attributes as indicated by the coefficient of variance exceeding the chosen threshold value of 2.

Table 6 Inferred (and stated) non-attendance

	Split 1 - Hypothetical		Split 2 – Real	
Attributes ignored	No.	Shares	No.	Shares
- Type of production	3 (6)	10% (20%)	7 (26)	10% (38%)
- Origin	3 (6)	10% (20%)	14 (24)	20% (35%)
- Color	2 (9)	7% (30%)	5 (45)	7% (65%)
- Taste and texture	0 (6)	0% (20%)	7 (8)	10% (12%)
- Price	2 (4)	7% (13%)	2 (25)	3% (36%)

Note: Numbers in parentheses report the stated non-attendance for comparison (previously shown in table 2).

It is apparent from table 6 that inferred non-attendance is generally much lower than stated non-attendance. This would suggest that respondents do indeed tend to exaggerate their stated non-attendance, or at least they interpret non-attendance in a wider sense than analysts typically do (Alemu et al. 2013). Examining the specific attributes, the results for public good character attributes like type of production and origin, are only partly in support of hypothesis 7, whereas the results of the private good character attributes, color, taste and texture (and price), are in contrast to

[13] See appendix A for model results.

hypothesis 8. It is though somewhat reassuring that the inferred approach suggests that only 3% of the respondent in the real treatment ignore the price attribute, while this is slightly higher at 7% in the hypothetical treatment. There is a remarkable difference in the extent to which respondents faced with a real incentive say that they ignore price (36%) and what the data and model tells us (3%). A behavioral interpretation could be that the respondents feel a social responsibility to participate in a survey like this (Loureiro and Lotade 2005). However, when offered real money to do so, they are subject to a sort of social desirability bias or interviewer bias in terms of wanting to signal that the money did not make a difference—at least when asked directly. While this might apply to the price attribute, it is difficult to see a similar behavioral argument explaining the huge difference between stated and inferred non-attendance for the Color attribute. It should of course be remembered that the inferred non-attendance approach should not necessarily be regarded as the 'true' non-attendance against which stated non-attendance should be benchmarked. As this approach rests on a number of modeling assumptions, e.g. the chosen threshold level for κ, as well as the quality of the data available in the given case, it is possible that this approach underestimates the 'true' level of non-attendance and, thus, exaggerating the differences in table 6. However, we have no way of assessing whether this might be the case.

6. Conclusion and discussion

In this paper we continue the line of research concerning impacts of real economic incentives in CE by investigating whether other behavioural aspects differ between hypothetical surveys and surveys with real incentives, and whether this potential can be explained by self-image impacts. More specifically, in a market good Choice Experiment setting we examine the potential impact of introducing a real economic incentive on preferences and WTP estimates as well as on respondents' decision making and attribute processing strategies. While the former has received a fair amount of attention in the literature, few have really focused on the latter. Besides focusing on attribute non-attendance, we also investigate impacts on error variance, protest behaviour, opting out of choices, and response latencies in order to further assess the influence of real incentives on the decision making process.

Based on an empirical case considering a relatively low-priced market good that is well-known to most consumers, namely apples, using a split sample approach we find that introducing a real economic incentive to a treatment sample in a CE survey has no significant effect on preferences and WTP compared to a control sample receiving an otherwise identical but purely hypothetical CE. We hypothesized that self-image effects could lead to stronger impacts of the real incentive on WTPs for attributes that have mainly public good character as compared to attributes with more private good character. We are however not able to find support for this hypothesis. One

explanation could be that hypothetical bias is completely absent in our empirical case. Hence, it would be interesting to further test this hypothesis in future research concerning non-market good cases where hypothetical bias is likely to be more present. Another explanation could relate to free-riding. Even though respondents might initially have a high WTP for a public good attribute due to the ethical dimension of self-image, this may be countered by another motivation in terms of free-riding. The individual can still enjoy the public good aspects of e.g. organic farming even if he is not buying organic products – and, furthermore, he might realize that others would have the same opportunity. This represent an incentive to free-ride which would tend to reduce the individual's WTP. Yet another explanation could be that the respondents do not perceive a given attribute to have as much public good character as the analyst would objectively assign to it, or vice versa. For instance, we have assumed that the "organic production" attribute in our case is primarily considered to have public good character in terms of reduced negative externalities from the production compared to conventional agricultural production, whereas private good traits like taste, nutrition and food safety are much less relevant since the objective differences between organically and conventionally produced food is not large. However, the respondents may see it differently if e.g. they believe that organically produced food is much healthier or taste better than conventionally produced food. If indeed respondents perceive the attribute as having mainly private good character, the ethical component of self-image effects would be less prevalent than we expect, and the WTP would thus be lowered.

Interestingly, we find marked impacts of the real economic incentive on issues pertaining to the behavioural decision making underlying respondents' choices. First of all, our data shows that introduction of the real incentive leads to a reduction in the share of respondents resorting to protest behaviour. If confirmed in future surveys, this tendency is clearly an additional benefit that should be taken into account when researchers in the early survey design phase is considering whether or not to use real incentives. As for response latency, unsurprisingly we find that respondents being subjected to the real incentive treatment generally spent longer time reading the scenario description preceding the choice tasks. Slightly surprising though, they were faster at answering the choice task than respondents in the hypothetical treatment sample. While one explanation could relate to the fact that they had more carefully read the scenario description and, thus, had a better idea of what the following choice tasks were about, another explanation could be that they were simple more prone to resort to simplifying decision making strategies that could make choices easier—though our *a priori* expectation was the opposite. One indication that would seem to weaken the latter suspicion is the fact that respondents in the real incentive treatment were significantly less prone to choose the opt-out alternative, i.e. they are apparently not resorting to the opt-out alternative as a simplifying heuristic, at least not to as large an extent as respondents in the hypothetical treatment. Rather, it would seem that another behavioural benefit of using real incentives could be a reduced tendency to use opting out as a heuristic. This is however in opposition to results in Lusk and Schroeder (2004), so more research on this issue is warranted. Another decision-making heuristic that could be causing the faster choice task responses obtained from respondents receiving the real incentive would be if respondents just made random choices, i.e. "quick-and-dirty" choices

(Campbell et al. 2012). Though not a significant difference, the fact that we find numerically lower error variance in the real incentives sample would suggest that this is however not the case. On the contrary, this would suggest that another behavioural benefit associated with the introduction of real incentives in CE might be reduced randomness in choices.

Surprisingly, looking at respondents' statements of attribute non-attendance we find what at first sight would seem to be a serious deficiency caused by the introduction of the real incentive: Rates of stated non-attendance are much higher for all but one attribute in the real incentive sample. Most notably, in this sample 36% of the respondents have stated non-attendance to the price attribute. If indeed one out of three respondents were ignoring the price attribute, consequences could be dire since this would result in a very low price parameter estimate that could potentially explode WTP estimates when entering the denominator. In the hypothetical sample the corresponding share ignoring price was only 13%. Considering that the real economic incentive is intended to direct respondents' attention to the monetary trade-offs, if anything, we would have expected the real incentive to reduce non-attendance to the price attribute rather than the opposite. The fact that we do not find extremely high WTP estimates in the real incentive sample suggests that respondents may have exaggerated their non-attendance statements. Moreover, comparing the results for the pure private good characteristic – taste and texture, and the public good characteristics – type of production and origin, the degree of stated non-attendance differ. For the attributes that have public good character the degree of stated non-attendance is larger in the real incentives sample and lower for the attributes that have private good character compared to the hypothetical setting.

Turning to the non-attendance that can be inferred from the analytical models based on respondents' actual choices, this is confirmed. Inferred non-attendance generally finds lower rates of non-attendance. Especially for the price attribute, the inferred non-attendance is lower in the real incentives sample than in the hypothetical sample. Moreover, for attributes with public good character, the results show larger degrees of inferred non-attendance in the real incentives sample, which is not in support of our hypothesis of self-image impacts in public good character attributes. However, when considering the private good character attributes, the similarities of the degree of inferred non-attendance between the hypothetical and the real incentives sample do support our hypothesis of self-image impacts within private good character attributes. If generalizable in future studies, this potential to reduce price non-attendance could also be considered an important behavioral benefit associated with using real economic incentives in CE.

The question of whether we can trust respondents' statements with respect to non-attendance or whether the inferred non-attendance approach is a more suitable way of identifying non-attendance is not settled in the literature even though it would seem that there is growing skepticism regarding the use of stated non-attendance. In line with others comparing statements of attribute non-

attendance with that inferred from analytical models, our findings confirm that respondents generally tend to exaggerate how often they ignore attributes, or at least they interpret the non-attendance question in a wider sense than analysts typically do (Alemu et al. 2013). Somewhat surprisingly however, we find that this tendency is markedly amplified when introducing a real economic incentive. Maybe most surprising is the fact that while more than one out of three respondents in the real incentives sample state that they have ignored the price attribute, inferring non-attendance from the analytical models suggests that it is less than one out of thirty respondents who have actually ignored price. In other words, the difference is in the magnitude of a factor ten. In the hypothetical sample price non-attendance is 'only' exaggerated by a factor of two. A possible explanation of this surprising result could be a sort of social desirability bias or interviewer bias in the non-attendance statements, but more research on the behavioral reasons underlying price non-attendance statements is needed in order to properly investigate this. Nevertheless, if we assume that inferred non-attendance is the appropriate measure against which stated non-attendance should be benchmarked, our results suggest that researchers should be relatively more skeptical towards relying on respondents' statements of non-attendance when real economic incentives are provided, because as we show, these could also be prone to hypothetical bias – even when real incentives are provided.

References

Alemu, M. H., M. R. Mørkbak, S. B. Olsen, and C. L. Jensen. (2013). Attending to the reasons for attribute non-attendance in Choice Experiments. *Environmental and Resource Economics* 54(3):333-359

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):1050-61

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, XCVII:1447–1458

Balcombe, K., M. Burton, and D. Rigby. (2011). Skew and attribute non-attendance within the Bayesian mixed logit model. *Journal of Environmental Economics and Management* 62:446-461.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Broadbent, C. D., J. B. Grandy, and R. P. Berrens. (2010). Testing for hypothetical bias in a choice experiment using a local public good: riparian forest restoration. *International Journal of Ecological Economics* 19(F10)

Brownstone, D., and K. Train (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* 89:109-129

Cameron, T. A., G. L. Poe, R. G. Ethier, and W. D. Schulze. (2002). Alternative non-market value-elicitation methods: Are the underlying preferences the same? *Journal of Environmental Economics and Management* 44:391-425.

Campbell, D., G. Hutchinson, and R. Scarpa. (2008). Incorporating discontinuous preferences in to the analysis of discrete choice experiments. *Environmental and Resource Economics* 41:401 – 417

Campbell, D., and V. S. Lorimer. (2009). *Accommodating attribute processing strategies in stated choice analysis: do respondents do what they say they do?* European association of environmental and resource economists annual conference, Amsterdam.

Campbell, D., M. R. Mørkbak, and S. B. Olsen. (2012). *Response latency in stated choice experiments: impact on preference, variance and processing heterogeneity*. Manuscript under review.

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

Carlsson, F., M. Kataria, and E. Lampi. (2010). Dealing with ignored attributes in choice experiments on valuation of Sweden's environmental quality objectives. *Environmental and Resource Economics* 47:65-89.

Carlsson, F., and P. Martinsson. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments? *Journal of Environmental Economics and Management* 41:179-192.

Carlsson, F., H. He and P. Martinsson. (2012). Easy come, easy go - The role of windfall money in lab and field experiments, *Experimental Economics*, Forthcoming

Carson, R. T., N. E. Flores, and N. F. Meade. (2001). Contingent Valuation: Controversies and Evidence. *Environmental and Resource Economics* 19:173-210.

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

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

Gowdy, J., M., and K. Mayumi. (2001). Reformulating the foundations of consumer choice theory and environmental valuation. *Ecological Economics* 39:223-237.

Gracia, A., M. L. Loureiro, and R. M. jr. Nayga. (2011). Are valuation from nonhypothetical choice experiments different from those of experimental auctions? *American Journal of Agricultural Economics* 93(5):1358-1373.

Gravelle, H., and R. Rees. (1992). *Microeconomics*. Harlow, England: Pearson Education Ltd.

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

Greene, W. H. (2003). *Econometric Analysis*. New York: Prentice-Hall International, Inc.

Greene, W. H., and D. A. Hensher. (2007). Heteroscedastic control for random coefficients and error components in mixed logit. *Transport Resesearch part E* 43(5):610-623

Hensher, D.A., and W. H. Greene. (2003). The mixed logit model: the state of practice. *Transportation* 30(2):133–176

Hensher, D. A., J. M. Rose, and W. H. Greene. (2005). The Implication on willingness to pay of respondents ignoring specific attributes. *Transportation* 32:203 – 222.

Hensher, D.A., and J. M. Rose. (2009). Simplifying choice through attribute preservation or non-attendance: Implication for willingness to pay. *Transportation Research Part E* 45:583-590.

Hensher, D. A. (2007). *Attribute Processing in Choice Experiments and Implications on Willingness to Pay*. In Kanninen, B. J. (ed) *Valuing Environmental Amenities Using Stated Choice Studies*, Springer Netherlands, pp. 135-157.

Hensher, D.A. (2010). Attribute Processing, Heuristics and Preference Construction in Choice Analysis. (*Invitational Keynote Paper for Choice Modelling Conference*, Leeds UK. March 30-

April 1 2009, linked ARC DP 0770618; 07-09). in Hess, S. and Daly, A. (eds.) *State-of Art and State-of Practice in Choice Modelling*, Emerald Press, UK., 35-70.

Herriges, J. A., and D. Phaneuf. (2002). Inducing patterns of correlation and substitution in repeated nested logit models of recreation demand. *American Journal of Agricultural Economics* 84(4):1076-1090

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

Hess, S., and D. Hensher. (2013). Making use of respondent reported processing information to understand attribute importance: a latent variable scaling approach. *Transportation* 40(2):397-412

Hess, S., A. Stathopoulos, D. Campbell, V. O'Neill, and S. Caussade. (2012). It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. *Transportation*, forthcoming

Hole, A.R., J. R. Kolstad, and D. Gyrd-Hansen. (2012). *Inferred vs stated attribute non-attendance in choice experiments: a study of doctors' prescription behaviour*. Sheffield Economic Research Paper Series. SERP number 2012010. University of Sheffield, United Kingdom.

Huber, J., and K. Train. (2001). On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths. *Marketing Letters* 12(3):259- 269

Johansson-Stenman, O., and H. Svedsäter. (2012). Self-image and valuation of moral goods: Stated versus actual willingness to pay. *Journal of Economic Behavior & Organization* 84(3):879-891

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.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives* 5(1):193-206

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *The Journal of Political Economy* 74(2):132-157.

Levy, S. J. (1981). Interpreting Consumer Mythology: A Structural Approach to Consumer Behavior. *Journal of Marketing* 45:49-61

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

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.

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(3): 697-712.

Loureiro, M.L., and J.Lotade. (2005). Interviewer Effects on the Valuation of Goods with Ethical and Environmental Attributes. *Environmental and Resource Economics* 30:49-72

Luce, R. D. (1959). *Individual choice behaviour*. New York: Wiley.

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(2):467-482.

McFadden, D. (1974). *Conditional Logit Analysis of Qualitative Choice Behavior*. In: P. Zarembka (ed) Frontiers in Econometrics. Academic, New York.

Mentzakis, E., and J. Zhang. (2012). *An investigation of individual preferences: consistency across incentives and stability over time*. Working paper no. 70. University of Zurich. Department of Economics, Working Paper Series

Meyerhoff, J., Mørkbak, M.R., & Olsen, S.B. (2013): A meta-study investigating the sources of protest behavior in stated preference surveys. ***Environmental and Resource Economics.*** DOI 10.1007/s10640-013-9688-1**.**

Mitchell, R.C., and R. T. Carson. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Johns Hopkins University Press, Baltimore, MD.

Mjelde, J.W., Y. H. Jin, C. K. Lee, T. K. Kim, and S. Y. Han. (2012). Development of a bias ratio to examine factors influencing hypothetical bias. *Journal of Environmental Management* 95:39-48.

Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead. (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics* 30:313-325.

Neill, H. R., R. G. Cummings, P. T. Ganderton, and G. W. Harrison. (1994). Hypothetical surveys and real economic commitments. *Land Economics* 70(2):145-154.

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

Rose J. M., A. T. Collins, M. C. Bliemer, and D. A. Hensher. (2009). Ngene 1.0 Stated Choice Experiment Design Software, University of Sydney.

Rosenberger, R. S., G. L. Peterson, A. Clarke, and T. C. Brown. (2003). Measuring dispositions for lexicographic preferences of environmental goods: integrating economics, psychology and ethics. *Ecological Economics* 44:63-76.

Scarpa, R., S. Ferrini, and K. G. Willis. (2005). Performance of error component models for status-quo effects in choice experiments. In: Scarpa R, Alberini A (eds) Applications of simulation methods in environmental and resource economics. Springer Publisher, Dordrecht, The Netherlands

Scarpa, R., K. G. Willis, and M. Acutt. (2007). Valuing externalities from water supply: status-quo, choice complexity and individual random effects in panel kernel logit analysis of choice experiments. *Journal of Environmental Planning and Management* 50(4):449-466.

Scarpa, R., M. Thiene, and F. Marangon. (2008). Using flexible taste distributions to value collective reputation for environmentally-friendly production methods. *Canadian Journal of Agricultural Economics* 56:145-162

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.

Sinden, J.A. (1988). Empirical tests of hypothetical bias in consumers' surplus surveys. *Australian journal of agricultural economics* 32(2.3):98-112.

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* 30:305-314.

Taylor, L. O., M. D. Morrison, and K. J. Boyle. (2010). Exchange rules and the incentive compatibility of choice experiments. *Environmental and Resource Economics* 47:197–220.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.

Train, K., and M. Weeks. (2005). *Discrete Choice Models in Preference Space and Willing-to-pay Space*. In: Scarpa R, Alberini A (eds) Applications of Simulation Methods in Environmental and Resource Economics. The Economics of Non-Market Goods and Resources, vol. 6. Springer, Dordrecht, The Netherlands, pp1-16

Volinsky, D., W. 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*.

57:205-221.

Yeu, C., and C. Tong. (2009). Organic or local? Investigating consumer preferences for fresh produce using a choice experiment with real economic incentives. *HortScience* 44 (2):366-371.

Appendix A - Random parameter logit models used for inferring non-attendance.

	Split 1 – hypothetical		Split 2 – real incentive	
	Coefficient	Std.err.	Coefficient	Std.err.
Random parameter mean estimates				
Organic produce (OP)	0.412	0.206	0.402	0.121
Produced in Denmark (DK)	0.258	0.208	0.105	0.132
Multicoloured (MC)	-1.307	0.301	-0.737	0.147
Crisp (CR)	-1.241	0.329	-0.880	0.230
Price (PR)	-0.037	0.014	-0.030	0.008
Fixed parameter estimates				
ASC - SQ	-1.485	0.344	-2.328	0.238
Cholesky matrix diagonal values				
Organic produce (OP)	0.550	0.185	0.466	0.123
Produced in Denmark (DK)	0.092	0.228	0.174	0.146
Multicoloured (MC)	0.152	1.148	0.028	0.308
Crisp (CR)	0.034	0.461	0.081	0.300
Price (PR)	0.0004	0.010	0.002	0.024
Cholesky matrix below-diagonal values				
DKxOP	0.309	0.202	-0.093	0.140
MCxOP	-0.588	0.290	0.201	0.204
MCxDK	-0.390	0.506	-0.425	0.167
CRxOP	0.654	0.362	-0.788	0.240
CRxDK	-0.439	0.581	-0.275	0.228
CRxMC	-0.219	1.418	0.105	0.333
PRxOP	-0.023	0.013	0.004	0.010
PRxDK	-0.008	0.013	-0.017	0.008
PRxMC	-0.003	0.020	0.005	0.012

PRxCR	0.000	0.011	0.002	0.009
N	359		828	
LL at convergence	-263.1		-604.6	
Pseudo R^2	0.333		0.335	

Note: For reasons of simplicity as well as for identification purposes, levels for the multilevel attributes have been collapsed into two mutually exclusive levels so that a single dummy variable can describe the variation in a specific attribute. This simplification obviously comes at a loss of design efficiency and reduced comparability to the models in table 5. Nevertheless, we believe this approach is acceptable considering the purpose which is merely to enable estimation of individual-specific conditional parameter estimates in order to calculate the coefficients of variance and thus infer non-attendance for each attribute for each respondent.

Behavioral implications of providing real incentives in stated choice experiments

Highlights:

- A theoretical model incorporating self-image effects at the attribute level
- No hypothetical bias for a low-priced, low-involvement marketed good
- introducing an economic incentive affects behavior beyond hypothetical bias
- Some of these effects are desirable while others are not
- Special caution should be taken if relying on stated non-attendance