University of Reading

Examining the relationship between visual attention and stated preferences: a discrete choice experiment using eye-tracking

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Balcombe, K., Fraser, I., Williams, L. and McSorley, E. ORCID: https://orcid.org/0000-0002-2054-879X (2017) Examining the relationship between visual attention and stated preferences: a discrete choice experiment using eye-tracking. Journal of Economic Behavior & Organization, 144. pp. 238-257. ISSN 0167-2681 doi: 10.1016/j.jebo.2017.09.023 Available at https://centaur.reading.ac.uk/72729/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.jebo.2017.09.023

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

University of Reading

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Examining the Relationship Between Visual Attention and Stated Preferences: A Discrete Choice Experiment Using Eye-Tracking

Kelvin Balcombe
University of Reading

Iain Fraser*
University of Kent
&
La Trobe University

Louis Williams
University of Reading

Eugene McSorley
University of Reading

September 2017

Acknowledgements

The authors wish to thank the editor and two anonymous referees for extremely useful comments and suggestions provided on earlier versions of the manuscript. They also acknowledge the funding from UK Department for Environment Food and Rural Affairs (Defra) for the research project (FA0156) on which this article is based. The views expressed in the article are entirely those of the authors and do not necessarily represent the views of Defra.

Address for correspondence:
School of Economics
University of Kent
Canterbury
Kent, CT2 7NP
UK
i.m.fraser@kent.ac.uk

Examining the Relationship Between Visual Attention and Stated Preferences: A Discrete Choice Experiment Using Eye-Tracking

Abstract

We examine the relationship between visual attention and stated preferences derived from a discrete choice experiment. Focussing on consumer preferences regarding country of origin food labels, we employ a Bayesian infinite mixture Logit to derive results that reveal patterns of respondent heterogeneity that would not be captured assuming that random parameters take a specific distributional form. Our results reveal weak relationships between the eye-tracking data, our stated preference results and various attribute use questions. Although respondents with higher levels of visual attendance value specific attributes more highly, the strength of the relationship is fairly weak. Therefore, whilst we maintain that eye-tracking is useful, we argue that there needs to be greater clarity about the aims and purpose of using eye-tracking in stated preference research.

Key Words: Discrete Choice Experiment, Eye-Tracking, Bayesian infinite-mixtures Logit.
JEL: D83, Q18, C99

1. Introduction

There is a rapidly growing literature examining the relationship between stated preference results, and associated measures of visual attention collected by employing eye-tracking (ET) technology (e.g., Balcombe et al., 2015; Oviedo and Caparros, 2015; Van Loo et al., 2015; Rasch et al., 2015; Meißner et al., 2016; Spinks and Mortimer, 2016; Uggeldahl et al., 2016; Krucien et al., 2017). This literature is attempting to understand the extent to which ET data can be used to improve stated preference research, specifically discrete choice experiments (DCE) (and conjoint analysis), both in terms of understanding respondent behaviour as well as model results. In particular, there is a large literature examining the extent to which survey respondents employ all of the information presented in a DCE (see Balcombe et al., 2016b). This behaviour is referred to as attribute non-attendance (ANA) and existing methods of assessing ANA, such as stated ANA or inferred ANA are considered imprecise. It has been argued that ET data can help to improve our understanding of ANA and its impact on model results (Balcombe et al., 2015) and it has been further argued that visual attention (rather than attendance) should be used to improve preference estimates (Krucien et al., 2017).

In this paper, we add to the literature by statistically analysing ET data collected as part of a DCE examining consumer preferences for country of origin (COO) information for meat products. Specifically, we address a central question: do 'higher value' attributes attract more visual attention (*ceteris paribus*)? If the answer to this question is "yes", then there is an obvious appeal for researchers in collecting ET data as an additional source of information about consumer preferences. However, it is tempting to think that if the answer to this question is yes, that i) attributes paid more visual attention than others implies they are of more value; and ii) if one person pays more attention to an attribute compared to somebody else then

they value that attribute more. Neither of these conclusions follow. This is because visual attention (however measured) is driven by many factors including visual salience, clarity of the visual cues, the complexity of the problem context along with differences in the nature of the individual that are not associated with their valuation of the attribute (non-value related visual respondent heterogeneity) (Orquin and Muller Loose, 2013). Therefore, whether people pay more visual attention to high value attributes is important, but whether that attention can be used to infer higher value across attributes or individuals, is an equally important but different question.

To ensure that we address these questions in a consistent and statistically sound way, we employ a robust econometric specification. To date, within the DCE literature preference heterogeneity has most commonly been modelled using two distinct types of models: the Mixed Logit (ML); and the Latent Class Logit (LCL). Both have a Bayesian analogue, with the former sometimes being called the Hierachical Bayes Logit (HBL) (e.g., Balcombe et al., 2015 and 2016b) and the latter being called a finite mixture Logit (FML). With the ML/HBL preference parameters are modelled as continuous distributions. With the FML/LCL approach, there are two or more groups, with each individual having a probability of belonging to each group and their preferences modelled accordingly. Both approaches are subject to limitations. With the ML/HBL the use of continuous distributions to model preference heterogeneity may be inappropriate if the underlying distribution is a finite number of distinct mixtures. In contrast, when employing the LCL/FML approach statistical criteria (i.e., AIC, BIC, etc.) as well as subjective judgements of researchers are used to identify the 'correct' number of mixtures (N^{o}M) when it is difficult to identify or does not really exist.[1]

To deal with these limitations, we employ the Bayesian 'infinite mixtures' Logit (BIML) to estimate our DCE model specifications. The use of the BIML within economics is currently very limited (e.g., Burda et al., 2008; Daziano, 2013) which is surprising given the benefits it can yield to the researcher.[2] Specifically, the appeal of the BIML is that it provides us with the means to address the identified weakness of both the ML/HBL and the LCL/FML. In particular, it is no longer necessary to select random parameter distributions *ex-ante* for each attribute as the BIML 'non-parametrically' estimates these distributions. This means that respondent heterogeneity will not be mis-specified as might be the case if we simply rely on the use of a pre-specified distribution, such as the normal which is the standard choice in the literature. To implement this approach, the BIML employs a Dirichlet Process (DP) prior, such that individual preference parameters are formed as a mixture of distributions that exist within a given iteration of the BIML algorithm, meaning that individual preference parameters have a mixture. The estimates for a given individual are constructed as a weighted sum of the distributions that compose the mixture, where the weights are the probabilities that a person belongs to a particular 'class'. The DP prior shrinks the N^{o}M to a number less than the number of individuals in a sample of data. This means that the BIML is able to endogenously identify

[1] In this paper, we will refer to the N^{o}M although this is frequently referred to in the literature as the number of classes.

[2] We note that our implementation of the BIML extends aspects of earlier research. Specific details are discussed in Section 4.

the N^{o}M from the data without imposing any limit on the N^{o}M.[3]

The issue that we examine is consumer preferences towards COO information on food labels. The use of mandatory labels on various food types in the EU has been growing.[4] To better understand consumer attitudes towards COO, we undertook a hypothetical DCE simultaneously employing ET methods so as to assess how survey respondents engaged with the DCE, building on the research of Balcombe et al. (2015).

Our study contributes to the literature in a number of related ways. First, we highlight that the BIML reveals aspects of respondent heterogeneity that can be missed as a result of employing standard econometric specifications. This is particularly important in this study given the highly individual and specific nature of ET data. For example, ET data measuring an individual's visual attention is unique (Orquin and Muller Loose, 2013). As such we need to employ a flexible and robust modelling approach of individual choice so that we can effectively investigate the relationship between ET and respondent attitudes to COO. The flexibility that the use of the BIML brings to our analysis helps to ensure that aspects of behaviour revealed by the ET are not missed as a result of making inappropriate random parameter distribution choices. Importantly, our results reveal parameter distributions that are unlikely to be captured by the typical pre-specified choice of continuous distributions used within a ML/HBL. This finding supports those reported by Burda et al. (2008) and Train (2016) and it suggests that future examination of DCE data needs to explore more flexible econometric strategies so as to avoid the possibility of model mis-specification.

Second, we examine how ET data can be used to test various hypotheses that exist in relation to DCE. In particular, we find a significant but relatively weak relationship between our ET results and the stated ANA data provided by respondents which is in keeping with the findings reported by Balcombe et al. (2015). We also find that over the sequence of choices made that average attention (i.e., dwell time per attribute) declines which is in keeping with a result reported by Meißner et al. (2016). Finally, we find that there is a statistically positive but weak relationships between our willingness-to-pay (WTP) estimates for the DCE attributes and how long a respondent dwells on a specific attribute. This result is similar to those reported by Meißner et al. (2016) who consider attribute importance and product attractiveness. Therefore, there is some (albeit weak) evidence to support the collection of dwell time using ET as a means to better understand the choices made by survey respondents.

Although these contributions are incremental, collectively they have implications for how ET is to be used in he future. First, we believe that progress in the use and application of ET will be through accumulation of evidence that points to similar (or different) findings, regardless of the particular nature of the DCE. Second, there has to some extent been a confounding of goals within the current ET literature and we hope our findings prompt other researchers into considering these issues at this juncture. Third, our results raise various questions regarding

[3] For more details on the DP see McAuliffe et al. (2006).

[4] In April 2015 the scope of COO was extended with the inclusion of fresh, chilled and frozen meat of swine, sheep, goats and poultry. Specifically, following Impact Assessments, Commission Implementing Regulation (EU) No 1337/2013 of 13 December 2013 laying down rules for the application of Regulation (EU) No 1169/2011 of the European Parliament and of the Council as regards the indication of the country of origin or place of provenance for fresh, chilled and frozen meat of swine, sheep, goats and poultry sets out the requirements for COO for these species and COO labels became mandatory for these products from 1 April 2015.

the potential longer term benefits from employing ET within a DCE. Currently, ET is a resource intensive exercise and there are technological barriers preventing it from being used pervasively within DCEs. If ET only provides marginally better information about respondent preferences, the costs of its large scale use may outweigh the benefits. Thus, we would discourage researchers from simply advancing research that solely examines the "best" way to use ET data in estimation, when in fact this is largely a non-issue for those implementing DCEs in the field, though of course this may change with technological advancement. Finally, ET offers an avenue to improve our understanding of respondent engagement both generally and in relation to specific DCE studies. Accordingly, we ask if one of the most effective uses of ET is as an input into the design of a DCE such that higher levels of attribute use are achieved and as such the extent of ANA reduced *ex-ante*? We discuss this question in the final section of the paper in light of our general findings and the use of ET within DCE.

The structure of our paper is as follows. In section 2, we begin with brief overview of ET and a review of applications within DCE studies that yield several outstanding research questions. Then in Section 3 we describe our DCE and this is followed by a description of the BIML model. In section 5, we present our model results that are composed of BIML results, the ET results and finally how the ET data and BIML results correlate. Finally, in Section 6, we discuss the implications our findings and conclude.

2. Eye-Tracking, Economics and Stated Preference Methods

ET is not a new technology and it has been extensively used in marketing and psychology for many years (Orquin and Muller, 2013; Lahey and Oxley, 2016). Its use has been more common in psychological research that is attempting to explain the underlying processes that determine choice, which is in contrast with economics that often puts more emphasis on explaining outcomes rather than the processes behind these outcomes. However, the increasing use of ET in economics reflects a growing interest in the profession to examine process data that sheds light on cognitive constraints as a way to improve our understanding of how choices are made (Woodford, 2014, Caplin and Dean, 2015, Caplin, 2016 and Geng, 2016). This idea is neatly expressed by Caplin (2016) as follows:

> *Yet standard economic data does not reveal what was noticed or considered, only what was chosen. This issue is fundamental and essentially universal. Observing final choices alone is inadequate once one makes allowance for incomplete information and its attentional grounding....we will have no choice but to face up to this limitation explicitly, and to work on data enrichments that liberate separate understanding of learning and choice.* (p2).

Given the existing use of ET a number of important findings have emerged regarding what ET data can and cannot reveal about individual decision making. For example, Orquin and Muller Loose (2013) in a very comprehensive review of ET, explain why ET data is considered so useful in helping to explain choice:

> *One of the theories which have had significant influence on eye tracking based*

decision research is the eye-mind assumption. The eye-mind assumption suggests a strong causal relationship between working memory and attention. (page 192).

Thus, to rephrase this point: choices and attention are in some way interrelated as has been demonstrated in the literature. More generally Orquin and Muller Loose (2013) summarise the literature on attention and eye movements as follows:

1) eye-movements during decision making are controlled both by top down and bottom up processes, 2) learning significantly influences the speed and accuracy of fixations, 3) decision makers trade-off between fixations and working memory, and 4) fixated information influences decision making more than non-fixated information. (page 193).

In another survey of the ET literature with a specific focus on food choice, Krajbich and Smith (2015) explain that decisions will typically take longer when a survey respondent is indifferent. They review the literature from a psychological perspective and make specific reference to the drift decision model (DDM) and more recent advances.[5] In particular, they observe that more dwell (or gaze) time on a specific item will typically yield a higher probability of the item being selected and that attention determines preference based choice in binary choice settings at least. Interestingly, in food choice research they note that dwell duration does not correlate with how the food is rated. Krajbich and Smith (2015) conclude by observing that more research is required to understand how information accumulation interacts with attention, and there may be path dependency in the sense that the pattern (or path) of attention influences the processing of the information that has been obtained.[6]

2.1. Eye-Tracking and Stated Preference Research

Increasing numbers of stated preference studies are making use of ET data, including DCEs (e.g., Balcombe et al., 2015; Van Loo et al., 2015; Spinks and Mortimer, 2016; Uggeldahl et al., 2016), conjoint analysis (e.g., Meißner et al., 2016; Rihn et al., 2016; Krucien et al., 2017), contingent valuation (e.g., Oviedo and Caparros, 2015) and experimental auctions (e.g., Lewis et al., 2016; Rihn and Yue, 2016). In the case of DCEs, somestudies motivate the use of ET data because of issues surrounding ANA.[7] Within the DCE literature the economic models that have examined ANA using ET have been referred to as 'relaxed rational models' (Orquim and Muller Loose, 2013). They suggest that these models although making no theoretical prediction about attention have included fixation (and or dwell time) frequencies as well as visual ANA within model specifications so as to improve model fit.

An example of the use of ET data to examine ANA is provided by Balcombe et al. (2015). As they note, ideally all DCE attributes should receive high levels of visual attendance/attention,

[5] The potential uses of the DDM in economics is further considered by Caplin (2016) and Caplin and Martin (2016).

[6] By attempting to understand information accumulation model performance is frequently evaluated on the basis of predicting outcomes as opposed to generating measures of the values associated with the final choices made. The different emphasis on model objectives is in part why DDM have as yet not been used within DCE, although several researchers allude to information accumulation in their writing.

[7] We do not review the more general use of ET within economics. For a discussion of this literature see Balcombe et al. (2015) and Lahey and Oxley (2016).

especially the price/cost attribute because high levels of attention are more likely to yield reliable WTP estimates and if the price/cost attribute is really being ignored then the WTP estimates can be grossly inflated. As such ET provides another means by which to assess the degree to which respondents engage with the survey instrument as well as assessing if answers to debriefing questions that are frequently employed within DCEs are meaningful.

To ensure that ET data can be understood and interpreted within the context of DCE analysis, Balcombe et al. (2015) distinguish between visual attention and attendance, although the two concepts are related:

- **Visual attendance** requires a respondent to "fix" on an attribute across all choice options on a given choice set (card). The requirement that all options are fixed upon (where neither are blank) is needed for a proper comparison of attribute levels to have been made.
- **Visual attention (dwell time)** is measured by the total number of fixations (i.e., times looked at) on a given attribute or total "dwell time" on a particular attribute (i.e., how long looked at).

The definitions of visual attendance and attention (dwell time) employed by Balcombe et al. (2015) allowed them to condition the choice data prior to estimation to capture the extent to which attributes enter the consideration set of respondents. However, this is far from the only way ET data can be used within stated preference research. For example, visual attention and attendance can be compared and correlated with the stated ANA responses of respondents (e.g., Spinks and Mortimer, 2016). This is a relatively simple piece of analysis to undertake but it can reveal important differences between how a respondent thinks they have used information compared to that which they have seen.

Alternatively, the ET data can and has been explicitly included within model specifications in an effort to improve model performance. For example, Van Loo et al. (2015) used a measure of visual attention by attribute and included it within the utility specification. Similarly, Grebitus et al. (2015) use dwell time within a utility function to help explain choice in a DCE examining COO in addition to various other food related attributes. In contrast, Uggeldahl et al. (2016) use ET data on the number of times visual attention moved (gaze shift) between alternatives to examine respondent certainty of choices in a spirit similar to that found in the contingent valuation literature. Interestingly, they report that response time required to make a choice significantly improved model performance compared to gaze shift. We also note that Oviedo and Caparros (2015) use ET within a DCE but in this case they use the ET data along with a comparable contingent valuation study so that they can assess convergent validity. Most recently Krucien et al. (2017) examine visual attention and ANA by employing a latent processing model specification that is similar to Hess and Rose (2012). In this paper an alternative definition of visual attention to that employed by Balcombe et al. (2015) is used, highlighting that there are as yet no agreed approaches to how researchers define particular constructs from ET data. Although Krucien et al. (2017) implement an interesting and promising econometric approach to the use of ET data the results presented do not appear to statistically support the latent processing model they introduce. This in many ways reflects

the current dilemma researcher faces when collecting ET. ET data is time consuming to collect, and how it is used is still the subject of much debate. There remains a non-negligible possibility that its inclusion within an econometric specification will not improve estimates of WTPs or Marginal Utilities.

Moving beyond DCEs, ET data has been collected during the implementation of experimental auctions by Lewis et al. (2016) and Rihn and Yue (2016). In both studies efforts have been made to assess if visual attention is positively or negatively related to WTP bids. The evidence presented by both studies is mixed regarding the extent to which the degree of attention paid to specific attributes correlates with value. These results differ significantly to those reported by Meißner et al. (2016) who report that within a series of three conjoint studies that attention is positively correlated with more important attributes (See Figure 5 on page 8 for details). This difference in findings may in part be a result of the difference in experimental method in that the experimental auction of Lewis et al. (2016) showed respondents a series of changing products whereas the conjoint showed the same task 12 times such that respondents are able to learn and as a result more accurately express their preferences and this may well be why the positive correlation between attributes and utility (value) is more likely to be detected.

2.2. Issues Emerging from the Literature

Based on the existing applications of ET in the literature, several important issues emerge. First, there are several forms that ET data can take (what is measured) and by and large areas of interest (AOI) have proven the most common way in which to assess user engagement. Second, given an AOI, various measures have emerged that have become popular within the literature: visual attendance; visual attention/dwell time; and the number of changes in viewing alternatives. Clearly, the reason for the use of these various measures stems from the fact there is as yet no specific or preferred use of ET data within a DCE. Third, currently the reasons that are being used to justify the use of ET methods draw on different aspects of economics and psychology that we believe are to some extent confounded.

The reason why there is a lack of clarity in application stems from the fact that ET is being used to understand a variety of issues including:

1. respondent behaviour and engagement;
2. respondent understanding;
3. to improve econometric performance; and
4. to improve the design of survey instruments.

Understanding that the application of ET within stated preference studies is motivated by this variety of issues is important if we are to make advances in the use of ET. Although we appreciate that the use of ET data to improve econometric performance is understandable, the practical benefits may be marginal. Current studies published contain a number of justifications for using ET data, and there is room for improved clarity and refinement both in terms of the goals of this research and the links with theory that underpins it. This in large part explains

why Orquim and Muller Loose (2013) see the need for 'relaxed rational models' to consider in more detail (i.e., develop a stronger theoretical link) the processes behind ANA.

In summary, given the current developments and acknowledged limitations in the literature, we are of the opinion that there are (at least) three questions that emerge for the use of ET data within stated preference research and DCE applications specifically. In each case the word 'attention' can be substituted with 'attendance' to obtain a slightly different question.

Question 1: Do 'higher value' (more attractive) attributes attract more visual attention (*ceteris paribus*)?

Question 2: If individuals or groups have higher visual attention (relative to other attributes) can we infer that those individuals or groups value that attribute more highly than other attributes?

Question 3: If one individual has higher visual attention towards a particular attribute (relative to other individuals) can we infer that this individual values that attribute more highly than other individuals?

As argued in the Introduction, the answers to questions 2) and 3) do not follow from the answer to 1). In addition, the literature is at present unclear about the specific relationship between the value placed on an attribute and the extent to which this is captured by visual attention. Therefore, to address these questions, we consider how measures of visual attention (dwell time) relate to the attributes in our DCE. Our view is that further consideration needs to be given as to relationships between ET data, stated measures (such as reported attendance and importance), as well as estimates such as WTP.

3. Eye-Tracking and DCE Implementation

3.1. Attribute Selection and COO Treatments

There is continued and growing interest and use of COO on food labels and packaging. In this study, we wished to examine UK consumer preferences regarding COO for meat.[8] Specifically, we consider how COO information for a 12 inch pepperoni pizza affects choices made. In this case the COO information only relates to the meat ingredient in the product.

Our choice of attributes to employ within the DCE was determined by reference to the antecedent literature and the views of policy makers. The set of attributes used in the DCE are as follows:

- **Price** - For this attribute, the range of values was determined by reference to product size and description and by reference to those most commonly on sale in UK shopping outlets. The set of prices used in the DCE are £2.00, £2.95, £3.75 and £5.25.
- **COO** - This attribute had four possible options: UK, USA, Italy and EU. The choice of countries (USA and Italy) reflects potential sources for imports of pepperoni and the UK and EU capture home country and a generic source indication that is used on existing food products.
- **Product Quality** - We selected three levels for this attribute: Basic, Choice and Premium. The inclusion of an attribute to describe product quality meant that we could implicitly

[8] The research reported here is part of a wider study on COO. See Balcombe et al. (2016a) for more details.

capture aspects of the product that relate to taste or other quality related characteristics. This attribute acts a cue indicating all of those characteristics that constitute "quality", but are not stated explicitly.

- **Farming System (Production)** - This attribute was either Organic or Conventional. These two production systems capture the majority of meat production on the market and is familiar to consumers.
- **Quality Assurance** – For this attribute our levels are: No Label, Freedom Food and the International Quality mark.

Currently, legislation permits, but does not mandate, the use of flags to signify COO and we considered it important to see if a flag effects our WTP estimates for COO. To assess this issue, we employed two experimental treatments to investigate the impact of text and text and flag in relation to COO:

Treatment 1 (T1) - presented the choice cards with only text to describe the COO.

Treatment 2 (T2) - employed both text and the appropriate flag.

Our respondents were randomly assigned to either treatment with one group completing the DCE with a text only COO version and the second group a version with text and flag.[9] An example of a choice card used in the DCE is presented in Figure 1.

{Approximate Position of Figure 1}

Given the choice of attributes and associated levels, we designed our choice sets in a standard manner. We employed an efficient design assuming a Multinomial Logit utility specification and employing D-error as the measure of design selection. Our design was produced using Ngene version 1.1.1 (Choice Metrics 2012) assuming null priors on our model coefficients. In total, we generated 24 choice cards for the DCE. Each of the participants completed the 24 choice cards which where randomised by task order.

In addition to the choice cards, the survey instrument provided respondents with briefing materials prior to taking part in the experiment and after all the choice cards had been completed a series of debriefing questions concerning their ranking of attributes and which they ignored when making their choices. The ranking question asked respondents to rank each attribute in terms of importance of them in making their choices. The ANA question asked respondents to indicate which if any attribute they ignored when completing the choice tasks. This data was collected using a standard question format: *"Which of the following attributes did you ignore when completing the choice task? (You can tick none or as many as required)".*

In designing and implementing this specific DCE, we did not include a no-choice option. As is well understood in the literature, a no-choice option is important when calculating welfare measures, but within this specific DCE, we did not see this as critical. We took this view because the sample of respondents we employed are not all a representative sample of consumers. This is because this specific DCE was implemented with a specific focus on user engagement with the survey instrument as assessed and gauged by use of ET data so as to feed into the design of a much larger DCE.

[9] Uggeldahl et al. (2016) report that employing pictures and not text to describe attributes reduces the degree of error variance which they attribute to a reduction in complexity in the DCE implementation.

3.2. DCE Sample

100 participants completed the ET DCE although one individual was dropped from the data analysis because of incomplete data collection. Our sample was recruited using University of Reading mailing lists, with respondents being incentivized by £10 for participation.[10]

In terms of the gender mix our sample was composed of 47 males and 53 females. The sample was also composed of a wide range of ages, but with a larger proportion of young people than in the population as a whole as well as very few participants over 55 years of age. Importantly, all participants consumed meat. Almost all (96) indicated that they were either the main shopper in the household (60) or shopped for meat some of the time (36). The majority of these participants indicated that they bought fresh meat more commonly than frozen, usually at least once a week and that they shopped in the expected range of supermarkets.

50 participants completed the treatment without a flag and 50 completed the treatment with text and a flag. All participants were eye-tracked while making their choices with choices being recorded by the selection of one of two buttons on a console (option A or B).

4. The Bayesian Infinite Mixtures Logit

As is standard in the DCE literature, we assume that the utility (U_{ijs}) for the jth person from the ith option in the sth choice set is:

$$\begin{aligned} U_{ijs} &= u\left(x_{ijs}, \beta_j\right) + e_{ijs} \\ &\text{for} \\ i &= 1, ..., N \\ j &= 1, ..., J \\ s &= 1, ..., S \end{aligned} \tag{1}$$

where x_{ijs} is a $(K \times 1)$ vector of known attributes, β_j is a vector of parameters characterising the systematic preferences of an individual and $u\left(x_{ijs}, \beta_j\right) = x'_{ijs}\beta_j$. As is standard, we also assume an extreme value error e_{ijs} that is independent across, i, j and s implying that the probability of choosing option i for the jth person in the sth choice set is (where β_j is known):

$$p_{ijs}\left(\beta_j\right) = \frac{e^{u\left(x_{ijs}, \beta_j\right)}}{\sum_i e^{u\left(x_{ijs}, \beta_j\right)}} \tag{2}$$

Two broad classes of models that allow for preference heterogeneity are the ML/HBL model under which a prior distribution is assigned to the latent parameters β_j, and the LCL/FML approach whereby β_j is assumed to be drawn from a discrete set of possible values. Under the ML/HBL approach it is most commonly assumed that $\beta_j \sim N\left(\mu, \Omega\right)$. Thus, the closer the true distribution of preference parameters is to the normal distribution, the better this model

[10] Our sample size is comparable with other studies reported in the literature. For example, Van Loo et al. (2015) employed 81 respondents, Oviedo and Caparros (2015), employed 26, Grebitus et al. (2015) had 130 respondents, Uggeldahl et al. (2016) 190 respondents and Krucien et al. (2017) 58.

will work. Under the LCL/FML approach, there are a countable set $\{\beta_m\}_{m=1}^{M}$ of vectors such that for every person j there is a mixture (i.e., classes) $c_j = m$ indicating that $\beta_j = \beta_{c_j}$.

To develop the BIML, we first present the Bayesian FML (with M mixtures) for the logit model in equation (3).

$$\begin{aligned} y_{ijs}|c_j, \{\beta_m\} &\sim p_{ijs}\left(\beta_{c_j}\right) \\ c_j|\{\pi_m\}, C &\sim Discrete\left(\pi_1, ..., \pi_M\right) \\ \beta_m &\sim G_0 \\ (\pi_1, ..., \pi_M) &\sim Dirichlet\left(\frac{\alpha}{M}, ..., \frac{\alpha}{M}\right) \end{aligned} \tag{3}$$

In equation (3) π_m are the probabilities that an individual falls into a specific mixture m, G_0 is a prior distribution for the parameters (possibly normal) and α is a hyperparameter, governing the 'concentration' of mixtures. The limitation of the FML specification is that it requires the pre-selection of M, the N^{o}M. However, a way to do this endogenously within the estimation process, is by using a DP.

The DP model is a generalisation of the FML model such that the number of potential mixtures can (in principle) be infinite although there will always be a finite N^{o}M generated within a Monte Carlo Markov Chain (MCMC) algorithm. Most importantly, by employing a DP, the prior probability that an individual falls within a given mixture becomes conditional on the size of the mixture (i.e., number of members).

Formally, the DP model with an extreme value error (i.e., logit) for individual $j's$ responses $\{y_{ijs}\}_{i,s}$ ($y_{ijs} = 1$ if individual j selects the ith option in the sth choice set, otherwise 0), denoted Y_j is:

$$\begin{aligned} Y_j|\beta_{c_j} &\sim f\left(Y_j|\beta_{c_j}\right) = \prod_{s=1}^{S}\prod_{i=1}^{N}\left(p_{ijs}\left(\beta_{c_j}\right)\right)^{y_{ijs}} \\ \beta_{c_j}|G &\sim G \\ G &\sim DP\left(G_0, \alpha\right) \\ \alpha &\sim Gamma(s, v) \end{aligned} \tag{4}$$

where G is a discrete distribution and $DP\left(G_0, \alpha\right)$ represents a DP with base distribution G_0 (which can be continuous). Unlike in equation (3) α, the concentration parameter, is determined within the estimation procedure such that an increase in α increases the average N^{o}M, and it is estimated using the hierachical priors shown in equation (4), as outlined in Escobar and West (1995).[11]

Before we proceed, it is briefly worth explaining several attractive and important features of this model. First, if we assume a given distribution (e.g., normal) when employing a HBL then it will tend to deliver parameter estimates that are consistent with that distribution. And if one generates data with preferences from two groups in a manner that is inconsistent with normally distributed parameters (e.g., highly bimodal), then the HBL tends to deliver latent distributions that (while they depart from normality or some other distribution) will tend to

[11] The DP model presented by Burda et al. (2008) only noted the possiblity of using Escobar and West (1995).

be skewed towards a "normalish" looking distribution. They may reflect signs of bimodality but the assumed nature of the distribution plays a large role. Burda et al. (2008) discusses this point at length. Conversely, the BIML does not do this to the same extent. Indeed, if we generate data from a normal distribution, the BIML tends to give normal looking posteriors, though the posteriors will likely be less accurate that the HBL that assumes normality. But clearly, if normality is not an appropriate choice then the BIML does a much better job at reproducing the underlying distribution. This issue has also been examined by Train (2008, 2016) within a classical setting using a logit function for the mixing distribution as well as the choice probabilities. As is evident from the discussion in Train (2016) reducing the need to make parametric assumptions in model estimation is very desirable but there are a host of questions that emerge from the methodological approach presented.

Second, although the BIML will estimate a N^{o}M there is not necessarily an "optimal" N^{o}M. The N^{o}M can be considered as a random variable in much the same way as parameters are. What this means is that throughout the estimation process the N^{o}M changes. Thus, although *ex-post* we can identify a mean or a median N^{o}M, the BIML does not, in contrast to the FML, assume that there is an optimal N^{o}M. For example, if the underlying parameter distribution is multivariate normal, there is in one sense no 'correct' N^{o}M. Moreover, during the BIML estimation procedure a given mixture can switch labels in the sense that at one point the sampler of 'mixture 1' which is labelled 'one', will later on become 'mixture 2' while retaining its label 'one' etc. Thus there is immediate meaningful identification of mixtures, as they can appear and disappear throughout the sampling process, although the mixtures with the most membership tend to stick around (Burda et al., 2008). Importantly, this presents no problem unless ones' aim is to assign individuals to classes (as in the classical latent class approach), which is not something we aim to do.

This approach described above is different to the classical way of choosing the 'correct' N^{o}M which is generally to try and maximise a penalised likelihood (information criteria).[12] However, choosing the correct N^{o}M is not the aim of DP models. Indeed, they are not predicated on the idea that there is a correct N^{o}M. As we have explained, the N^{o}M is a parameter that has a posterior distribution just like all parameters and the DP imposes a prior structure on the N^{o}M. Naturally, some values for the N^{o}M will have higher posterior probability than others. Having sampled the N^{o}M posterior, researchers typically report a point estimate in terms of the mean, median or mode The classical instinct would perhaps be to go further and insist that estimation should then be restricted to the modal N^{o}M. However, the Bayesian response would be that this is simply restricting inference about parameters to a subspace of the posterior without any particular *a priori* reason and as such could introduce a specification bias that has a resulting impact on model performance and results.

Finally, as is becoming more common within the DCE literature, we estimate our model as a utility in WTP space specification (e.g., Train and Weeks, 2005). This approach to model specification is a simple transformation using the price coefficient such that our model will yield WTP estimates for all attributes. Starting with the standard utility specification,

[12] For an extensive discussion of the Classical LCM see Train (2008).

$$U_{ijs} = -\exp(\beta_{1j})x_{1js} + \gamma_{2j}x_{2js} + ... + \gamma_{kj}x_{kjs} + e_{ijs} \tag{5}$$

where x_{1js}, the price attribute, has a coefficient that is exponentially transformed so as to ensure that the coefficient is negative. Following Train and Weeks (2005), we transform equation (5) as:

$$\begin{aligned} U_{ijs} &= \exp(\beta_{1j})\left(-x_{1js} + \frac{\gamma_{2j}}{\exp(\beta_{1j})}x_{2js} + ... + \frac{\gamma_{kj}}{\exp(\beta_{1j})}x_{kjs}\right) + e_{ijs} \\ &= \exp(\beta_{1j})\left(-x_{1js} + \beta_{2j}x_{2js} + ... + \beta_{kj}x_{kjs}\right) + e_{ijs} \end{aligned} \tag{6}$$

It now follows that the quantities $\beta_{2_j}, \beta_{3_j}, ..., \beta_{k_j}$ are the marginal rates of substitution (MRS), and with the numeraire being the first attribute, which in this case is price, means they are also WTP estimates. In addition, to the reasons typically stated in the literature for employing this approach, we view the ability to be able to employ priors that are meaningful in terms of the WTP estimates as an important issue. Thus, the difference between this specification and a preference space model, is that in preference space priors need to be set taking account of scale and this will always vary between model specifications and data sets.

4.1. BIML Priors

To establish the priors used in model estimation, we first employed an empirical Bayesian approach to select G_0. Given that the model was estimated in WTP space, as explained above, the priors could be set so that they could reflect that actual price variation observed in the products we examine in the DCE i.e., Pepperoni Pizza. Thus, we initially estimated a standard logit specification assuming non-informative priors and this lead us to specify $G_0 = N(0, g_0 \times I)$ where we subsequently considered three values for the standard deviation: $\sqrt{g_0} = (1, 2 \text{ and } 3)$. For the product we examined with our DCE the total variation in our prices were £3.25 (£2 to £5.50). As we would expect, these priors did have an impact on the final results with the more diffuse priors yielding larger WTP estimates, but with the relativities being preserved. Given that we consider our most informative prior (unity) sufficiently diffuse so as not to dominate the data it was used for the subsequent analysis.

Second, for estimation we experimented with two sets of hyperparameters for the mean and variance for the gamma distribution for α. For two of the runs we set the prior mean equal to five, with two sets of standard deviations of $\sqrt{5}$ and $\sqrt{20}$. The results were not that sensitive to the setting of these priors.

4.2. BIML Estimation

The estimation of the BIML requires that we combine the methods employed for the HBL and the DP. Neal (2000) outlines a number of algorithms that achieve this. In this paper, we employ algorithm five introduced by Neal (2000). This algorithm is a standard Metropolis

Hastings (MH) with additional DP steps allocating individuals to mixtures such that new mixtures can be introduced or eliminated until the model converges. Thus, for a given point in the Markov chain (which can be initially specified), we define $\{m\}_{m=1}^{M}$ as the set of non-empty mixtures (mixtures with at least one individual allocated to each class) along with $\{\beta_m\}_{m=1}^{M}$ and $\{c_j\}_{j=1}^{J}$. Let n_m be the number of individuals allocated to mixture m and $n_{-j,m}$ the number of individual allocated to mixture m having excluded the jth individual. Following Neal (2000), we then define the following proposal distribution for new mixtures for each individual

$$\text{Prob}\left(c_j^* = m^* | m^* \in \{m\}_{m=1}^{M}\right) = \frac{n_{-j,m^*}}{J-1+\alpha} \tag{7}$$

$$\text{Prob}\left(c_j^* = m^* | m^* \notin \{m\}_{m=1}^{M}\right) = \frac{\alpha}{J-1+\alpha} \tag{8}$$

The top probability is that which is proposed (as distinct from the probability that this proposal is accepted) that the jth individual is allocated to an existing mixture. The lower probability is that the individual is proposed a mixture that does not currently exist. The algorithm used for updating mixtures is:

- – for $j = 1, ..., J$;
  - – draw candidate $c_j^* = m^*$ from the proposal distribution above;
  - – if c_j^* is not from an existing mixture then draw β^* from G_0 otherwise $\beta^* = \beta_{m^*}$;
  - – accept c_j=c_j^* with probability (with $f\left(Y_j|\beta_{c_j^*}\right)$defined in [4])

$$u = \max\left(1, \frac{f\left(Y_j|\beta_{c_j^*}\right) G_0\left(\beta_{c_j^*}\right)}{f\left(Y_j|\beta_{c_j}\right) G_0\left(\beta_{c_j}\right)}\right); \text{and} \tag{9}$$

  - – redefine $\{m\}_{m=1}^{M}$, $\{\beta_m\}_{m=1}^{M}, \{c_j\}_{j=1}^{J}$ if necessary (that is add new non-empty mixtures and eliminate empty ones)

This algorithm (or repetition) is then nested within a MH algorithm for which values of $\{\beta_m\}_{m=1}^{M}$ are updated conditional on the existing N$^{\text{o}}$M. The algorithm is initiated at an arbitrary starting point and the algorithm is repeated for $t = 1, 2, ..., T + T_0$ times with the first T_0 points being discarded (the 'burn in').

At each iteration t (that is after completing the assignment of all individuals to mixtures and the draw of the parameters associated with those parameters) there is also a draw of α from its posterior distribution which is a function of the hyperparameters in the prior for α, and N$^{\text{o}}$M and number of units (J) only. This step is outlined in Escobar and West (1995). Importantly, at each iteration t, we can estimate each individual's parameters. We do this by constructing the posterior probabilities for mixture membership under a uniform prior probability that they belong to any one of the mixtures. Denoting $\omega_{j,m}^{(t)}$ as the posterior probability that the jth individual is in mixture m at iteration t, we take the estimate of the individual's parameters to be $\beta_j^{(t)} = \sum_{m=1}^{M(t)} \omega_{j,m}^{(t)} \beta_m^{(t)}$, and record these for all individuals at $t = T_0 + 1, ..., T + T_0$. Convergence in the Bayesian sense can be determined by visual and statistical inspection of the

sequence of draws for the parameters which should be a stationary distribution. For the case in hand, we present these statistics in Appendix A.3.

5. Survey Results

5.1. BIML Model Results

We first begin with an examination of BIML results for the DCE data. Given that we have collected data for two COO treatments we needed to decide how best to use our data. Based on model results for both treatments, we found that all model estimates are very similar. For this reason we focus on the pooled data model specification.[13]

The model specification we have estimated takes the following form:

$$\begin{aligned} U_{ijs} &= \exp(\beta_{1j})(-\Pr ice_{ijs} + \beta_{2j} Choice_{ijs} + \beta_{3j} \Pr emium_{ijs} + \beta_{4j} Organic_{ijs} + \beta_{5j} UK_{ijs} \\ &\quad + \beta_{6j} EU_{ijs} + \beta_{7j} USA_{ijs} + \beta_{8j} FreedomFood_{ijs} + \beta_{9j} International_{ijs}) + e_{ijs} \end{aligned} \quad (10)$$

There are several comments that we need to make about this specification. First, the model specification does not include an alternative specific constant (ASC) as we did not employ a no choice option. Second, at each point during the estimation procedure, a respondent is allocated to a given mixture, but they have a probability of belong to each of the mixtures. Their parameters, and those which we report below, are then calculated as a weighted sum of these weights at each iteration. Their overall estimates are then the means of these estimates. The estimates below were constructed from 10,000 draws from the sampler, where each draw was taken after 100 iterations (known as 'thinning') so as to decrease serial correlation. Prior to taking draws there the sampler had 1 million iterations so as to reach a high density region. Convergence of the sampler was determined by formal tests and visual observation of the sequence of draws. Using modified t-tests there were no significant differences (at the 5% level) between the mean of the estimates between of draws in the first and second half of the sequence (see Appendix A3).

We begin by presenting our WTP results for our DCE estimated using our BIML specification in WTP space. As outlined above, the BIML model introduces and removes new mixtures (groups) as part of the MCMC algorithm. As noted we have 99 participants in the sample and we have data for 24 choices. The model results are presented in Table 1:

[13] The frequency of responses with regard to A and B were virtually identical across the two experiments and a simple t-tests for differences between the treatments (Flag and No Flag) for the COO attribute revealed a small but significant impact of the flag effect with a positive effect in relation to the UK versus Italy and a negative effect with regard to the USA versus Italy, but not significant with regard to the EU versus Italy. All model and test results are available on request.

Table 1: WTP Estimates for COO DCE

	Mean	Median	St Dev	Min	Max
Price (Scale)	0.820	0.787	0.384	0.210	1.745
Quality					
Choice (v Basic)	0.208	0.212	0.108	-0.001	0.431
Premium (v Basic)	0.810	0.763	0.418	-0.021	1.621
Farming System					
Organic (v Conventional)	0.910	0.940	0.718	-0.553	3.340
COO					
UK (v Italy)	0.413	0.363	0.253	-0.099	1.036
EU (v Italy)	-0.127	-0.081	0.174	-0.801	0.181
USA (v Italy)	-0.837	-0.910	0.227	-1.154	-0.181
Quality Assurance					
Freedom Food (v None)	0.653	0.654	0.193	0.149	0.970
International (v None)	0.536	0.542	0.270	-0.046	1.124

Notes: Results are pooled data (Flag and No Flag)

The results presented in Table 1 have been generated by a BIML that produced on average 10 (in terms of the posterior mode) mixtures, which is also the mean and mode for the N^{o}M distributions. The distribution of the N^{o}M is shown in Figure 2. As can be seen that a non-zero probability as been assigned to as few as 2 mixtures and as many as 23.

{Approximate Position of Figure 2}

The fact that we have generated our best model results with these N^{o}M is important. While 10 mixtures is not a large number, it is significantly bigger than the typical model results reported for FML/LCL research. Thus, had we relied on a FML/LCL specification it is highly unlikely that we would have arrived at a model specification indicating 10 mixtures.[14]

In terms of interpreting the results presented in Table 1, we can see that the least preferred combination of attributes is: Basic plus + Conventional plus USA plus none. In contrast the most preferred combination of attributes is: Premium plus Organic plus UK plus Freedom Food. Thus, the difference in payment between these two products is calculated as: 0.81 + 0.91 + (0.41 + 0.84) + 0.65 = £3.62. Given that our hypothetical price range is £3.25 this estimate is plausible in magnitude.

An alternative way in which to understand how the BIML has influenced the results reported in Table 1 is to consider the distribution of WTP for all of the attributes. These results are shown in Figure 3:

{Approximate Position of Figure 3}

[14] Train (2008) discusses the potential number of mixtures/classes that can be derived and demonstrates that by using the EM algorithm the limitations frequently encountered in the literature can be overcome.

What we can observe from Figure 3 is that the distributions of WTP by attribute vary significantly in terms of shape. This is important as it indicates that the use of a HBL/ML approach that would attempt to select a random parameter distribution *ex-ante* for each attribute will struggle to capture the differences observed in Figure 3. These results therefore support the use of the BIML to reveal this aspect of respondent heterogeneity that might be mis-specified if we simply rely on the use of a normal distribution which is the standard choice in the literature.

5.2. ET Results

We now turn to our ET results and the specific focus of this DCE. The first thing to report is that there is a very high positive correlation (approx 0.99) between dwell time and total number of fixations on attributes. As a result we only discuss our results in relation to visual attention in terms of dwell.[15]

The first set of results we consider are the extent of dwell by attribute for all 24 choice tasks in the sequence with which they were presented. This information is reported in Figure 4, where total dwell time is reported for each attribute with two sequences representing the Flag and No Flag experiments. In these diagrams there is what we would characterizes as a small to moderate but nonetheless systematic drop off in visual attention.

{Approximate Position of Figure 4}

Next, Table 2 gives the results from a random effects (RE) model where log of total dwell on the *kth* attribute in the *sth* card ($\ln(Dwell_{k,s})$) is explained by the log of the card number (Card no) (i.e., where the card lies in the sequence presented to respondents) and an intercept. Both factors are random effects conditioned on the attributes. This yields a model of the following form:

$$\ln(Dwell_{k,s}) = \theta_{1k} + \theta_{2k}\ln(s) + e_{ks} \text{ where } (\theta_{1k}, \theta_{2k}) \sim N((\theta_1, \theta_2), \Psi) \tag{11}$$

In the model, we have logged the variables so as to make the coefficient θ_{2k} interpretable as an elasticity, although its significance is roughly unchanged if the data is not logged. Results are reported in Table 2.

Table 2: Proportional Dwell and Time

	Coeff	**St.Err**	**P value**
Intercept (θ_1)	10.663	0.229	0.000
ln($Card\ No$) (θ_2)	-0.126	0.051	0.013
Intercept RE (Ψ_{11})	2.022	1.522	0.184
Intercept RE x Card No RE (Ψ_{12})	0.240	0.281	0.393
Card No RE (Ψ_{22})	0.041	0.075	0.582

$R^2 = 0.605$

Note: RE - random effect; St Err - standard error

[15] All of our fixation results are available on request.

As we can see from Table 2, the parameter estimate on Card No is negative (-0.126) and the associated p value is 0.013 indicating that 'Card No' has a significant negative effect. Interpreting this as an elasticity implies that a 10% increase in cards leads to around a 1.2% decrease in visual attention. Thus, overall, on the basis of these results there is a relatively slow, but significant drop off in attention that would be consistent with respondent fatigue. However, although significant in the statistical sense 24 choice cards is a relatively long sequence compared to that usually employed in DCE studies where potential respondent fatigue is one of the reasons used to introduce fewer cards (i.e., fewer than 12). This finding is consistent also with the results reported by Meißner et al. (2016) who state that because of the repetitive experimental format being used with DCEs, respondents may become more effective (i.e., quicker) at completing discrete choice tasks. Thus, perhaps requiring less visual attention to complete the same tasks.[16]

Importantly, in examining our ET data, we note that the data for both the quality assurance (*QAS)* and organic (*ORG)* attributes have been 'corrected' in the sense that they have been scaled upwards because one of their alternatives was sometimes 'blank' and thus did not reflect a fair comparison. It was quite stark that when respondents were faced with an attribute level that was signified by a blank, they simply did not fix or dwell on these regions. This can be seen from the *ORG* plot in Figure 4, where *ORG* received virtually no visual attention on the third card. This is because it was the sole card where there was a choice between two conventionally produced Pizzas where conventional production was signified by the lack of an organic label and was thus left blank. Faced with no 'visual' stimulus, respondents did not pay visual attention to this space rather than fixing or dwelling on the blank region. For all the other cards except the first, one or other of the options were *ORG* and *QAS*, no other cards had both options 'blank'. This in itself, highlights the difficulty in inferring 'value' from attention. If we had of placed text or label saying 'conventional' then respondents would have fixed and dwelled upon these regions, but without an 'informational' advantage in the sense that is quite clear they were able to ascertain whether the option was Organic or conventional without this information.

Next we turn to the extent of visual attendance by attribute. These results are presented in Figure 5. In order for a card to be deemed visually attended, we required that a respondent fixed on the levels of both options, except where they were blank.

{Approximate Position of Figure 5}

The vertical axis indicates the number of individuals and the horizontal axis measures the total trials visually attended for each attribute. Thus, for example, the results shown that Price (*PR*) was visually attended on average at 95% for all 24 cards, by all survey participants (with a median of 100%). Thus, the vast majority of individuals visually attend Price for nearly all the cards. As is expected with stated preference studies, however, there are, at least

[16] Interestingly, for the COO attribute there is very little difference in dwell time regardless if presented with the Flag or No Flag treatment. Our expectation was that the flag would act as a visually 'salient' feature attracting attention as previously observed by Uggeldahl et al. (2016). However, there is little evidence that this is the case, and this aspect is consistent with our WTP estimate.

in this case, a very small minority of individuals that seem to ignore Price. This is important from a DCE perspective, since ANA of Price can undermine WTP estimates. In contrast, for some of the other attributes, we see much lower levels of attendance although these are not that low except for the Product Quality (PQU) attribute.

The investigation of importance of stated attendance has been an important component of the literature to date, and stated ANA data (i.e., where respondents reported that they ignored an attribute) was collected as part of the survey. This data is shown in Figure 6.

{Approximate Position of Figure 6}

The results presented in Figure 6 show that a high proportion of respondents (i.e., almost 80%) indicate that they do not attend at least one attribute throughout the DCE. However, the number who claim not to have attend more than one attribute drops very rapidly. There also tends to be a relationship between stated ANA, and other measures such as the stated rank importance of a given attribute. Here we extend this comparison by including the ET data and comparing it to the stated measures of attribute ANA and rank importance. The relationship between various measures of attribute use for the various data types we have collected is shown in Figure 7.

{Approximate Position of Figure 7}

What we observe in Figure 7 for Price and COO, is that the various measures appear to yield very similar results. However, in contrast we have very high levels of visual attendance for ORG but much lower measures for the other ANA data measures collected. This indicates that at least for ORG many respondents claim to not consider this attribute even when we see very high levels of visual attention.

We can enhance our understanding of these figures by examining 'consistency' statistics. Respondents would be labelled consistent if according to one measure (e.g., stated attendance) they then do not declare or exhibit a behaviour according to another measure (e.g., visual attention) that indicates that it is less important than another attribute. For our data, we can report that 89% of respondents are consistent in their stated ANA and rankings; 36% are consistent in their ANA and dwell; 36% are also consistent in their stated ANA and total fixations; and finally, 51% are consistent in their ANA and visual ANA (i.e., that is 49% visually attend attributes to a greater extent than ones they state they do not attend).[17]

Another way to examine the relationship between these measures is to estimate correlation coefficients between visual attendance and the stated measures of attribute use. These estimates are reported in Table 3:

[17] Note, when we examine the same relationships for the flag and no flag treatments, we find very similar levels for all measures and in all cases we find the same relationship between stated ANA and dwell and total fixations.

Table 3: Correlations between Visual Attendance and Stated Measures

	S-Att Corr	P value	Rank Corr	P Value
Price	0.304	0.002	0.372	0.000
COO	0.348	0.000	0.382	0.000
QAS	0.341	0.000	0.144	0.162
PQU	0.327	0.001	0.344	0.001
ORG	0.122	0.228	0.083	0.423

Note: S-Att - stated attendance; Corr - Correlation

As can be seen in Table 3 there is a positive correlation in each case and in the majority of cases this correlation is statistically significant at the 5% level. The exceptions are for the measures for the *ORG* attribute, which has an insignificant correlation with both stated attention and rank, and *QAS*, but in terms of our rank measure only. The correlations are, however, very low across the board.

Overall, these results are consistent with previous findings (Balcombe et al. 2015) which found a weak relationship between visual attendance and stated preferences. From the findings above, we would conclude that at the aggregate level there is a general tendency for attributes that have a high stated importance or attendance to be attended visually. However, this aggregate level tendency proves very weak when looking at individual behaviour. Knowing that somebody has visually attended an attribute, is only a weak stochastic signal that they will have stated that they attend that attribute (and vice versa). This observation can be further supported by the examination of particular individuals. While we cannot present them in the context of the paper, eye movements are recorded in the form of 'movies'. An examination of these proves most intriguing. For example, one individual stated that they did not attend Price, but attended all other attributes. However, this individual can be seen on multiple cards clearly fixing on Price, moving away to other attributes and then returning to Price again. Since we see no particular incentive to deliberately misreport their attendance, it would be interesting to understand why individuals can be so clearly orientated towards Price, yet perceive that they ignore it. In future work, we would suggest that key individuals should be invited back to review their responses and eye movements with a view to understanding the nature of these responses.

5.3. ET and WTP

The final, and most important piece of analysis we present are the results that assess the extent to which ET and WTP estimates are potentially correlated. For this component of the analysis we first compare the WTP estimates for each individual by attribute with their associated visual attention. Our measure of visual attention was 'proportional dwell' which was the proportion of the total time that an individual fixed on a given attribute. We note that substantively similar results are obtained by using proportional fixations, if absolute rather

than proportional measures are used, or if visual attendance is substituted for dwell. We do not present these results due to length constraints.

As we discussed in the Introduction, the potential value of ET data collection within a DCE is seriously increased if we can establish a relationship between ET data on dwell time (i.e., number of fixations) and WTP estimates. The pattern of this relationship for all of the attributes used in this study are shown below in Figure 8. The dependent variable here is the WTP for the difference between the most preferred attribute level relative to their least preferred for each of the four non-monetary attributes. We also include the coefficient of Price such that $WTP_{0,}$ is not actually a WTP, but rather the scale coefficient (i.e., the coefficient of Price).

{Approximate Position of Figure 8}

Within Figure 8, we present the scatter plots along with two regression lines, one using OLS and another from the 50% quantile regression, since it is evident that there are some large outliers in some of the regressions and the quantile approach is robust to these. What we can observe from Figure 8 is that for all of the attributes there is a (weak) positive relationship between WTP and dwell time for each attribute. The only exception to this appears to be for the QAS attribute. The R^2 and Psuedo R^2 reflect very small correlations although for the majority of cases these are significant (4 out of 5 for OLS and 3 out of 5 for the quantile regression).

An alternative way in which we can investigate the overall relationship between visual attention and WTP is to estimate a mixed model regression specification (i.e., containing both fixed and random effects). In this case, we assume random effects for individuals on each of the attribute intercepts. The regression we consider takes the following form:

$$WTP_{k,j} = \sum_{k} \phi_{k,j} + \omega(\text{Prop Dwell}_{kj} + e_{ki} \text{ where } \left(\phi_{1,j}, ...\phi_{K,j}\right) \sim N\left(\phi, \Theta\right) \tag{12}$$

where WTP_{kj} is the jth person's WTP to receive the most preferred level over the least preferred level of attribute k, and Prop Dwell is the proportional dwell on the kth attribute by the jth person. The main results of model are reported in Table 4:[18]

Table 4. Mixed Model Results for WTP

	Coeff	**Std.Err**	**P value**
Prop Dwell (ω)	0.812	0.094	0.000
Price (ϕ_o)	0.567	0.044	0.000
COO (ϕ_1)	1.038	0.043	0.000
QAS (ϕ_2)	0.602	0.028	0.000
PQU (ϕ_3)	0.687	0.039	0.000
ORG (ϕ_4)	0.864	0.064	0.000
$R^2 = 0.96$			

[18] Although, we only report the main effects herethe full set of results including interactions are available on request.

As can be seen in Table 4 the proportional dwell coefficient is positive and highly significant reinforcing the results we have presented in Figure 8.

Taken together, the results presented in Figure 8 and Table 4 allow us to address the questions we previously raised regarding the use of ET data within a DCE. Let us start with our third question that considered, if an attribute is more highly attended (or paid more attention) by an individual (relative to other individuals) does that mean that an individual values it more highly than other people. The positive relationships observed in Figure 8 would suggest that the answer is, yes. Of course, the extent to which this relationship is revealed by our data is not that strong but it does suggest that dwell time does reveal something about how an individual values a specific attribute.

Turning to our first question, we considered if 'higher value' attributes attract more visual attention (*ceteris paribus*) than others. Our general results would appear to give qualified support to this contention, given the tendency for highly visually attended items to have higher stated attendance and rankings and given the weak correlation of attention with WTPs. However, the weak positive relationship does not mean that relatively higher valued attributes are more attended. Indeed, the range of WTP estimates for the *ORG* attribute is greater than any others but the relative level of visual attention (i.e., dwell time) is relatively low. Thus, in this case even though this attribute has a high (absolute) WTP (and very high in some cases) the relative level of dwell time is not high. This then means that there need not be any reason why attributes that are relatively highly visually attended are valued more highly than others. This in turn means that the answer to our second question is a qualified 'no' in the sense that the underlying signal being provided by attention or attendance can be swamped by other factors. From the point of view of an individual, attention or attendance may be governed by not only the importance of the information provided but by its form and whether this can be assimilated easily. Therefore, it would be highly speculative to assume that high visually attended items are of more value as measured by our WTP estimates.

6. Discussion and Conclusions

In this paper, we have examined consumer preferences regarding COO for meat using a DCE. We have analysed our data using a BIML specification. In addition, we have employed ET methods so as to assess the extent to which visual attendance and attention of attributes help to explain reported engagement with the survey instrument as well as individual WTP results.

First, the BIML model specification suggests that the shape of the WTP distributions are non-standard in the sense that it would be difficult to *a priori* to select a set of distributions to capture this aspect of the data. These results have emerged from a BIML model specification with an average of 10 mixtures. Thus, by employing the BIML specification, we have reduced possible model specification biases that might emerge from using existing econometric specifications. These findings support those reported by Train (2016) and related earlier research. In particular, the ability of the researcher to no longer need to specify *ex-ante* a particular probability distribution with which to model preference heterogeneity reduces another source of possible bias in model specification. Of course, with regard to the BIML there are assumptions

that are made with regard to priors but there are also many assumptions made by Train (2016). Clearly, both approaches warrant further scrutiny so that we better understand the importance of these model assumptions on model performance and results.

Second, our results reveal the extent to which stated measures of DCE attribute use and non-use relate to the measures provided by the ET data. We find that there is, overall, a reasonable correspondence between ET data and other measures of attribute use such as the frequently employed debriefing questions that have become widely reported in the literature. But, our results confirm once again, that at the individual level stated attendance is a very weak signal in relation to visual attendance and vice versa.

Third, we find evidence of longer engagement with high value attributes, as measured by total dwell time as well as total number of fixations. This relationship exists, but it is quite weak. This result bolsters existing work that suggests that ET data does reveal something about how respondents value the attributes used in a specific DCE. However, does this result represent a sufficient increase in our understanding of choice behaviour to warrant the extensive use of ET? As already discussed, the ET data has proven very useful in extending our understanding of ANA but less revealing in terms of insights gained with regard to WTP for the attributes.

So what do these results tell us about how to employ and use ET data within a DCE in the future? If ET was costless and easily implemented, we would recommend that it was employed universally. However, this is currently not the case. As is clear from the literature ET data collection is time consuming, sample sizes are relatively small and in contrast to standard DCE applications many aspects of implementation are still subject to debate about what exactly to measure and how. Thus, if the purpose of generating ET data is to improve the efficiency of estimation then we would recommend increasing sample size as a better strategy to pursue. Therefore, given the associated costs of using ET technologies and the potential benefits on offer, at this stage in the development and use of ET, we argue that ET research efforts must either be directed at improving our understanding of decision making at the process level, and/or as an activity that is largely undertaken as part of the improved design and piloting of a DCE.

With regard to decision making at the process level our thoughts about the benefits from undertaking ET are part of a wider debate about how to use process data in economics. Even though the use of ET is no longer considered to be an issue within economics there remain unanswered questions with regard to what we should and can do with the ET data. As Orquim and Muller Loose (2013, page 201) argue, the DCE ANA models will only start to make an increased contribution to the literature when economics researchers attempt to explain why ANA occurs and this will require an understanding of decision (cognitive) processes. This observation is important as the generation of ET data as part of a DCE is rapidly growing, yet there appear to be issues regarding what exactly researchers should be doing with the data. If we are to move beyond the current *ad hoc* applications then one possible avenue of research that warrants more consideration is how process data (generally defined) is being used to understand cognitive limitations as they relate to choice (Caplin, 2016; Seng, 2016). However, attempting to draw on this research will not be easy as the choice context considered within DCE is far more complex than those typically considered within the attention literature in economics to date.

Turning to how to make use of ET as part of the survey design process, this perspective stems from the observation that amongst our sample of respondents there are a number who provided inconsistent stated and ET information. There is reason to assume that this has possibly occurred because the respondents in question have failed in some way to fully appreciate or understand aspects of the task in hand. Indeed, it would be potentially interesting to gain a qualitative understanding of why these inconsistencies have emerged by asking these respondents additional de-briefing questions. The expectation would then be that in light of this enhanced understanding, we re-design the survey so as to try and minimise inconsistent behaviour prior to the survey being implemented in full. This use of ET could provide a way in which we can effectively incrementally design a DCE so as to reduce (or minimise) the extent of visual ANA or perhaps maximise attention across all attributes.

Beyond these two proposed uses of ET data there are also a number of important ways in which ET data collection might be implement to enhance our understanding of visual attention. For example, an important feature of the results we present is that, due to the nature of the experiment, we are unable to conclude about the causality of the relationship between visual attention and attributes valuation. Although, we have discussed our results assuming that attributes' values are driven visual attention (i.e., the more I like it, the more I look at it), we cannot rule out the possibility that in fact the opposite holds (i.e., the more I look at it, the more I like it). If we are to examine this specific issue then it would be necessary to design a DCE such that, for example, we employ a mechanism that allows to expose a group of survey respondents to certain visual clues in advance of the DCE and then another group who are not exposed to the same visual clues.

There are also additional reasons why we need to pay more attention to experimental design when using ET within a DCE. As we discuss, visual attention is driven by many factors with attributes' value being only one component. Of the other factors that might impact on visual attention, it is almost certainly the case that we would need to explore between- and within-subjects variability in measures of visual attention to see if we could isolate specific factors.

Another important issue that deserves more attention are the different ways we can examine ET data to examine attribute use. For example, it has become reasonably common to define an AOI but it is unclear how the AOI relates to the design of the survey instrument in a graphical sense. Almost certainly beyond the DCE literature there is an understanding of how a specific pictorial or visual design will help with a sharper delineation of an AOI and drawing on this research will very likely reduce noise in ET data that may in part be a function of survey instrument design. In fact, in would make far more sense to identify AOI *ex-ante*, as part of the survey design process and not *ex-post* once the survey has been completed.

Finally, there is also a need to establish best practice when it comes to the how we define specific measures from our ET data. At present there is much heterogeneity in the literature which simply occurs because of the relatively rapid adoption and use of ET within DCE. This means that because current practice is very varied and this makes like for like comparisons of papers/research difficult. For example, there are different ways we can use ET to assess ANA e.g., Balcombe et al. (2015) and Krucien et al. (2017). At this stage this heterogeneity is useful as we strive to improve our understanding of what to do with ET data, but at some point if

researchers can agree on appropriate methods of measurement then it is likely that replication will yield an enhanced understanding of decision making in DCE.

References

Balcombe K.G., Fraser I.M. and McSorely, E. (2015). Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments. **Journal of Applied Econometrics**, 30(3): 447-467.

Balcombe, K.G., Bradley, D., Fraser, I.M. and Hussein, M. (2016a). Consumer Preferences Regarding Country of Origin for Multiple Meat Products. **Food Policy**, 64: 49-62.

Balcombe K.G., Fraser, I.M., Souza-Monteiro, D. and Lowe, B. (2016b). Information Customization and Food Choice, **American Journal of Agricultural Economics,** 98(1): 54-73**.**

Burda, M., Harding, M. and Hausman, J. (2008). A Bayesian Mixed Logit-Probit Model for Multinomial Choice, **Journal of Econometrics**, 147: 232-246.

Caplin, A. (2016). Measuring and Modeling Attention. **Annual Review of Economics,** 8: 379-403.

Caplin, A. and Dean, M. (2015). Enhanced Choice Experiments. In **Handbook of Experimental Economics Methodology**. Edited by G.F. Frechette and A. Schotter. Oxford University Press.

Caplin, A. and Martin, D. (2016). The Dual-Process Drift Diffusion Model: Evidence from Response Times. **Economic Inquiry**, 54(2): 1274-1282.

Daziano, R.A. (2013). Conditional-Logit Bayes Estimators for Consumer Valuation of Electric Vehicle Driving Range. **Resource and Energy Economics**, 35: 429-450.

Escobar, M.D. and West, M. (1995). Bayesian Density Estimation and Inference Using Mixtures. **Journal of the American Statistical Association**, 90(430): 557-558.

Findlay, J. M. (2009). Saccadic eye movement programming: Sensory and attentional factors. **Psychological Research**, 73: 127-135.

Geng, S. (2016). Decision Time, Consideration Time, and Status Quo Bias. **Economic Inquiry**, 54(1): 433-449.

Grebitus, C., Roosen, J. and Seitz, C.C. (2015). Visual Attention and Choice: A Behavioural Economics Perspective on Food Decisions. **Journal of Agricultural and Food Industrial Organization**, 13(1): 73-81.

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

Krajbich, I. and Smith, S.M. (2015). Modeling Eye Movements and Response Times in Consumer Choice. **Journal of Agricultural and Food Industrial Organization**, 13(1): 55-72.

Krucien, N., Ryan, M. and Hermens, F. (2017). Visual Attention in Multi-Attribute Choices: What Can Eye-Tracking Tell Us? **Journal of Economic Behavior and Organization**, 135: 251-267.

Lahey, J.N. and Oxley, D. (2016). The Power of Eye Tracking in Economics Experiments. **American Economic Review: Papers and Proceedings**, 106(5): 309-313.

Lewis, K.E., Grebitus, C. and Nayga Jr, R.M. (2016). The Impact of Brand and Attention on Consumers' Willingness to Pay: Evidence from an Eye Tracking Experiment. **Canadian Journal of Agricultural Economics**, 64: 753-777.

Matin, E. (1974). Saccadic suppression: a review and an analysis. **Psychological Bulletin**, 81:899-917.

McAuliffe, J. D., Blei, D. M. and Jordan, M. I. (2006). Nonparametric empirical Bayes for the Dirichlet process mixture model. **Statistics and Computing**, 16(1): 5-14.

Meißner, M., Musalem, A. and Huber, J. (2016). Eye Tracking Reveals Processes that Enable Conjoint Choices to Become Increasingly Efficient with Practice. **Journal of Marketing Research**, LIII(February): 1-17.

Neal, R.M. (2000). Markov Chain Sampling Methods for Dirichlet Process Mixture Models. **Journal of Computational and Graphical Statistics**, 9(2): 249-265.

Orquin, J.L. and Muller Loose, S. (2013). Attention and Choice: A Review on Eye Movements in Decision Making. **Acta Psychologica**, 144: 190-206.

Oviedo, J.L. and Caparros, A. (2015). Information and Visual Attention in Contingent Valuation and Choice Modelling: Field and Eye-Tracking Experiments Applied to Reforestations in Spain. **Journal of Forest Economics**, 21: 185-204.

Rasch, C., Louviere, J.J. and Teichert, T. (2015). Using Facial EMG and Eye Tracking to Study Integral Affect in Discrete Choice Experiments. **Journal of Choice Modelling**, 14: 32-47.

Rayner, K. (2009). Eye movements and attention in reading, scene perception and visual search. **The Quarterly Journal of Experimental Psychology**, 62: 1457-1506.

Rihn, A., Khachatryan, H., Campbell, B., Hall, C. and Behe, B. (2016). Consumer Preferences for Organic Production Methods and Origin Promotions on Ornamental Plants: Evidence from Eye-Tracking Experiments. **Agricultural Economics**, 47: 599-608.

Rihn, A.L. and Yue, C. (2016). Visual Attention's Influence on Consumers' Willingness-to-Pay for Processed Food Products. **Agribusiness**. 32(3): 314-328.

Spinks, J. and Mortimer, D. (2016). Lost in the Crowd? Using Eye-Tracking to Investigate the Effect of Complexity on Attribute Non-Attendance in Discrete Choice Experiments. **BMC Medical Informatics and Decision Making**, 16: 14.

Train, K.E. (2008). EM algorithms for nonparametric estimation of mixing distributions. **Journal of Choice Modelling,** 1:40-69.

Train, K.E. (2016). Mixed logit with a flexible mixing distribution. **Journal of Choice Modelling**, 19:40-53.

Train, K.E. and Weeks, M. (2005). Discrete Choice Models in Preference Space and Willingness-to-Pay Space. Chapter 1, pages 1-16, **Applications of Simulation Methods in Environmental and Resource Economics**, Edt. By Alberini, A. and Scarpa, R., Kluwer Academic Publishers.

Uggeldahl, K., Jacobsen, C., Lundhede, T.H. and Olsen, S.B. (2016). Choice Certainty in Discrete Choice Experiments: Will Eye-Tracking Provide Useful Measures? **Journal of Choice Modelling**, 20: 35-48.

Van Loo, E.J., Caputo, V., Nayga Jr., R.M., Seo, H-K., Zhang, B. and Verbeke, W. (2015). Sustainability Labels on Coffee: Consumer Preferences, Willingness-to-Pay and Visual Attention to Attributes. **Ecological Economics**, 118: 215-225.

Woodford, M. (2014). Stochastic Choice: An Optimizing Neuroeconomic Model. **American Economic Review: Papers & Proceedings**, 104(5): 495-500.

Figure 1: Example Choice Card

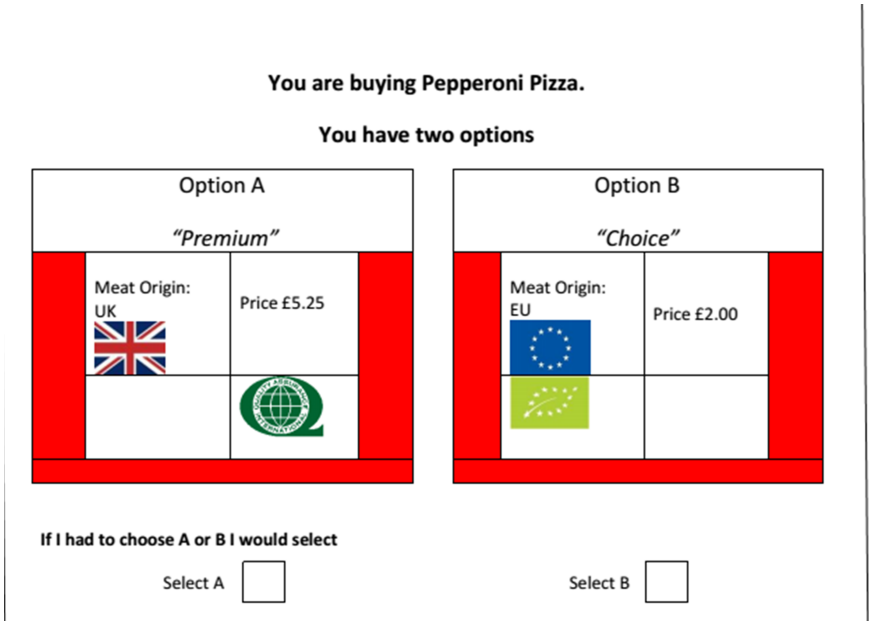


Figure 2: Distribution of Number of Mixtures

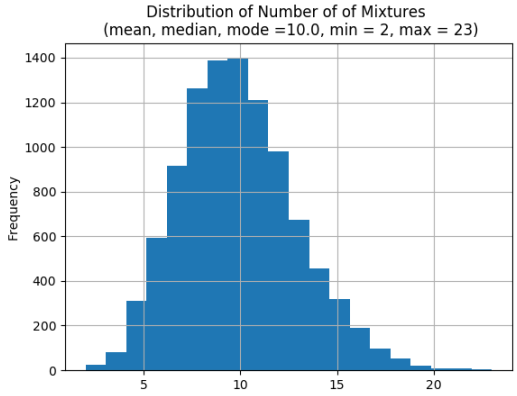
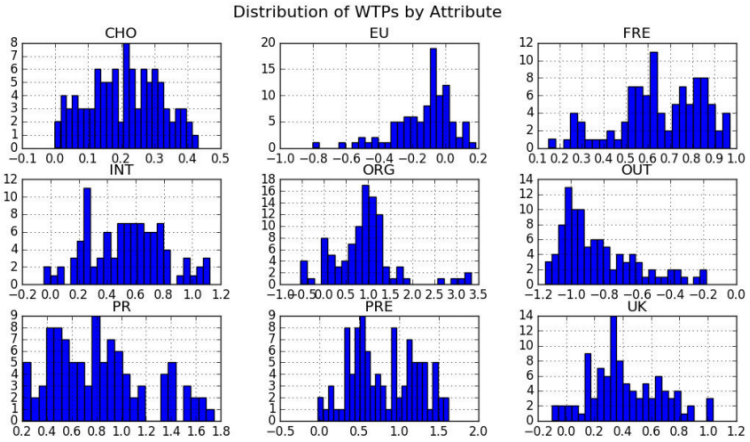
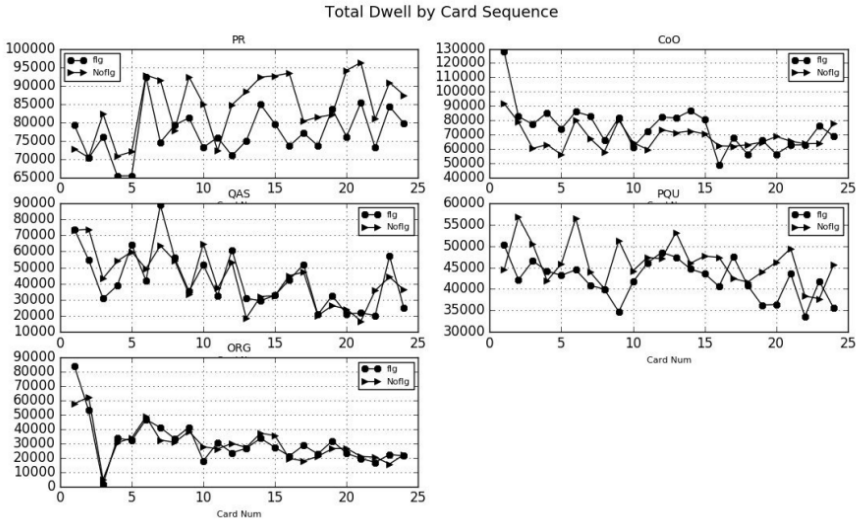


Figure 3: Distribution of WTPs by Attribute



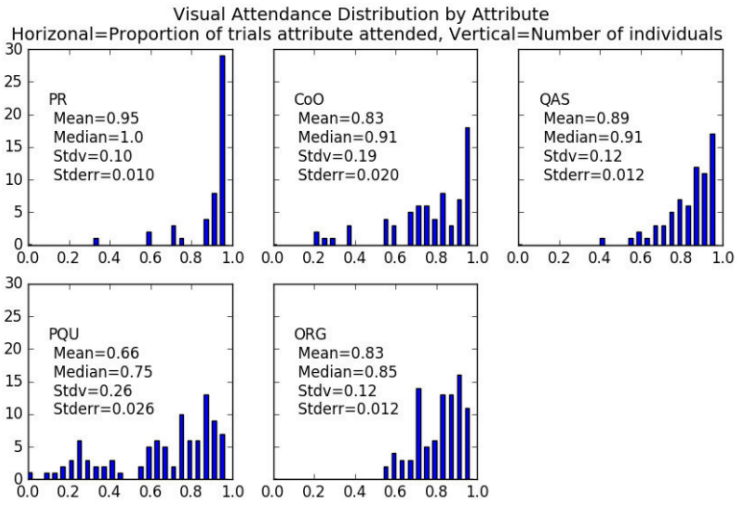
Note: CHO - Choice (Product Quality); EU - European Union (COO);
FRE - Freedom Food (Quality Assurance); INT - International (Quality Assurance);
Org - Organic (Production System); OUT - Outside of the EU (COO);
PR - Price; PRE - Premium (Product Quality); UK - United Kingdom (COO).

Figure 4:Total Dwell by Choice Set



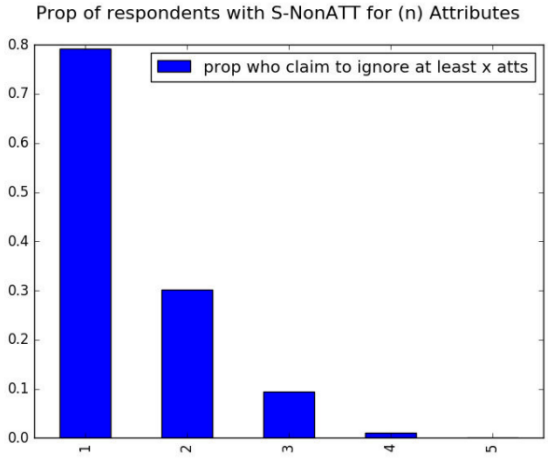
Note: PR - Price; COO - Country of origin; QAS - Quality assurance; PQU - Product quality; Org - Organic.

Figure 5: Visual Attendance By Attribute



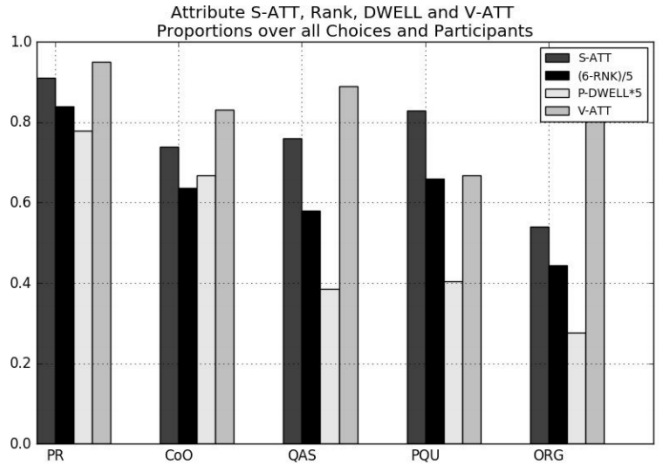
Note: PR - Price; COO - Country of origin; QAS - Quality assurance; PQU - Product quality; Org - Organic.

Figure 6: Proportion of Respondents with Stated ANA for DCE Attributes



Note: S-NonAtt - Stated non-attendance; prop - proportion

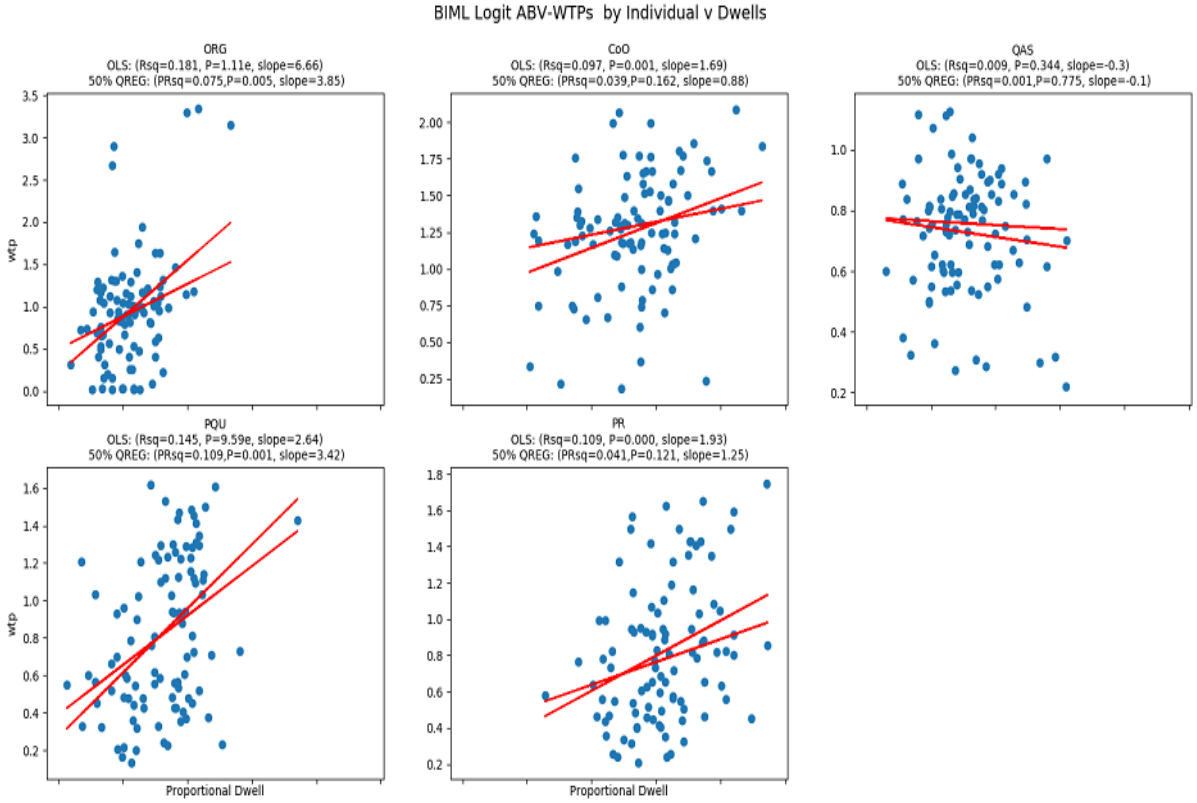
Figure 7: Attribute Stated Attendance, Rank, Dwell and Visual Attendance



Note: PR - Price; COO - Country of origin; QAS - Quality assurance; PQU - Product quality; Org - Organic;

S-ATT - stated attendance; V-ATT - visual attendance.

Figure 8: Relationship Between Dwell Time per Attribute and WTP.



Note: PR - Price; COO - Country of origin; QAS - Quality assurance; PQU - Product quality; Org - Organic.

Appendix A

A1: What is Eye-Tracking

Individuals will their eyes when confronted with visual stimuli because sharpness across the retina quickly diminishes with distance from the fovea (this is the part of the eye responsible for processing detailed visual information). Given that only two percent of a respondents' visual field will be projected onto the fovea, for a respondent to 'attend' an object, their eyes must move. It is for this reason that examination of eye movements can (in principle) help in our understanding of how information is obtained.

It is standard to classify eye movements into two types:

1. Fixations – these occur when eye movements are relatively still with durations of between 200-500 milliseconds such that a contiguous area is projected onto the fovea leading to detailed visual processing.

2. Saccades – very rapid movements of between 20-40 milliseconds that help project specific locations of a given visual scene onto the fovea.

Eye-tracking research aim to understand how the brain deals with visual information received. This information, which is transmitted via the optic nerve, is dealt with using various attentional mechanisms that aid in the selection of a subset of relevant information that is subject to enhanced processing. This means that the brain is simultaneously enhancing and suppressing information.

In normal viewing situations attention and eye movements are intimately linked and move in tandem to the same visual location (Findlay, 2009). This comes from evidence examining the close correspondence between eye movements and higher-order cognitive processes (e.g., Rayner, 2009). As such eye-tracking research has yielded insights into the control of visual attention (Findlay, 2009).

In practice, much eye-tracking research looks for patterns based on fixations and saccades. The eye-tracking technology records patterns of movements and pauses, while respondents consider a visual stimulus. These patterns can then collected together in what is referred to as a scanpath which provides spatial-temporal data on the spatial distribution of attention across the visual stimulus. Therefore, eye fixation is in principle a good indicator of visual attention because:

(i) acuity deteriorates rapidly outside the fovea;

(ii) little visual information can be obtained during saccades (Matin, 1974); and

(iii) fixation and attention are naturally yoked.

A2: Technical Details of Implementation

The choice cards were presented on a 21 inch colour monitor with a refresh rate of 75 Hz (DiamondPro, Sony). Each card subtended 16.7 by 12.7 degrees of visual angle as viewed from 1 metre. All stimuli were presented on a white background. Eye movements were recorded using a head-mounted, video-based, eye-tracker with a sampling rate of 500 Hz (Eyelink II, SR Research), recording monocularly from the respondents' right eye.

Head movements were constrained with a chin-rest, which held the participant so that their eyes were in line with the horizontal meridian of the screen. Choices were recorded through a response gamepad. The eye-tracker was calibrated using a standard 9 point grid, carried out at the beginning of the experiment. Calibration was accepted only once there was an overall difference of less than 0.5 degrees between the initial calibration and a validation retest. In the event of a failure to validate, calibration was repeated. In order to ensure that accuracy was maintained throughout the CE a drift correction was carried out between each card viewing. Participants were asked to view a spot stimulus and press a button when they were fixating its centre. The drift correct stimulus consisted of a small black annulus that gave the appearance of a small black spot (0.5 centimetres in diameter) with a smaller white spot in the centre (0.25 cm diameter) shown in the upper left quadrant of the screen off set from the centre by 5.12 degree horizontally and 3.86 degree vertically. This procedure minimized the effects of slight movement of the head impacting on the accuracy of the eye-tracking.

Once participants were comfortable in the eye-tracker and their eye movements calibrated, they were presented with the choice cards. Participants viewed the cards for as long as the y wished while we tracked their eye movements. They responded with a button press indicating which basket they selected. A drift correct stimulus was then shown until a button press from the participant indicated they were looking at it. The next choice card was then shown.

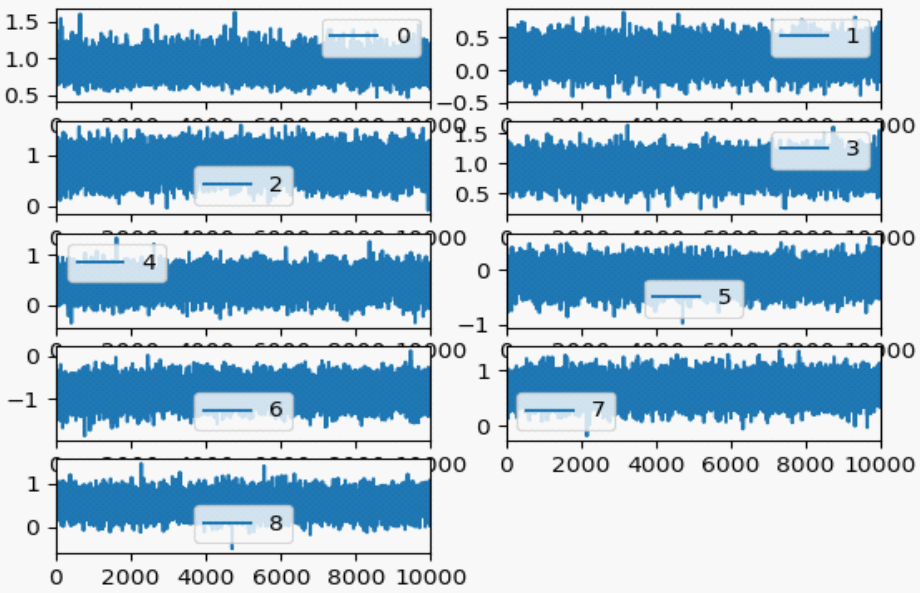
A3: Convergence Diagnostics

Table A.3 Convergence Diagnostics for BIML parameters

	*Diff	t-val	P-Val	Autocorr
Price (Scale)	0.0057	1.6214	0.0531	0.2043
Choice (v Basic)	0.0006	0.1565	0.4379	0.0089
Premium (v Basic)	0.0032	0.6264	0.2658	0.0672
Organic (v Conventional)	0.0025	0.2591	0.6469	0.0484
UK (v Italy)	-0.0004	-0.1036	0.4588	0.0363
EU (v Italy)	0.0012	0.2914	0.3855	0.0144
USA (v Italy)	0.0049	0.8940	0.1861	0.0772
Freedom Food (v None)	-0.0026	-0.6221	0.2672	0.0378
International (v None)	0.0020	0.4345	0.3322	0.0466

Note: *Diff is the mean difference between separated draws from first and second half of the sampler The t and P-Values test for a zero difference, Autocorr is first order autocorrelation

Figure A3: The Trace Plots for the 9 coefficients in DCE



Note: 0-8 indicate the same ordering of as in Table A.3