Inferring Attribute Non-Attendance Using Eye Tracking in Choice-Based Conjoint Analysis

Narine Yegoryan (Humboldt University Berlin)
Daniel Guhl (Humboldt University Berlin)
Daniel Klapper (Humboldt University Berlin)

Discussion Paper No. 111

August 3, 2018

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Inferring Attribute Non-attendance using Eye tracking in Choice-based Conjoint Analysis

Narine Yegoryan[a,*] Daniel Guhl[a] Daniel Klapper[a]

[a] *Institute of Marketing, School of Business and Economics, Humboldt University Berlin, Spandauer Str. 1, 10178 Berlin, Germany.*

[*]Corresponding author: Tel.: +49 30 2093 99484.
E-mail addresses: narine.yegoryan@hu-berlin.de (Narine Yegoryan), daniel.guhl@hu-berlin.de (Daniel Guhl), daniel.klapper@hu-berlin.de (Daniel Klapper).
Funding: Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged

Abstract

Traditionally, the choice-based conjoint analysis relies on the assumption of rational decision makers that use all available information. However, several studies suggest that people ignore some information when making choices. In this paper, we build upon recent developments in the choice literature and employ a latent class model that simultaneously allows for attribute non-attendance (ANA) and preference heterogeneity. In addition, we relate visual attention derived from eye tracking to the probability of ANA to test, understand, and validate ANA in a marketing context. In two empirical applications, we find that a) our proposed model fits the data best, b) the majority of respondents indeed ignores some attributes, which has implications for willingness-to-pay estimates, segmentation, and targeting, and c) even though the latent class model identifies ANA well without eye tracking information, our model with visual attention helps to better understand ANA by also accounting for differences in attribute processing patterns.

Keywords: Attribute non-attendance; Eye tracking; Discrete choice modeling; Choice-based conjoint analysis

1 Introduction

Choice-based conjoint analysis (CBC) is a popular tool in marketing used to elicit consumer preferences, predict consumers' response to new product introductions, identify segments that similarly value product attributes, and optimize product design, targeting and pricing strategies (Rao 2014). The basis for analyzing the observed choices from CBC is the random utility maximization (RUM) model, with consumer "rationality" as its key behavioral pillar (McFadden 2001). More explicitly, under RUM, consumers are considered to have stable preferences, process all available information, and select the option that maximizes their utility. However, the validity of these assumptions has been largely challenged (e. g., DellaVigna 2009). Particularly regarding the assumption of full information processing, it has long been argued that due to limited cognitive capacity, individuals often simplify their choices (e. g., Payne, Bettman, Coupey, and Johnson 1992) and ignore some information about product attributes or alternatives (e. g., Orquin and Loose 2013).

The marketing literature has been mainly interested in the case where consumers neglect some of the available alternatives, which is conventionally labeled as "choice set" formation (e. g., Swait and Ben-Akiva 1987; Bronnenberg and Vanhonacker 1996). This issue is specifically common in revealed preference data (i. e., market data), where consumers face dozens or even hundreds of product options. In typical CBC settings, alternatives are usually restricted to a manageable number, e. g., three to four, and include a higher number of attributes (Rao 2014). However, some respondents may not deem all the attributes to be relevant or may tend to ignore them due to choice task characteristics, e. g., complexity (Hensher 2014). Hence, in the CBC, it seems more plausible that respondents ignore attributes, conventionally termed "attribute non-attendance" (ANA)[1], rather than alternatives. In turn, in a given context, non-attended attributes do not contribute to the utility of a particular individual, implying that the corresponding preference parameters in the utility

[1]Note that the term "non-attendance" comes mainly from transportation science and in this context is used as a synonym for "ignoring" or "not considering" an attribute in the decision-making. To avoid being at odds with the main body of ANA literature, we adopt the corresponding terminology.

specification should be zero. In light of ANA, models assuming the use of all attributes in decision-making (hereafter "full (attribute) attendance"[2]) may result in biases in parameter estimates and, subsequently, in the derived willingness-to-pay (WTP) and welfare estimates (e. g., Gilbride, Allenby, and Brazell 2006; Scarpa, Gilbride, Campbell, and Hensher 2009). Given the plausibility and implications of ANA, in the following paper, we focus on its prevalence in a marketing context.

One of the most promising ideas for tackling ANA, mainly issuing from fields such as transportation, environmental, and health economics, is the use of a latent class model, where each a priori defined class represents a specific attribute attendance/non-attendance pattern (hereafter "attribute processing strategy") and hence a different utility specification (e. g., Hole 2011; Hess, Stathopoulos, Campbell, and Caussade 2013). The main advantage of this approach is that ANA can be inferred based on the observed choices alone instead of relying on auxiliary information such as respondents' self-stated measures (e. g., Hensher 2006) or proxies of ANA derived from process-tracing techniques (e. g., Currim, Mintz, and Siddarth 2015; Meißner, Scholz, and Decker 2011). Such information can still be useful, and using it to augment the models for inferring ANA is a promising area, allowing a better understanding of the underlying individual behavior. For example, Hole, Kolstad, and Gyrd-Hansen (2013) and Hess et al. (2013) demonstrate the benefits of using a stated ANA measure coupled with the latent class model. Nevertheless, an open question remains: how beneficial would process-tracing measures, specifically eye tracking, be within such a modeling framework for identifying ANA? This is precisely the focus of the current paper.

We build on Hole et al. (2013) and use a latent class approach that allows for the simultaneous inference of ANA and preference heterogeneity and, in addition, integrates information from eye tracking. In doing so, our objective is to better understand and capture different attribute processing strategies individuals may apply when making choices. The measures derived from eye tracking, being representative of underlying cognitive processes (e. g., Wedel

[2]Note that as we assume all alternatives are considered, "full (attribute) attendance" is equivalent to full information processing or full compensatory decision rule.

and Pieters 2008) and indicating the relevance of information (e. g., Meißner and Oll 2017), are best suited for this purpose. Moreover, as Meißner, Musalem, and Huber (2016) demonstrate, individuals become more efficient and selective in information processing during the CBC tasks. Thus, eye tracking can be particularly informative in uncovering ANA in a CBC. Furthermore, eye movements are driven by both top-down (e. g., consumers' goals, traits, emotions) and bottom-up (e. g., salience, location, features of the stimuli) factors (Wedel and Pieters 2008). As such, without making a distinction, they may allow capturing different drivers of ANA, i. e., person- (e. g., true irrelevance) and task-related (e. g., complexity).

Our second objective is to understand the effect of visual attention on consumers' actions, i. e., choice. In particular, we model this relationship so that visual attention affects the likelihood of attending an attribute. Subsequently, the attended attributes enter the utility function, are traded-off against each other and affect choice. Additionally, we investigate whether this relationship varies across attributes.

Third, we are interested in understanding the prevalence of ANA in a marketing context, where typically we observe varying levels of complexity in the choice task (e. g., many product features and alternatives), consumer involvement, knowledge, and risks associated with the product category (e. g., buying a car involves higher stakes than buying a pair of headphones). Additionally, we aim to assess the consequences of neglecting ANA for managerially relevant measures, such as WTP.

In two empirical applications, we indeed find evidence that individuals ignore attributes, with the majority attending to only three to four out of six available and almost no one attending to all. We demonstrate that neglecting ANA results in substantial biases in preference parameters and, accordingly, in derived aggregate and individual-level measures such as the relative importance of attributes and WTP. Moreover, we find a positive and significant effect of visual attention on the likelihood of attending an attribute and demonstrate that eye tracking is helpful in determining the allocation of individuals into and the size of the segments, which describe specific attribute processing strategies.

We, therefore, contribute to the literature in several ways. First, we provide further empirical evidence of ANA in two different marketing contexts and outline the implications of the strict assumption of full attribute attendance. Second, we contribute by proposing a novel framework of how visual attention might affect choices through its implicit link to the relevance of and subsequent attendance to attributes. This further allows investigating attribute-specific differences in how attention translates into attendance. Third, we provide further validation of methods for inferring ANA based on the observed choices. We find that relying only on the observed choices can be sufficient for recovering general patterns in applied attribute processing strategies and the distribution of WTP.

The rest of the manuscript is structured as follows. In Section 2, we review the relevant literature on existing approaches to incorporating ANA as well as on the use of eye tracking in studying decision-making and choice. Subsequently, we describe the methodology, including our main models as well as benchmark models, the derivation of the visual attention measure, the estimation procedure as well as measures of interest derived from the obtained parameter estimates. In Section 4, we present and discuss the results of two empirical applications. The paper concludes with a summary and an outline of avenues for future research.

2 Relevant Literature

Our study links eye tracking with discrete choice models that account for ANA. Therefore, in the following section, we concisely review the literature on methods that explicitly account for ANA and outline general trends and recent developments in this area. Furthermore, we provide an overview of the eye tracking research in relation to decision-making and choice and outline some important findings.

2.1 Methods to Account for ANA

To date, in the existing literature, two main approaches accommodating ANA can be outlined (Hensher 2014). One approach, which we will refer to as exogenous, solely relies on supplementary data collected during an experiment such as stated ANA (e. g., Hensher 2006), attribute importance ranking (e. g., Hess and Hensher 2013) from debriefing questions, click data from Mouselab experiments (e. g., Currim et al. 2015) and measures derived from eye tracking (e. g., Balcombe, Fraser, and McSorley 2015; Meißner et al. 2011). The preference parameters in the random utility specification are then conditioned on these measures by setting them to zero (e. g., Hensher, Rose, and Greene 2005), rescaling them downwards ("shrinking"; e. g., Balcombe et al. 2015), or estimating a separate set of parameters for non-attenders (e. g., Hess and Hensher 2010). Each of the auxiliary measures used has certain limitations. In particular, stated ANA is a subjective measure and depends on the recall, belief, and motivation of the respondents (Hess and Hensher 2010). On the other hand, Mouselab experiments may influence the respondents' information search process (andreas and Betsch 2008). In contrast, in the case of the more objective eye tracking measure (Meißner and Oll 2017), one needs to derive a discrete measure for use as a proxy for ANA. For example, Balcombe et al. (2015) use fewer than two fixations as an indicator of ANA in a given choice task. If the attribute was not attended to in more than half of the choice tasks, it is considered non-attended throughout the choice experiment. However, the choice of the cutoff in each and across all choice tasks may influence the model outcomes. Moreover, attribute-specific cutoffs might be more suitable, as some attributes may require more "looking", depending on how they are presented (e. g., as a picture, number, text, font size, etc.)[3]. In the case of all of the measures, a common limitation of the exogenous methods remains their deterministic use (i. e., assuming a one-to-one relationship with ANA).

To address this limitation, several scholars have proposed inferring ANA from the ob-

[3]One way of identifying optimal cutoffs could be by, e. g., employing a grid-search. However, the optimization problem can become rather complex in the case of attribute-specific cutoff values.

served choices rather than solely relying on additional data (Hensher 2014). We refer to this class of approaches as endogenous methods. Within this framework, e. g., Hess and Hensher (2010) suggested inferring ANA on the basis of high dispersion of the individual-level conditional parameter distribution. In contrast, e. g., Scarpa et al. (2009) and Hole (2011) propose a latent class approach probabilistically allocating individuals into a priori defined classes that are based on (many or) all possible attendance/non-attendance combinations, i. e., attribute processing strategies. This approach was shown to outperform the exogenous approach relying on the stated ANA measure (e. g., Scarpa, Zanoli, Bruschi, and Naspetti 2013). Endogenous models were further developed to simultaneously accommodate heterogeneity in individual preferences (e. g., Gilbride et al. 2006; Hess et al. 2013; Hole et al. 2013). As Hess et al. (2013) and Hole et al. (2013) demonstrate, neglecting preference heterogeneity may result in an overstatement of the amount of ANA, as the model may not correctly distinguish between zero and low sensitivity.

A few scholars have proposed further augmenting these models by conditioning class allocation on auxiliary information, thus bridging the gap between the exogenous and endogenous approaches. For example, Swait, Popa, and Wang (2016) have used complexity measures, potentially capturing task-driven ANA. However,Alemu, Morkbak, Olsen, and Jensen (2013), using debriefing questions, establish that true irrelevance of the attributes is a common reason for ANA. Heidenreich, Watson, Ryan, and Phimister (2018) further find higher levels of ANA for respondents who are more familiar with the particular product category. In contrast, Hole et al. (2013) and Collins, Rose, and Hensher (2013) use stated ANA as a covariate, which should capture different drivers of ANA. Nevertheless, the objectivity and reliability of this measure remain an issue.

We build upon the outlined developments in the ANA literature and adopt an endogenous approach for incorporating ANA. We simultaneously account for heterogeneity in preferences (Hole et al. 2013) and, in contrast to the existing literature, condition the class allocation on a measure of visual attention derived from eye tracking. Notably, the main body of research on ANA streams from other fields, including transportation, environmental and health economics. However, ANA has practical relevance and importance in marketing, given the large variation in characteristics for the choice situations consumers face.

2.2 Eye tracking, Decision-making and Choice

Eye tracking has a long history in psychology and marketing research and has been used in diverse settings to understand attentional processes, search behavior and choice (Wedel and Pieters 2008). As eye movements are considered to be representative of covert attention and cognitive processes (e. g., Wedel and Pieters 2008), it has been paramount in studying consumer decision-making (see Orquin and Loose 2013 for a comprehensive review).

For example, Shi, Wedel, and Pieters (2013) study the information acquisition of consumers on comparison websites. Notably, they find that not all alternatives and attributes receive attention or are discarded at the decision stage. Meißner et al. (2016) explore attentional processes in CBC. They show that repeated choices reinforce the ease of finding relevant information and that through the sequence of choice tasks, respondents become more selective and faster at acquiring that information. Orquin, Chrobot, and Grunert (2018) further demonstrate that predictability of the location of the information, which is the case in the CBC, increases (decreases) the likelihood of looking at information of high (low) relevance. That is, while eye movements are generally a result of both bottom-up (e. g., size of the stimuli), and top-down (e. g., consumer goals) factors, due to the learning that occurs in repeated choices, the latter seems to prevail (Orquin, Bagger, and Loose 2013).

Other studies use eye tracking to relate attention to preferences (e. g., Toubia, De Jong, Stieger, and Füller 2012), consideration set formation (e. g., Chandon, Hutchinson, Bradlow, and Young 2009), as well as choice (e. g., Pieters and Warlop 1999). Furthermore, it has been essential in studying and modeling consumer search behavior (e. g., Van der Lans, Pieters, and Wedel 2008; Reutskaja, Nagel, Camerer, and Rangel 2011; Liechty, Pieters, and Wedel 2003). Most notably, several studies have proposed joint models of information search and choice. For example, Stüttgen, Boatwright, and Monroe (2012) jointly model search with a satisficing choice rule, i. e., where consumers stop the evaluation process as soon as they find a satisfactory product. Yang, Toubia, and De Jong (2015) propose a dynamic search model, where information acquisition process represents a cognitive cost that needs to be compensated.

In contrast, we do not model information search. Instead, we are interested in the link between visual attention and the underlying attribute processing strategy and treat it as an exogenous indicator of the relevance of the attribute information (Meißner and Oll 2017). From this perspective, the studies of Balcombe et al. (2015), Meißner et al. (2011), Krucien, Ryan, and Hermens (2017), and Van Loo, Nayga, Campbell, Seo, and Verbeke (2018) use eye tracking in the ANA context. However, they adopt an exogenous approach, which suffers from the limitations outlined in Section 2.1. Furthermore, e. g., Krucien et al. (2017) allow a continuous measure of visual attention to update the preference parameters. In contrast, we utilize an endogenous approach that allows us to link eye fixations to the underlying ANA strategies in a probabilistic manner, avoiding any explicit assumptions about a causal effect of eye movements on preferences as outlined by Orquin and Loose (2013).

3 Methodology

We start the following section by describing our main model – the mixed endogenous attribute attendance (MEAA) model – which explicitly allows us to accommodate both ANA and preference heterogeneity as well as to connect visual attention from eye tracking to the consumers' applied attribute processing strategy. After that, we will discuss how we derive the measure of visual attention to be used in the MEAA model, and last, we explain the calculation of the measures (e. g., the relative importance of attributes, WTP) and quantities (e. g., posterior probabilities) that are obtained as a transformation of the parameter estimates and used to generate insights in the empirical application.

3.1 Mixed Endogenous Attribute Attendance Model

The MEAA model (Hole et al. 2013) is a confirmatory latent class approach (Hess et al. 2013) that relaxes the assumption of full information processing. In particular, individuals can ignore any number and combination of attributes. Given K attributes, there exist 2^K possible attendance/non-attendance combinations or attribute processing strategies (e. g., Hess et al. 2013). In the MEAA model, for each of the possible attribute processing strategies, we have a corresponding latent class s ($s = 1, \ldots, S$) that can be described by a K-dimensional column vector $\lambda_s = [\lambda_{s1}, \ldots, \lambda_{sK}]'$ of zeros and ones, indicating the attributes that are ($\lambda_{sk} = 1$) and are not included ($\lambda_{sk} = 0$) in the specific class s.

In this model, the utility individual i ($i = 1, \ldots, I$) obtains from alternative j ($j = 1, \ldots, J$) in choice task t ($t = 1, \ldots, T$) is class-specific:

$$U_{ijt|s} = x_{ijt} \cdot \beta_{is} + \epsilon_{ijt}, \tag{1}$$

where x_{ijt} is a K-dimensional row vector of attribute values describing alternative j in choice task t for individual i, β_{is} is a column vector of corresponding preference parameters, and ϵ_{ijt} is an identically distributed type I extreme value error term. The subscripts i and s indicate that the vector of preference parameters is individual- and class-specific. The former allows incorporating preference heterogeneity, assuming the individual parameters β_i are distributed multivariate normal: $\beta_i \sim N(\beta, \Sigma)$. The class-specific parameters are obtained via the elementwise multiplication of λ_s with the individual-specific vector of parameters: $\beta_{is} = \lambda_s \circ \beta_i$. For the attributes not included in class s, the corresponding elements in λ_s set the preference parameters to zero. We use effects coding for all categorical attributes (e. g., brand) and linear specification for price-related attributes.[4] If multiple elements in x_{ijt} are related to an attribute ("attribute-levels"), we map the λ_s vector onto the correct parameter dimension.

[4]Note that dummy coding is not an option here because then the reference level of an attribute has a utility of zero and cannot be differentiated from ANA (Gilbride et al. 2006).

To obtain a better idea of different utility functions in each of the classes, we provide a simple example, where products are described by only three attributes. This results in $S = 2^3 = 8$ possible classes. As the parameter estimates are switched on and off, each class is characterized by a different linear (additive) utility function, which we present in Table 1. As a result, several decision rules are incorporated: (class 1) full compensatory (i. e., full attendance), (class 8) random choice, (classes 5-7) (a probabilistic version of) lexicographic, and (classes 2-4) semicompensatory, i. e., the compensatory rule applies only within the particular subset of attributes.

[Insert Table 1]

Even though, we allow for ANA, we assume that individuals are utility maximizers. In doing so, we follow the bounded rationality literature, which states that individuals can still act rationally by maximizing their utility but do so based on partial and imperfect information (Rasouli and Timmermans 2015). Given the distribution of the error term, within each class s, the probability of individual i choosing alternative j in choice task t is:

$$P_{ijt|s} = \frac{\exp(x_{ijt} \cdot \beta_{is})}{\sum_{j' \in J} \exp(x_{ij't} \cdot \beta_{is})}. \tag{2}$$

Following Hole (2011), we assume the likelihood of attending a particular attribute is independent of attending other attributes (independence of attribute attendance (IAA)). Thus, the class probabilities τ_{is} (where $0 \leq \tau_{is} \leq 1$ and $\sum_{s=1}^{S} \tau_{is} = 1$) can be modeled as a mapping from the attribute attendance probabilities, π_{ik}, which are parametrized as a logistic function:

$$\tau_{is} = \prod_{k=1}^{K} \pi_{ik}^{\lambda_{sk}} \cdot (1 - \pi_{ik})^{1-\lambda_{sk}} \tag{3}$$

$$\pi_{ik} = \frac{\exp(z_{ik} \cdot \gamma)}{1 + \exp(z_{ik} \cdot \gamma)}, \tag{4}$$

where z_{ik} is a $K+E$-dimensional row vector of K attribute-specific intercepts and (possibly)

E attribute-specific individual-level variables (e. g., revealed or stated ANA) with corresponding parameter vector γ. Note that the additional variables entering z_{ik} are optional, and the MEAA model can be estimated without any extra information. In this case, the attribute attendance probabilities π_{ik} and respective class probabilities τ_{is} are common across individuals, and the subscript i can be dropped. The submodel in Equation (3) is closely related to the model proposed by Swait and Ben-Akiva (1987) for modeling choice set heterogeneity and to the concomitant latent class models of Kamakura, Wedel, and Agrawal (1994). Although the IAA assumption seems restrictive, it ensures parsimony and the practicality of the model because the number of parameter rises linearly with K, not S, as would be the case if we used a multinomial logit model for τ_{is} with $S-1$ class-specific intercepts (Hole 2011; Hole et al. 2013). In addition to the loss of parsimony, relaxing the IAA assumption may reduce model stability (i. e., potential issues with local maxima) while offering only marginal improvements in fit (Hess et al. 2013).

The unconditional probability of individual i choosing alternative j in choice task t can be derived by combining Equations (2) and (3). τ_{is} can be interpreted as the size of class s and is the prior probability of finding individual i in class s.

$$P_{ijt} = \sum_{s=1}^{S} \tau_{is} \cdot P_{ijt|s}. \tag{5}$$

The utility function in Equation (1) allows a straightforward derivation of restricted models, which do not include either ANA, preference heterogeneity or both. Figure 1 summarizes the relationships between different model versions.

[Insert Figure 1]

For example, by setting $\tau_1 = 1$ (i. e., everyone belongs to class 1 with full attendance), the MEAA model is reduced to the mixed multinomial logit (MMNL) model. By setting $\Sigma = \mathbf{0}$ while retaining all S classes, the MEAA reduces to the endogenous attribute attendance (EAA) model proposed by Hole (2011). Combining both restrictions leads to the multinomial logit (MNL) model. Hence, MNL, EAA, and MMNL are all special cases of the MEAA model and will serve as benchmarks in the empirical application.

3.2 Estimation Procedure

For statistical inference, we employ maximum likelihood estimation with sample log-likelihood $LL(\theta) = \sum_{i \in I} \ln(L_i)$ where L_i is the likelihood of individual i, and θ denotes the vector of unknown parameters $\theta = [\beta, \Sigma, \gamma]'$. For the MEAA model, L_i is the weighted sum of the respective class-specific likelihoods, i. e., $L_i = \sum_{s=1}^{S} \tau_{is} \cdot L_{i|s}$. The latter represents the sequence of observed choices for individual i conditional on class s because the data have a panel structure (i. e., t choice tasks for each individual i), and we assume that individuals do not change the attribute processing strategy across tasks:

$$L_{i|s} = \int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|s}^{y_{ijt}} \phi(\beta_i | \beta, \Sigma) d\beta_i, \tag{6}$$

where y_{ijt} is a dummy indicating whether alternative j was chosen by individual i in choice task t, and ϕ is the density of the normal distribution. For the MMNL model, no weighting by class probabilities is necessary. For the EAA model, preference parameters are homogenous, and therefore no integration over the parameter distribution is required. For MNL, neither integration over the parameter distribution nor weighting by class probabilities is required.

As the integral over the density of β_i in Equation (6) has no closed-form solution, we adopt the simulated maximum likelihood approach and approximate it using 500 Halton draws (Train 2009). We estimate all parameters simultaneously using the gradient-based Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (see Train 2009, p. 225). Because latent class models may have multiple local optima (see Wedel and Kamakura 2000), multiple starting values were tested to find the global optimum (Dayton and Macready 1988).

3.3 Postestimation Measures

Based on the estimated parameters, we derive several measures and quantities that are central for our empirical analysis.

In the (M)EAA models, we can simultaneously segment respondents into specific classes characterizing particular attribute processing strategies. We therefore use the estimated parameters $\hat{\theta}$ in Equations (3) and (6) to obtain posterior class probabilities via Bayes' rule (Wedel and Kamakura 2000):

$$\hat{\tau}_{is}^{\text{post.}} = \frac{\hat{\tau}_{is} \cdot L_{i|s}}{\sum_{s=1}^{S} \hat{\tau}_{is} \cdot L_{i|s}}, \tag{7}$$

Here, the prior class probabilities are reweighted by the estimated likelihood of each individual i conditional on class s. The resulting posterior class probabilities represent a "fuzzy" segmentation criterion, and individuals can be fractional members of (multiple) attribute processing strategy segments. Given our assumption that each individual has a specific attribute processing strategy, we opt for a nonoverlapping segmentation and assign individuals to the class where the value of $\hat{\tau}_{is}^{\text{post.}}$ is the highest (cf. Desarbo, Ramaswamy, and Cohen 1995). To assess the degree of overlap, we use an entropy-based measure (e. g., Wedel and Kamakura 2000):

$$\text{entropy} = 1 + \frac{\sum_{i=1}^{I} \sum_{s=1}^{S} \hat{\tau}_{is}^{\text{post.}} \cdot \ln\left(\hat{\tau}_{is}^{\text{post.}}\right)}{I \cdot \ln(S)}. \tag{8}$$

This measure is bound between 0 and 1, with values of zero indicating complete overlap between class allocations (i. e., all posterior class probabilities are equal) and values close to one implying a more certain class assignment with minimal overlap.

Furthermore, we derive some key measures that represent a transformation of the estimated preference parameters such as the relative importance of attributes and WTP, which have practical significance for marketing managers (Rao 2014). We base these measures on the conditional individual-level estimates $\hat{\beta}_{is}$. That is, we utilize all the available information (e. g., observed choices and other individual-level information) in a submodel of class probabilities in Equation (3)) to increase the accuracy of the preference estimate for a given individual (e. g., Hensher, Rose, and Greene 2015).

After obtaining the class for each individual with the highest value of $\hat{\tau}_{is}^{\text{post.}}$), denoted by $\hat{s}$, we employ Bayes' rule again to condition on the observed choices and compute the posterior means of the parameters on the individual level (Train 2009):

$$\hat{\beta}_{i\hat{s}}^{\text{post.}} = \frac{\int \hat{\beta}_{i\hat{s}} \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|\hat{s}}^{y_{ijt}} \phi(\beta_i|\beta, \Sigma) d\beta_i}{\int \prod_{t=1}^{T} \prod_{j=1}^{J} P_{ijt|\hat{s}}^{y_{ijt}} \phi(\beta_i|\beta, \Sigma) d\beta_i}, \tag{9}$$

where we again employ a simulation method with Halton draws to approximate the integrals. We use this definition of individual estimates because it preserves the value zero on the individual level if i is classified as a non-attender. This feature is of central importance in marketing applications because it also translates into the derived measures, such as relative importance and WTP (as described next), and thus helps the analyst target on the individual level. Note that EAA model preference parameters are the same for all individuals. Therefore, the subscript i in $\hat{\beta}_i$ in Equation (9) can be dropped. For the MMNL model, we do not have to condition on a specific class.

The relative importance of an attribute is the ratio of its utility range to the sum of the utility ranges of all attributes (Rao 2014). We compute the utility ranges on the individual level from $\hat{\beta}_{i\hat{s}}^{\text{post.}}$. Furthermore, we use two aggregate measures for the (M)EAA model (similar to Gilbride et al. 2006). First, we calculate the mean across all individuals, which also includes individuals who did not attend to attribute k and have a corresponding relative importance of zero. Second, we are interested in assessing this measure for individuals who, in fact, attended to the specific attribute. Therefore, we further compute the average relative importance for this subset of individuals. For targeting specific segments, this measure of relative importance entails essential and relevant information for managers.

We compute individual WTP values from $\hat{\beta}_{i\hat{s}}^{\text{post.}}$ by dividing the parameters (respectively, the differences in attribute-level related parameters) of nonprice related attributes by the negative price parameter. Hence, conveniently, we rescale the utility of each attribute in monetary units based on the marginal rate of substitution (Rao 2014). Following a similar logic as that used for the relative importance computation, we derive an average WTP for an attribute across all and one across the subset of individuals who attended to the attribute.

3.4 Measure of Visual Attention

Subsequently, we use a continuous measure of visual attention derived from eye tracking as an additional exogenous individual-level information, which enters z_{ik} in Equation (4) in the (M)EAA model specification. This is similar to the inclusion of consumer descriptors (concomitant variables) in a latent class model for segmentation purposes (e. g., Kamakura et al. 1994; Wedel and Kamakura 2000). For example, Gupta and Chintagunta (1994) use demographics in the context of brand choice using scanner panel data. In a similar manner, Swait and Adamowicz (2001) employ complexity measures when modeling choices from a stated choice experiment. Generally, the supplementary information used as a covariate in the submodel of class probabilities proves to increase model fit and aid with the identification of the latent classes (Dayton and Macready 1988).

We use eye fixations as an input for our metric. This choice from among the possible eye tracking metrics is motivated by the fact that the number of fixations is one of the most commonly used proxies indicating information acquisition and attention (Wedel and Pieters 2008; Holmqvist et al. 2011) and has been previously used in the context of ANA (e. g., Balcombe et al. 2015). Furthermore, in line with previous literature, we also find it to be highly correlated with fixation duration. In CBC, the information in each choice task is presented in a matrix form, where attributes are typically presented as rows and alternatives as columns. Given K attributes and J alternatives, this results in a $K \times J$ matrix, with each (k, j) element characterizing attribute k for alternative j. We define each of the (k, j) elements as a separate area of interest. In a given task t for each individual i, we count the number of fixations on each area of interest, i. e., the (k, j) element. We further aggregate the fixation counts on each area of interest across all J alternatives and T choice tasks. Next, we standardize this measure within each individual and label it as va_{ik}. This enables us to control for any potential heterogeneity in individual processing capacities. The motivation for standardization here is similar to that in Pieters and Warlop (1999), where they mean center the visual attention measure for each individual to control for differences in experimental conditions.

Using va_{ik} as additional information in z_{ik} should help to model π_{ik} and, therefore, τ_{is}. In particular, we expect a positive effect of va_{ik} on π_{ik}. Nevertheless, the probabilistic relationship between visual attention and the particular attribute processing strategy, in contrast to exogenous approaches to accommodating ANA, allows for the possibility that looking at given information does not guarantee that it is deployed in decision-making.

4 Empirical Application

We start the following section by describing the two datasets we chose for our empirical application. We then continue with a detailed discussion of the estimation results, e. g., model fit, parameter estimates, including the effect of visual attention, as well as the differences in subsequent individual class memberships, the relative importance of attributes and WTP.

4.1 Data

We employ two studies involving choices in different durable product categories: coffee makers and laptops, conducted by Meißner et al. (2016) and Yang et al. (2015), respectively. Both studies combine a CBC study with an eye tracking experiment, i. e., eye movements of the respondents were simultaneously tracked while they completed the choice tasks (for details on the eye tracking devices used and the experimental setup, we kindly refer the reader to the respective articles). We chose these two datasets because they represent typical CBC studies used in marketing research. Therefore, the results of our analysis are relevant to a broader marketing audience. However, they differ in terms of the product category and some features of the experimental setup and design presented in Table 2, which enables us to validate that the general patterns we find also hold across contexts.

[Insert Table 2]

In both cases, the respondents were sampled from students of European universities. After excluding responses with incomplete data or a straight-lining pattern, we obtain samples of 59 (coffee makers) and 70 (laptops) respondents, which is a typical sample size for eye tracking experiments. The studies vary in the number of choice tasks: 12 in coffee makers and 20 in laptops, i. e., fewer observations per respondent are available for the coffee makers study. Moreover, a "none" option was included in the coffee makers study, with an average choice share of 15.4%. The rest of the choice shares were equally distributed among the three alternatives. In the laptop study, the average distribution of the choice shares was slightly less balanced, ranging from 21–30%. Additionally, the laptop study was incentive-aligned, but without a "none" option, which makes it inappropriate for WTP calculation (Allenby, Brazell, Howell, and Rossi 2014). Another relevant difference for our application is the variation in the characteristics of the attributes included in the studies. Notably, some attributes in the coffee makers study contain pictorial information (e. g., design and system), while in the laptop study, all attributes are nonpictorial. We would expect the pictorial attributes to require more visual attention to be processed and incorporated into the decision-making. Furthermore, both studies include six attributes, which results in 64 possible attribute processing strategies. Therefore, we can investigate how well the models can identify the particular strategy applied by an individual given these many possibilities.

Regarding the eye tracking information, we observe that all respondents do fixate on all attributes across choice tasks. The standardized number of fixations and therefore our measure of visual attention varies substantially across attributes (mean ranging from -0.70

to 0.53 for coffee makers and from -0.95 to 1.13 for laptops), across individuals (standard deviation ranging from 0.75 to 0.89 for coffee makers and 0.24 to 0.79 for laptops), as well as within individuals (average range across individuals of 2.59 for coffee makers and 2.51 for laptops). This variation, therefore, allows identification of the parameter estimates, and hence we conclude that the datasets are well suited for our analysis.

4.2 Model Comparison

For each of the datasets, we have estimated six models, including the MNL, MMNL, EAA, EAA(va), MEAA, and MEAA(va), where "va" indicates that the models include the visual attention measure. In the initial solutions, the MEAA and MEAA(va) models had very large and positive intercept estimates in the submodel of the class probabilities in Equation (3) for the attributes *price* (coffee makers) and *support* (laptops). Note that this is not an issue and only shows that there is no ANA for these attributes after controlling for preference heterogeneity (Hole et al. 2013). Hence, we simply reestimated the models setting these attributes to 100% attendance. Consequentially, for both datasets, $5^2 = 32$ possible attribute processing strategies (or classes) are available. Please note that we also included a dummy variable for the "none" option in the coffee makers study, but we do not allow for ANA here because "none" is not a product attribute. Additionally, we estimated the heterogeneous models with a diagonal specification of Σ for reasons of parsimony. We report the final estimation results in Tables 3 and 4.

[Insert Table 3 and 4]

In general, we are interested in several comparisons. First, by contrasting homogeneous vs. heterogeneous models, we aim to validate the importance of accounting for preference heterogeneity, primarily due to potential confounding with ANA. Second, by comparing the models that assume full attendance vs. those including ANA, we outline the implications of neglecting the latter in a marketing context. Third, the prime focus is the comparison of models that include the measure of visual attention vs. "regular" ANA models (i. e., EAA and MEAA).

For the comparison of in-sample fit across models, we use log-likelihood (LL), Bayesian information criterion (BIC), and McFadden's R^2 (ρ^2) (see Wedel and Kamakura 2000). The BIC has the advantage over LL that it penalizes model complexity and we can further use it for the comparison of non-nested models.[5] ρ^2 offers an intuitive interpretation, with values between 0.2 and 0.4 indicating a very good fit (Louviere, Hensher, and Swait 2000).

First, the models fit the data in both applications quite well. Furthermore, as suggested by the smaller BIC, larger ρ^2, and LL-values, all heterogeneous models outperform their homogeneous counterparts in both applications. For example, the MEAA outperforms the EAA and the MEAA(va) outperforms the EAA(va). Moreover, the MMNL outperforms all homogenous models, including the best fitting EAA(va). Thus, relaxing the assumption of homogeneity in preferences is crucial, including for models that accommodate ANA. In addition, in both applications, there is more to gain by accounting only for preference heterogeneity vs. only for ANA. Second, the MEAA and MEAA(va) outperform the MMNL, while the EAA and EAA(va) outperform the MNL model, i. e., in general, accounting for ANA leads to considerable improvement in model fit across both applications. Finally, the MEAA(va) is, overall, the best fitting model and outperforms the MEAA in both studies. Similarly, the EAA(va) outperforms the EAA model. Hence, the visual attention measure is a useful indicator of ANA.[6]

To assess the predictive validity of the models, we additionally report hit rate and hit probability as common measures of out-of-sample fit (Gilbride et al. 2006). Using the individual-level posterior parameter estimates, we computed the measures from leave-one-out cross-validation (Maldonado, Montoya, and Weber 2015). In each fold, we randomly

[5]Note that the MEAA model nests all other models at the boundary of the parameter space, hence the LL ratio test is not applicable (McLachlan and Peel 2000).

[6]As an additional validation exercise, we have used the derived measure of visual attention in an exogenous approach. Using grid search, we determined the cutoff value for building the discrete indicator of ANA. The corresponding attributes are then set to zero in the MNL and MMNL models. These benchmark models (which can be estimated using standard software) outperform the MNL and MMNL and have a comparative fit to EAA and MEAA models. Results are available from the authors upon request.

left out one choice task for each respondent, repeated this procedure T times, and averaged the results. Hence, all observations are also used once in the validation, which increases the robustness of the results, however, at the cost of the need for T estimation runs. The hit rate is the average rate of correct predictions across the individuals. However, it does not convey any information on the "certainty" of the prediction. On the other hand, hit probability, as the average predicted probability of the chosen alternative across the sample, does. These measures confirm the result based on the in-sample fit, although the relative differences across models are somewhat smaller, but we also do not detect overfitting. In general, all the heterogeneous models fit very well both in-sample (ρ^2 values of over 0.39) and out-of-sample (hit rate of more than 0.64 and hit probability of more than 0.59 among four alternatives). Therefore, getting any additional improvement out-of-sample is challenging. We would, hence, consider the MEAA(va) model to be more reliable. As we will show, they generate somewhat different insights, particularly on the individual-level.

4.3 Parameter Estimates

The resulting parameter estimates are presented in the upper panels in Tables 3 and 4. In general, the estimates across all models in both studies have face validity (e. g., negative price parameters) and reasonable magnitudes. However, we do observe relevant differences in utility parameters, including partworth and price estimates, across models. In particular, we see both increases and decreases in the magnitudes when moving from the worse fitting (M)MNL models to the better fitting (M)EAA and (M)EAA(va) models. For example, in the MEAA models, the mean price parameter increases in magnitude compared to the MMNL model. At the same time, the standard deviation decreases. As we find some level of non-attendance to *price* in the laptop study, it is expected that some of the heterogeneity recovered in the MMNL is now captured by the non-attendance class, shifting the mean away from zero and implying less continuous heterogeneity. In the case of the coffee makers study, the shift in the price estimate initially seems counterintuitive considering the full attendance to this attribute. However, as Hess et al. (2013) state, such changes may also depend on the specification of other attributes. Along with the potential scale differences, the latter complicates the direct comparison of the utility estimates. Assuming that the true model includes preference heterogeneity and ANA, the results show that neglecting the latter leads to biased estimates.

Turning to the class parameters in the (M)EAA models, we see large differences in the intercepts across attributes in both datasets. This already indicates differences in the attribute attendance probabilities, and interestingly, the differences in intercepts persist in models including the visual attention measure. Regarding the latter, we observe a positive and significant effect in both applications, i. e., a higher level of visual attention generally results in a higher likelihood of attending an attribute. Notably, the magnitude of the effect increases in the MEAA(va) compared to the EAA(va) model. One potential explanation for this finding is the confounding of preference heterogeneity and ANA. As visual attention should be indicative of non-attendance rather than low sensitivity, by better isolating these two in the MEAA(va) model, the relationship between visual attention and attribute attendance becomes more pronounced. We find further support of the confounding effect when examining the average attribute attendance probabilities presented in Figure 2. In particular, in both studies for almost all attributes, attendance probabilities are higher in the MEAA models, becoming 100% for *price* in coffee makers and for *support* in laptops.

[Insert Figure 2]

The higher attribute attendance probabilities also translate into a higher probability of attending more attributes, as evident from the shift of the probability distribution of the number of attended attributes to the right for the MEAA models (see Figure 3). In studies of choices of prescription drugs and commuting routes, respectively, Hole et al. (2013) and Hess et al. (2013) also find that many attributes become 100% attended after accounting for heterogeneity. However, we still see a considerable amount of non-attendance for most of the attributes in our two applications. For example, except *price* and *price per cup*, the attendance probabilities for all other attributes in coffee makers remains below 50%, resulting in the majority of respondents attending to three out of six attributes in the MEAA models. One potential explanation is the differences in the level of involvement and the associated risk of the decision in the various contexts. We find further supporting evidence by noting that the levels of non-attendance are lower in the laptop study, potentially due to incentive alignment and a higher risk related to financial cost. Additionally, we observe a larger shift of the probability distribution for the number of attended attributes to the right in the MEAA models, with the majority attending to four out of six attributes. It is, however, noteworthy that despite incentive alignment, some respondents did not attend to *price*.

[Insert Figure 3]

Returning to the comparison of the MEAA(va) and MEAA models, some differences are visible in attendance probabilities for *price per cup* in the coffee makers study, where the MEAA(va) model retrieves 12.3 percentage points lower attendance probability. Likewise, for laptops, we find a 7.5 to 15.6 percentage points lower attendance probabilities for *capacity*, *size*, and *price*. As a result, the share of the respondents attending to three attributes for coffee makers and four attributes for laptops becomes larger, mainly on account of the smaller share of those attending to five or six attributes. All in all, we find ample evidence for ANA in both product categories, frequent use of semicompensatory (more than 98% in both studies), rare use of lexicographic (less than 2% in coffee makers and 0% in laptops) and full compensatory strategy (less than 1% in both studies), and no use of random choice. Considering the better fit and potential issues with confounding, from this point on, we focus on the heterogeneous models.

4.4 Visual Attention and Attribute Non-attendance

The relationship between visual attention and attribute (non)attendance merits a more detailed discussion. All in all, paying more attention to an attribute increases the likelihood of, but does not guarantee, using it when making choices. Moreover, due to variation in attribute intercepts, the same amount of visual attention results in different attendance probabilities across attributes, as illustrated in Figure 4.

[Insert Figure 4]

Here, we have calculated the attendance probabilities for a range of values of visual attention (observed in the datasets) using the estimated $\hat{\gamma}$ parameters. The negative (positive) values of attribute intercepts shift this sigmoid relationship to the right (left), such that for a given amount of visual attention, a higher attendance probability is implied for *brand*, followed by *material*, *price per cup*, *system* and *design* in coffee makers (left panel). The dots in Figure 4 are the average values of visual attention for each attribute and relate to the average attendance probabilities presented in Figure 2. The model is, therefore, capturing the heterogeneity in required processing capacity across attributes. It, in fact, distinguishes some patterns in visual attention related to the characteristics of the attribute. In particular, a distinction can be observed between pictorial (e. g., *design* and *support*) and nonpictorial attributes, with the former seemingly requiring higher levels of attention for processing and incorporating into decision-making. In the case of the laptop study (right panel), we also see differences in inferred attendance probabilities across attributes. However, the attribute-specific curves seem to be closer to each other than in the coffee makers study. One potential explanation is that the laptop study does not contain any pictorial information. However, the same level of visual attention results in a higher and similar probability of attending *price* and *capacity*, followed by *antivirus*, and much lower probability for attending *speed* and *size*.

4.5 Differences in Class Allocation

We now turn to the comparison of the MEAA models to investigate how much the visual attention measure helps with allocating people into classes. To this end, we compare the entropy measure calculated in Equation (8). We obtain values of 0.66 in the MEAA and 0.75 in the MEAA(va) for coffee makers and 0.76 in the MEAA and 0.84 in the MEAA(va) for laptops. Considering 32 classes in both applications, already the class allocation in the MEAA model appears to be quite good, and even more so in the MEAA(va) model. To illustrate some general patterns and critical distinctions between the models, we report an example of class allocation for two respondents in the coffee makers study in Figure 5. However, these are representative for most of the respondents in the analyses (64% and 70% for coffee makers and laptops, respectively), where the MEAA(va) compared to the MEAA model has a higher posterior probability for the identified class.

[Insert Figure 5]

More specifically, id = 14 (top panel) represents the case (34% and 26% for coffee makers and laptops, respectively) where the MEAA(va) and the MEAA model indicate the same class, but the former results in a higher posterior probability. By contrast, id = 28 (bottom panel) illustrates the case (30% and 44% for coffee makers and laptops, respectively), where the class allocation is different, with the MEAA(va) having a higher posterior probability for the identified class. Note that the classes are distinct in terms of the implied attribute processing strategy. Therefore, differences in the class allocation will have consequences for other individual-level measures. For instance, the MEAA(va) model suggests that id = 28, in addition to *price* (which is always attended), most likely attends to *design*. In contrast, the MEAA model suggests that this respondent only attends to *price* and potentially to *price per cup*. Therefore, the vector of parameters for id = 28 would be considerably different between the models, which will have further repercussions for the WTP. Given that the models suggest different class allocations (irrespective of which model leads to a higher posterior class probability) for a considerable proportion of the samples (47% and 67% for coffee makers and laptops, respectively), one can expect substantial discrepancies between the derived individual-level insights. Hence, the additional use of visual attention is not only important to improve (in-sample) fit on the aggregate level but also to obtain substantive results on the individual level, which in turn are relevant for target marketing.

4.6 Relative Importance of Attributes and Willingness-to-pay

The aggregate values of the relative importance of attributes for both studies are summarized in Table 5. At first glance, the MMNL results seem consistent (e. g., same ranking across attributes) with the MEAA models' implied weighted average values in columns 3 and 4. However, we still observe some meaningful differences in both studies. For example, for coffee makers, the importance of *design* drops approximately 4 and *price* increases approximately 6 percentage point in the model accounting for ANA. Similarly, for laptops, the MMNL model understates (overstates) the relative importance of *speed* (*antivirus*).

[Insert Table 5]

Nevertheless, firms might be more interested in understanding the importance of attributes for those individuals who, in fact, use them when making choices. These are presented in columns 5 and 6 in Table 5. Note that the sum of the relative importance across attributes is no longer 100% as they are based on different subsets of the sample. Here, we see substantial increases in the relative importance across all attributes except price. The magnitude of the difference is dependent on the amount of ANA for a given attribute. For example, the importance of *design* rises from 7.8% in the MMNL to approximately 32% in the MEAA models. Contrasting with the results of Gilbride et al. (2006), we find larger differences in the relative importance measure, which demonstrates that these are also context-specific.

Furthermore, the difference between the MEAA and MEAA(va) models appears to be less pronounced. Nevertheless, we find some variation that is meaningful for practitioners. In particular, even on average over the whole sample (column 3 and 4), the values differ by approximately 2 to 3 percentage points for *price per cup* for coffee makers as well as for *speed* and *price* for laptops. For the subsets of attenders (column 5 and 6), we find approximately 2 and 5 percentage point differences for *brand* and *price per cup* in the coffee makers study, respectively. Even larger differences are suggested for laptops, with approximately 7 and up to 12 percentage points for *antivirus* and *price*, accordingly.

We also find differences in implied WTP across the three models presented in Table 6.[7] Here, as well, for the MEAA models, we report the weighted average WTP across all and only the attending individuals.

[Insert Table 6]

For the majority of the attribute-level comparisons, the MMNL model seems to overestimate the average WTP (columns 2, 3 and 4). While in some cases the differences appear to be small (e. g., for Krups vs. Severin 22.10€ in MMNL compared to 18.98€ in MEAA and 20.45€ in MEAA(va)), in other cases they are considerable (e. g., for Philips vs. Severin 43.50€ in MMNL vs. 23€ in MEAA and 22.73€ in MEAA(va)). This finding is in line with the studies of Hole et al. (2013) and Hess et al. (2013). In particular, as 38.4% of the sample in the MEAA and 40.2% in the MEAA(va) ignore the brand attribute, they, subsequently have a zero WTP, which decreases the average value over the sample. By contrast, when we consider only the subsets of attenders to the specific attribute (columns 5 and 6), the MMNL model understates the WTP across almost all attribute-level comparisons by more than 2 times. Interestingly, the MEAA(va) shows here (in absolute terms) slightly lower WTP values (except for stainless steel vs. aluminum) compared to the MEAA.

To obtain a better understanding of the individual-level differences, we present the cumulative distribution of individual WTP values (see also Hensher, Collins, and Greene 2013) for selected attribute-level comparisons in Figure 6. The Krups vs. Severin comparison (upper panel) is representative of 6 out of 9 attribute-level comparisons, where the WTP stays (mostly) in the positive domain (i. e., for all individuals with nonzero WTP, Krups is preferred over Severin). By contrast, plastic vs. aluminum (lower panel) represents the other three cases, where the WTP in the MMNL model spreads across both positive and negative domains (i. e., some individuals prefer plastic over aluminum, while others prefer aluminum over plastic as a material for coffee makers).

[7] As the laptop study did not include a “none” option, we do not present the WTP calculations, as WTP values are not necessarily meaningful (Allenby et al. 2014). Nevertheless, the general patterns observed for coffee makers are also present in the laptop study. The results are available upon request.

[Insert Figure 6]

In line with the previous literature (e. g., Hess et al. 2013), the MEAA models in all attribute-level comparisons suggest a lower level of heterogeneity (i. e., the variance in the WTP distribution). For both datasets, the recovered heterogeneity in WTP is overstated in the MMNL model, and it seems to be driven mainly by extremes, for which we obtain (in absolute terms) unrealistically high WTP values (e. g., $|\text{WTP}| > 100$€ for plastic vs. aluminum). At the same time, the WTP for the rest shrinks towards zero (e. g., |WTP| of only approximately 10 to 20€ for a large fraction of the sample). In contrast, due to a high level of non-attendance to *brand* and *material*, many individuals in the MEAA models have WTP values of exactly zero, and therefore we obtain a lower mean value over the sample. However, for the rest of the subset, the WTP values are in many cases much higher than MMNL predicts. Hence, consistent with Gilbride et al. (2006), we also find supporting evidence that accounting for ANA is crucial for accurate identification of subsets of individuals with high (but realistic) WTP, i. e., the extremes of the preference distribution, and for the proper targeting of these segments following the suggestions of Allenby and Ginter (1995). Comparing the WTP distributions of the MEAA and MEAA(va) models, we see that the main difference in WTP stems from different subsets of individuals with a (non)zero WTP for a given attribute, as already discussed in Section 4.5. Note that, e. g., the MEAA(va) identifies more respondents with WTP values of approximately 20€ to 50€ for Krups vs. Severin. Given that the MEAA(va) provides the better class allocation and better model fit, we interpret the resulting differences in WTP as important. Nevertheless, the WTP distributions of the models with ANA are in general similar, and larger differences arise in comparisons using the MMNL model.

In sum, we conclude that it is crucial to control for heterogeneity and ANA in discrete choice models to obtain relative importance and WTP values that are realistic as well as meaningful and, at the same time, insightful for targeting in marketing applications.

5 Conclusion

In this paper, we show the prevalence of ANA in a marketing context, specifically in two applications where individuals choose durable products, such as coffee makers and laptops. Although one would expect that people are more careful when making product choices in durable categories due to higher stakes, we find that across the two applications, even after controlling for preference heterogeneity, the majority attends to only three to four (different) attributes out of the available six. Simultaneously, almost no one considers all, none, or only one attribute. In such cases, assuming full attendance can be misguided and lead to considerable biases in the derived implications.

Furthermore, we provide empirical evidence of the positive and significant effect of visual attention on the probability of attending a particular attribute. Our proposed model further captures variation in the required visual attention for processing and incorporating different attributes in decision-making. More specifically, our results may suggest that the same level of attention to nonpictorial vs. pictorial information results in a higher likelihood of attendance. The proposed model, in general, provides a framework for testing the efficiency of, e. g., framing effects in CBC. For example, Jonker, Donkers, de Bekker-Grob, and Stolk (2018) observe that more people seem to attend to an attribute when it is made more salient by color coding. Our framework can enable further investigation, whether such shifts are due to a higher level of visual attention or the ease of spotting salient information and whether they vary across attributes, e. g., depending on perceived relevance. Notably, we show that the use of eye tracking is informative in uncovering individual-level behavior. In particular, it helps to more clearly classify individuals into segments related to different attribute processing strategies. This implies less uncertainty in identifying the size of the segments that the firm might want to target and differences in the individual-level results (e. g., WTP). However, the model using the observed choices to infer ANA may already be sufficient for recovering the approximate patterns of attribute processing strategies as well as some key aggregate measures of interest (e. g., choice share predictions, relative importance and distributional characteristics of WTP).

Building upon our findings, several implications are noteworthy for marketing practitioners. First, we have demonstrated that ANA is plausible and applied by consumers in choice situations and there is much to gain from employing appropriate tools to support segmentation, targeting, and pricing decisions. Second, considering the general trend of decreasing prices for eye tracking (Wedel 2018), our proposed model, retaining a rather simple way of using this information, can be valuable for practitioners that want to engage in one-on-one targeting.

We see several limitations and potential extensions of the employed model. First, we have used the eye tracking information as a proxy for visual attention. However, it can instead be modeled as an outcome of an underlying latent process to avoid potential measurement error (which we would expect only to strengthen the effect of visual attention). Second, the model can be extended by relaxing the assumption of stability in applied attribute processing strategies across choice tasks. Several questions merit further investigation: whether such switching occurs and to what extent, whether the potential biases are substantial or assuming the stability of ANA strategies is acceptable. Last, the model can be easily extended to incorporate, e. g., alternative specifications of parameter distributions (e. g., lognormal vs. normal), and attribute-specific slopes for the effect of visual attention on attendance probabilities.

Additionally, contrasting respondents' self-reported (stated) ANA measures with our "revealed" measure from eye tracking might be interesting. Both work well in isolation (see Hole et al. 2013 for stated ANA and this paper for visual attention from eye tracking), but the question remains which of the measures is a more appropriate indicator of ANA. Furthermore, Balcombe et al. (2015), using a different modeling framework, suggest that these two measures might be complementary. Due to the flexibility of the MEAA model, both can be simultaneously incorporated to test related hypotheses, and we leave this for future research.

References

Alemu, M. H., Morkbak, M. R., Olsen, S. B., & Jensen, C. L. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and Resource Economics*, *54*(3), 333–359.

Allenby, G. M., Brazell, J. D., Howell, J. R., & Rossi, P. E. (2014). Economic valuation of product features. *Quantitative Marketing and Economics*, *12*(4), 421–456.

Allenby, G. M., & Ginter, J. L. (1995). Using extremes to design products and segment markets. *Journal of Marketing Research*, *32*(4), 392–403.

andreas, G., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*(5), 1055–1075.

Balcombe, K., Fraser, I., & McSorley, E. (2015). Visual attention and attribute attendance in multi-attribute choice experiments. *Journal of Applied Econometrics*, *30*(3), 447–467.

Bronnenberg, B. J., & Vanhonacker, W. R. (1996). Limited choice sets, local price response and implied measures of price competition. *Journal of Marketing Research*, *33*(2), 163–173.

Chandon, P., Hutchinson, J. W., Bradlow, E. T., & Young, S. H. (2009). Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing*, *73*(6), 1–17.

Collins, A. T., Rose, J. M., & Hensher, D. A. (2013). Specification issues in a generalised random parameters attribute nonattendance model. *Transport Research Part B: Methodological*, *56*, 234–253.

Currim, I. S., Mintz, O., & Siddarth, S. (2015). Information accessed or information available? The impact on consumer preferences inferred at a durable product e-commerce website. *Journal of Interactive Marketing*, *29*, 11–25.

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, *83*(401), 173–178.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, *47*(2), 315–72.

Desarbo, W. S., Ramaswamy, V., & Cohen, S. H. (1995). Market segmentation with choice-based conjoint analysis. *Marketing Letters*, *6*(2), 137–147.

Gilbride, T. J., Allenby, G. M., & Brazell, J. D. (2006). Models for heterogeneous variable selection. *Journal of Marketing Research*, *43*(3), 420–430.

Gupta, S., & Chintagunta, P. K. (1994). On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research*, *31*(1), 128–136.

Heidenreich, S., Watson, V., Ryan, M., & Phimister, E. (2018). Decision heuristic or preference? Attribute non-attendance in discrete choice problems. *Health Economics*, *27*(1), 157–171.

Hensher, D. A. (2006). How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics*, *21*(6), 861–878.

Hensher, D. A. (2014). Attribute processing as a behavioural strategy in choice making. In S. Hess & D. andrew (Eds.), *Handbook of choice modelling* (pp. 268–289). Cheltenham, UK: Edward Elgar Publishing.

Hensher, D. A., Collins, A. T., & Greene, W. H. (2013). Accounting for attribute non-attendance and common-metric aggregation in a probabilistic decision process mixed multinomial logit model: A warning on potential confounding. *Transportation*, *40*(5), 1003–1020.

Hensher, D. A., Rose, J., & Greene, W. H. (2005). The implications on willingness to pay of respondents ignoring specific attributes. *Transportation*, *32*(3), 203–222.

Hensher, D. A., Rose, J., & Greene, W. H. (2015). *Applied choice analysis* (2nd ed.). Cambridge, UK: Cambridge University Press.

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

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

Hess, S., Stathopoulos, A., Campbell, D., & Caussade, S. (2013). It's not that I don't care, I just don't care very much: Confounding between attribute non-attendance and taste heterogeneity. *Transportation*, *40*(3), 583–607.

Hole, A. R. (2011). A discrete choice model with endogenous attribute attendance. *Economic Letters*, *110*(3), 203–205.

Hole, A. R., Kolstad, J. R., & Gyrd-Hansen, D. (2013). Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour. *Journal of Economic Behavior & Organization*, *96*, 21–31.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures.* Oxford, UK: Oxford University Press.

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018). Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. *Value in Health*, *21*(7), 767 - 771.

Kamakura, W. A., Wedel, M., & Agrawal, J. (1994). Concomitant variable latent class models for conjoint analysis. *International Journal of Research in Marketing*, *11*(5), 451–464.

Krucien, N., Ryan, M., & Hermens, F. (2017). Visual attention in multi-attributes choices: What can eye-tracking tell us? *Journal of Economic Behavior and Organization*, *135*, 251–267.

Liechty, J., Pieters, R., & Wedel, M. (2003). Global and local covert visual attention: Evidence from a bayesian hidden markov model. *Psychometrika*, *68*(4), 519–541.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: Analysis and applications*. Cambridge, UK: Cambridge University Press.

Maldonado, S., Montoya, R., & Weber, R. (2015). Advanced conjoint analysis using feature selection via support vector machines. *European Journal of Operational Research*, *241*(2), 564–574.

McFadden, D. (2001). Economic choices. *American Economic Review*, *91*(3), 351–378.

McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Meißner, M., Musalem, A., & Huber, J. (2016). Eye tracking reveals processes that enable conjoint choices to become increasingly efficient with practice. *Journal of Marketing Research*, *53*(1), 1–17.

Meißner, M., & Oll, J. (2017). The promise of eye-tracking methodology in organizational research: A taxonomy, review and future avenues. *Organizational Research Methods*, *in press*.

Meißner, M., Scholz, S., & Decker, R. (2011). Using eyetracking and mouselab to examine how respondents process information in choice-based conjoint analysis. In *Proceedings of the Sawtooth Software conference 2010* (pp. 151–170).

Orquin, J. L., Bagger, M. P., & Loose, S. M. (2013). Learning affects top down and bottom up modulation of eye movements in decision making. *Judgment and Decision Making*, *8*(6), 700–716.

Orquin, J. L., Chrobot, N., & Grunert, K. G. (2018). Guiding decision makers' eye movements with (un)predictable object locations. *Journal of Behavioral Decision Making*, *31*(3), 341–354.

Orquin, J. L., & Loose, S. M. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206.

Payne, J. W., Bettman, J. R., Coupey, E., & Johnson, E. J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, *80*(1), 107–141.

Pieters, R., & Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing*, *16*(1), 1–16.

Rao, V. R. (2014). *Applied conjoint analysis*. Berlin, Germany: Springer.

Rasouli, S., & Timmermans, H. (2015). Models of bounded rationality under certainty. In S. Rasouli & H. Timmermans (Eds.), *Bounded rational choice behaviour: Applications in transport* (pp. 1–30). Bingley, United Kingdom: Emerald Group Publishing Limited.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *The American Economic Review*, *101*(2), 900–926.

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

Scarpa, R., Zanoli, R., Bruschi, V., & Naspetti, S. (2013). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics*, *95*(1), 165–180.

Shi, S. W., Wedel, M., & Pieters, R. (2013). Information acquisition during online decision making: A model-based exploration using eye-tracking data. *Management Science*, *59*(5), 1009–1026.

Stüttgen, P., Boatwright, P., & Monroe, R. T. (2012). A satisficing choice model. *Marketing Science*, *31*(6), 878–899.

Swait, J., & Adamowicz, W. (2001). The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research*, *28*(1), 135–148.

Swait, J., & Ben-Akiva, M. (1987). Incorporating random constraints in discrete models of choice set generation. *Transportation Research Part B: Methodological*, *21*(2), 91–102.

Swait, J., Popa, M., & Wang, L. (2016). Capturing context-sensitive information usage in choice models via mixtures of information archetypes. *Journal of Marketing Research*, *53*(5), 646–664.

Toubia, O., De Jong, M. G., Stieger, D., & Füller, J. (2012). Measuring consumer preferences using conjoint poker. *Marketing Science*, *31*(1), 138–156.

Train, K. E. (2009). *Discrete choice methods with simulation* (2nd ed.). Cambridge, UK: Cambridge University Press.

Van der Lans, R., Pieters, R., & Wedel, M. (2008). Eye-movement analysis of search effectiveness. *Journal of the American Statistical Association*, *103*(482), 452-461.

Van Loo, E. J., Nayga, R. M., Campbell, D., Seo, H.-S., & Verbeke, W. (2018). Using eye tracking to account for attribute non-attendance in choice experiments. *European Review of Agricultural Economics*, *45*(3), 333–365.

Wedel, M. (2018). Improving ad interfaces with eye tracking. In K. L. Norman & J. Kirakowski (Eds.), *The Wiley handbook of human computer interaction* (p. 47-76). John Wiley & Sons.

Wedel, M., & Kamakura, W. A. (2000). *Market segmentation: Conceptual and methodological foundations*. New York: Springer Science & Business Media.

Wedel, M., & Pieters, R. (2008). Eye tracking for visual marketing. *Foundations and Trends in Marketing*, *1*(4), 231–320.

Yang, L. C., Toubia, O., & De Jong, M. G. (2015). A bounded rationality model of information search and choice in preference measurement. *Journal of Marketing Research*, *52*(2), 166–183.

	Utility function								Decision rule
Class 1:	U_{ijt1}	$=$	$\beta_i^1 \cdot x_{ijt}^1$	$+$	$\beta_i^2 \cdot x_{ijt}^2$	$+$	$\beta_i^3 \cdot x_{ijt}^3$	$+$ ϵ_{ijt}	Full compensatory
Class 2:	U_{ijt2}	$=$			$\beta_i^2 \cdot x_{ijt}^2$	$+$	$\beta_i^3 \cdot x_{ijt}^3$	$+$ ϵ_{ijt}	Semi-compensatory
Class 3:	U_{ijt3}	$=$	$\beta_i^1 \cdot x_{ijt}^1$	$+$			$\beta_i^3 \cdot x_{ijt}^3$	$+$ ϵ_{ijt}	Semi-compensatory
Class 4:	U_{ijt4}	$=$	$\beta_i^1 \cdot x_{ijt}^1$	$+$	$\beta_i^2 \cdot x_{ijt}^2$			$+$ ϵ_{ijt}	Semi-compensatory
Class 5:	U_{ijt5}	$=$	$\beta_i^1 \cdot x_{ijt}^1$	$+$				ϵ_{ijt}	Lexicographic
Class 6:	U_{ijt6}	$=$			$\beta_i^2 \cdot x_{ijt}^2$	$+$		ϵ_{ijt}	Lexicographic
Class 7:	U_{ijt7}	$=$					$\beta_i^3 \cdot x_{ijt}^3$	$+$ ϵ_{ijt}	Lexicographic
Class 8:	U_{ijt8}	$=$						ϵ_{ijt}	Random choice

The upper notation indicates the preference parameter for a particular attribute (variable): β_i^k

Table 1: Class characteristics in case of 3 attributes

Study	Coffee Makers	Laptops
Number of respondents:	I = 59	I = 70
Number of choice tasks:	T = 12	T = 20
Number of alternatives:	J = 3 + none	J = 4
Number of attributes:	A = 6	A = 6
Attributes:	brand, material, system design, price, price per cup	speed, size, capacity support, antivirus, price
Number of potential classes:	C = 64	C = 64
Choice task design:	Orthogonal and level balanced	Random design
Randomization across subjects:	Yes	No
Incentive compatibility:	No	Yes
Source:	Meißner et al. (2016)	Yang et al. (2015)

Table 2: Description of datasets

		MNL	EAA	EAA(va)	MMNL	MEAA	MEAA(va)
Utility Parameters							
None:	b	-0.19 (0.11)	0.33 (0.13)	0.30 (0.14)	-2.46 (0.87)	-2.53 (0.96)	-2.64 (0.91)
	σ				4.39 (0.78)	4.48 (0.85)	4.67 (0.81)
Braun:	b	0.06 (0.09)	1.51 (0.49)	0.60 (0.50)	0.08 (0.14)	0.52 (0.47)	0.43 (0.32)
	σ				0.39 (0.16)	0.55 (0.37)	0.63 (0.33)
Krups:	b	0.01 (0.09)	0.69 (0.49)	0.38 (0.38)	-0.02 (0.12)	0.15 (0.39)	0.33 (0.41)
	σ				0.05 (0.26)	0.88 (0.45)	0.62 (0.33)
Philips:	b	0.23 (0.09)	0.99 (0.54)	0.62 (0.33)	0.38 (0.12)	0.90 (0.34)	0.68 (0.28)
	σ				0.26 (0.18)	0.16 (0.48)	0.71 (0.35)
Stainless Steel:	b	0.51 (0.07)	1.52 (0.19)	1.43 (0.19)	0.78 (0.11)	1.87 (0.30)	1.70 (0.22)
	σ				0.40 (0.13)	0.36 (0.26)	0.30 (0.19)
Plastic:	b	-0.51 (0.08)	-1.67 (0.25)	-1.51 (0.24)	-0.80 (0.12)	-2.01 (0.35)	-1.78 (0.27)
	σ				0.41 (0.16)	0.60 (0.40)	0.53 (0.27)
Pad:	b	0.22 (0.05)	1.73 (0.24)	1.50 (0.20)	0.33 (0.12)	1.99 (0.28)	1.90 (0.28)
	σ				0.79 (0.14)	0.16 (1.07)	0.51 (0.40)
Design A:	b	-0.29 (0.09)	-2.20 (0.75)	-1.81 (0.53)	-0.48 (0.13)	-2.61 (1.09)	-2.06 (0.58)
	σ				0.01 (0.21)	0.02 (0.84)	0.05 (0.48)
Design B:	b	0.03 (0.09)	0.32 (0.37)	0.25 (0.34)	0.05 (0.12)	0.31 (0.43)	0.23 (0.36)
	σ				0.09 (0.17)	0.17 (0.53)	0.19 (0.44)
Design C:	b	0.13 (0.09)	1.40 (0.41)	1.26 (0.30)	0.24 (0.12)	1.73 (0.72)	1.43 (0.35)
	σ				0.20 (0.28)	0.47 (0.52)	0.07 (0.40)
Price per cup:	b	-0.80 (0.07)	-1.76 (0.18)	-1.74 (0.16)	-1.34 (0.16)	-1.77 (0.28)	-1.90 (0.22)
	σ				0.83 (0.14)	0.86 (0.20)	0.64 (0.19)
Price:	b	-2.12 (0.17)	-3.72 (0.36)	-3.72 (0.33)	-3.45 (0.39)	-4.19 (0.48)	-4.13 (0.43)
	σ				1.78 (0.34)	1.61 (0.43)	1.59 (0.40)
Class Parameters							
Brand			-2.08 (0.69)	-0.08 (1.46)		-0.54 (0.82)	1.59 (1.48)
Material			-0.15 (0.35)	0.28 (0.52)		-0.03 (0.36)	0.86 (0.60)
System			-1.42 (0.38)	-1.84 (0.47)		-1.01 (0.33)	-2.02 (0.64)
Design			-1.76 (0.58)	-2.05 (0.66)		-1.72 (0.60)	-2.64 (0.87)
Price per cup			0.57 (0.37)	-0.13 (0.48)		2.08 (1.29)	0.72 (0.71)
Price			1.60 (0.53)	1.03 (0.58)			
Visual attention				1.94 (0.36)			2.88 (0.66)
In Sample Fit							
LL		-745.11	-688.89	-656.50	-600.81	-578.33	-540.66
BIC		1568.98	1495.90	1437.68	1359.11	1346.96	1278.19
ρ^2		0.24	0.30	0.33	0.39	0.41	0.45
Out-of-Sample Fit							
Hitrate		0.55	0.57	0.60	0.64	0.67	0.68
Hitprob		0.43	0.49	0.51	0.59	0.62	0.63

Note: standard errors are indicated in parentheses.

Table 3: Parameters and fit measures of estimated models for coffee makers

	MNL	EAA	EAA(va)	MMNL	MEAA	MEAA(va)
Utility Parameters						
1.6 Ghz: b	-1.69 (0.11)	-2.26 (0.19)	-2.72 (0.20)	-2.71 (0.21)	-3.16 (0.26)	-3.00 (0.24)
σ				1.15 (0.15)	0.89 (0.15)	1.08 (0.16)
1.9 Ghz: b	-0.69 (0.09)	-0.91 (0.12)	-1.14 (0.13)	-1.17 (0.13)	-1.46 (0.16)	-1.36 (0.16)
σ				0.24 (0.16)	0.56 (0.15)	0.07 (0.34)
2.7 Ghz: b	0.98 (0.09)	1.19 (0.12)	1.45 (0.13)	1.62 (0.13)	1.84 (0.16)	1.73 (0.15)
σ				0.39 (0.12)	0.01 (0.23)	0.24 (0.18)
26 cm: b	-0.31 (0.06)	-1.68 (0.24)	-2.04 (0.24)	-0.44 (0.14)	-0.38 (0.22)	-0.07 (0.28)
σ				1.34 (0.16)	2.25 (0.33)	2.69 (0.40)
35.6 cm: b	0.22 (0.08)	0.11 (0.17)	0.23 (0.18)	0.29 (0.13)	0.76 (0.23)	0.74 (0.22)
σ				0.73 (0.12)	0.84 (0.14)	0.97 (0.16)
40 cm: b	-0.03 (0.09)	0.41 (0.19)	0.52 (0.22)	0.00 (0.12)	0.13 (0.20)	0.06 (0.20)
σ				0.13 (0.16)	0.21 (0.25)	0.51 (0.27)
160 GB: b	-1.16 (0.10)	-1.99 (0.19)	-2.11 (0.19)	-1.82 (0.16)	-2.48 (0.26)	-2.40 (0.24)
σ				0.69 (0.11)	0.56 (0.17)	0.43 (0.17)
320 GB: b	0.05 (0.08)	0.23 (0.12)	0.15 (0.12)	0.05 (0.12)	0.08 (0.14)	0.05 (0.14)
σ				0.06 (0.13)	0.07 (0.24)	0.33 (0.13)
500 GB: b	0.61 (0.08)	1.00 (0.13)	1.12 (0.11)	0.99 (0.11)	1.25 (0.14)	1.22 (0.14)
σ				0.07 (0.10)	0.00 (0.12)	0.01 (0.10)
1 year support: b	-0.15 (0.10)	0.49 (0.26)	0.78 (0.37)	-0.29 (0.13)	-0.27 (0.13)	-0.24 (0.12)
σ				0.17 (0.13)	0.27 (0.13)	0.30 (0.13)
2 year support: b	0.17 (0.08)	0.54 (0.27)	0.37 (0.42)	0.23 (0.11)	0.15 (0.11)	0.14 (0.11)
σ				0.29 (0.11)	0.16 (0.16)	0.02 (0.19)
3 year support: b	0.03 (0.07)	-0.18 (0.22)	-0.24 (0.36)	0.06 (0.10)	0.05 (0.10)	-0.02 (0.10)
σ				0.03 (0.09)	0.01 (0.12)	0.00 (0.12)
30 days antivirus: b	-0.01 (0.06)	-3.28 (0.55)	-3.29 (0.54)	-0.05 (0.09)	-3.56 (0.81)	-3.60 (0.67)
σ				0.36 (0.11)	0.99 (0.97)	0.51 (0.73)
1 year antivirus: b	-0.10 (0.09)	0.25 (0.33)	0.24 (0.33)	-0.18 (0.12)	0.05 (0.45)	0.00 (0.39)
σ				0.12 (0.10)	0.19 (0.48)	0.30 (0.51)
2 year antivirus: b	0.19 (0.09)	1.95 (0.41)	1.81 (0.38)	0.29 (0.13)	2.07 (0.51)	1.99 (0.47)
σ				0.21 (0.13)	0.22 (0.59)	0.30 (0.38)
Price: b	-0.36 (0.03)	-0.98 (0.06)	-0.95 (0.07)	-0.56 (0.07)	-0.76 (0.11)	-1.08 (0.12)
σ				0.95 (0.10)	0.74 (0.07)	0.66 (0.08)
Class Parameters						
Speed		2.21 (0.77)	-1.28 (0.50)		2.65 (0.79)	-0.83 (1.01)
Size		-0.89 (0.40)	-2.78 (0.49)		0.36 (0.45)	-2.74 (0.74)
Capacity		0.82 (0.41)	0.44 (0.37)		1.28 (0.49)	1.16 (0.56)
Support		-1.37 (0.77)	0.08 (1.28)			
Antivirus		-2.57 (0.49)	-0.70 (0.56)		-2.47 (0.53)	0.72 (0.79)
Price		-0.18 (0.26)	-0.52 (0.36)		2.31 (2.20)	1.39 (0.60)
Visual attention			2.34 (0.35)			4.79 (0.84)
In Sample Fit						
LL	-1415.82	-1199.59	-1164.76	-1111.62	-1067.89	-1020.64
BIC	2947.56	2558.56	2496.14	2455.05	2403.81	2316.56
ρ^2	0.27	0.38	0.40	0.43	0.45	0.47
Out-of-Sample Fit						
Hitrate	0.56	0.61	0.63	0.70	0.70	0.70
Hitprob	0.43	0.54	0.55	0.62	0.64	0.64

Note: standard errors are indicated in parentheses.

Table 4: Parameters and fit measures of estimated models for laptops

	across all individuals			across individuals with attribute attendance	
	MMNL	MEAA	MEAA(va)	MEAA	MEAA(va)
Coffee makers:					
Brand	9.2%	6.8%	8.0%	23.7%	21.5%
Material	16.6%	16.3%	15.9%	32.0%	30.2%
System	10.5%	8.1%	8.8%	30.0%	30.5%
Design	7.8%	3.8%	4.4%	32.4%	32.2%
Price per cup	25.1%	28.5%	25.7%	30.6%	35.3%
Price	30.8%	36.4%	37.3%	36.4%	37.3%
Laptops:					
Speed	35.4%	37.9%	39.9%	39.6%	43.0%
Size	15.0%	13.6%	14.6%	25.7%	30.1%
Capacity	19.9%	19.6%	18.4%	24.9%	27.5%
Support	4.0%	3.5%	3.4%	3.5%	3.4%
Antivirus	4.2%	2.2%	2.7%	30.4%	37.8%
Price	21.5%	23.2%	21.0%	23.2%	34.9%

Table 5: Average relative importance of attributes

	across all individuals			across individuals with attribute attendance	
	MMNL	MEAA	MEAA(va)	MEAA	MEAA(va)
Brand					
Braun vs. Severin	28.98	20.99	21.51	72.83	57.67
Krups vs. Severin	22.10	18.98	20.45	65.87	54.83
Philips vs. Severin	43.50	23.00	22.73	79.83	60.94
Material					
Stainless Steel vs. Aluminum	55.42	23.61	25.04	46.44	47.66
Plastic vs. Aluminum	-25.42	-27.63	-24.84	-54.33	-47.27
System					
Pad vs. Capsule	28.81	29.79	30.25	109.85	105.00
Design					
A vs. D	-33.56	-10.40	-9.97	-87.61	-73.50
B vs. D	-7.68	-0.96	-0.84	-8.10	-6.17
C vs. D	0.54	3.88	4.11	32.67	30.30

Table 6: Average willingness-to-pay for coffee makers

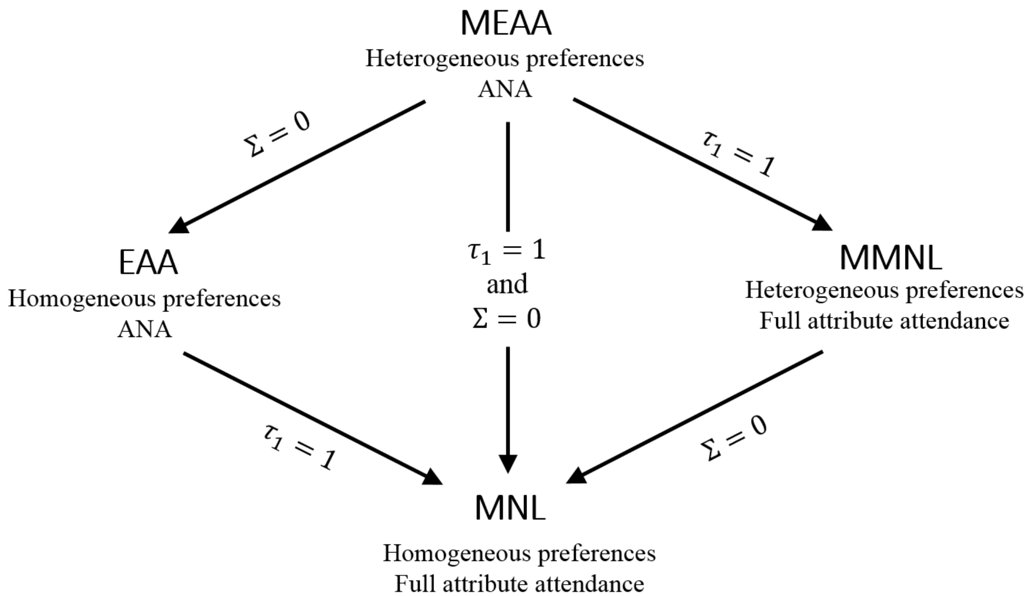


Figure 1: The main and restricted models

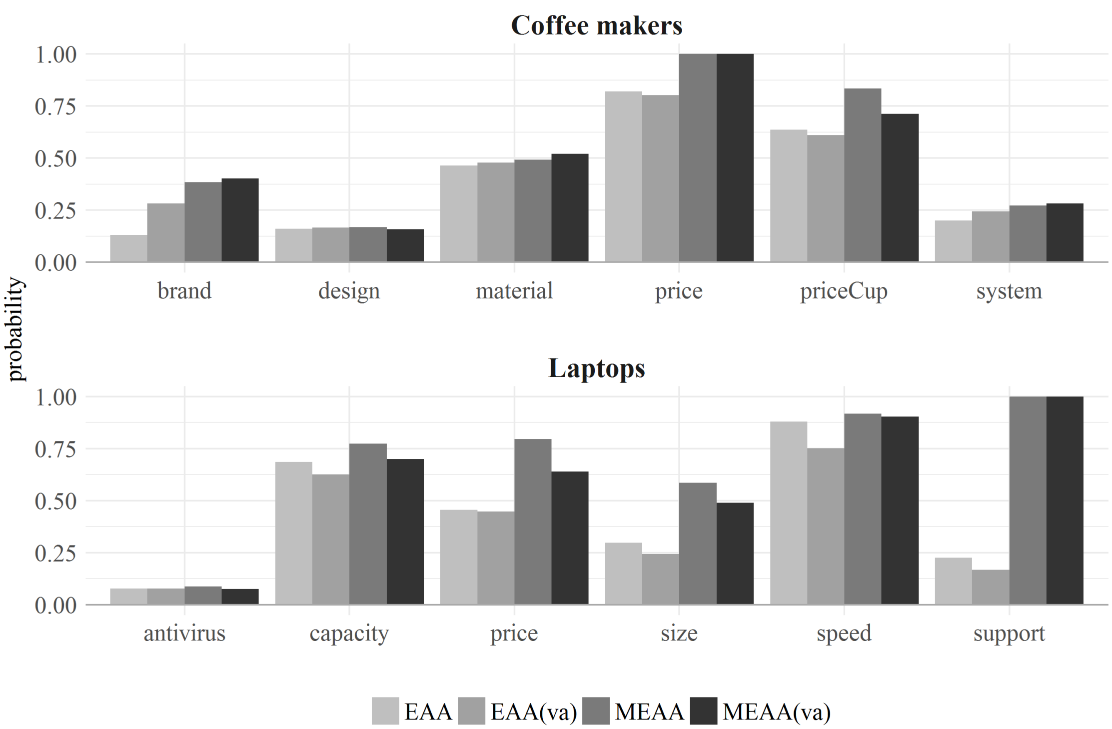


Figure 2: Attribute attendance probabilities

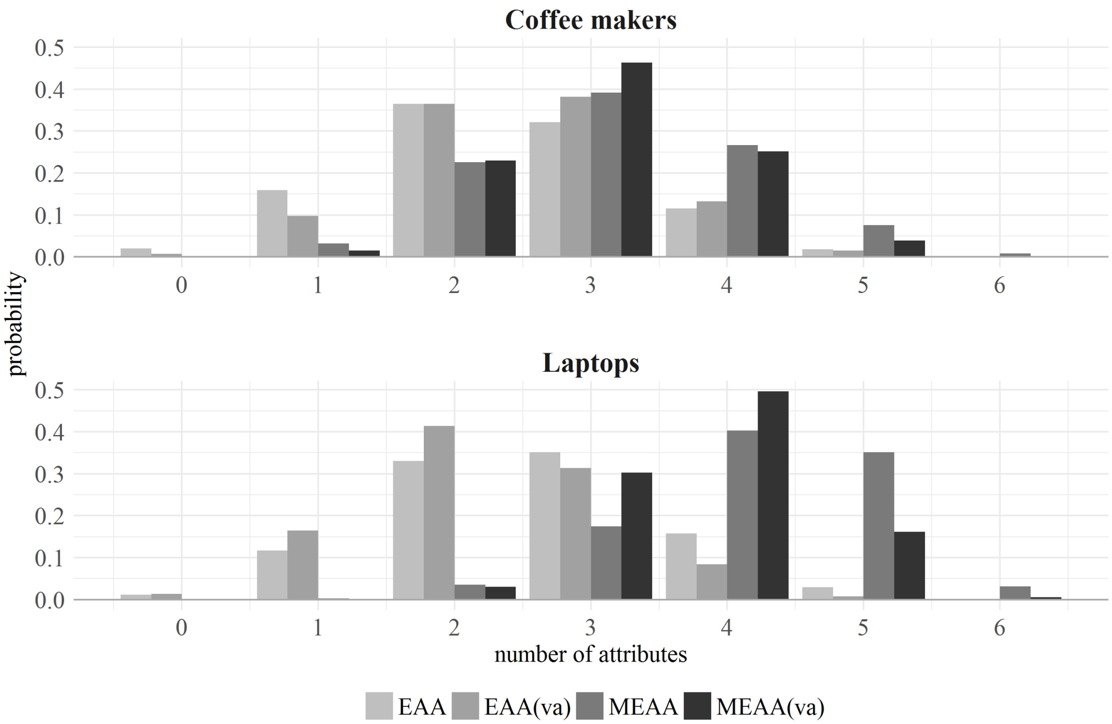
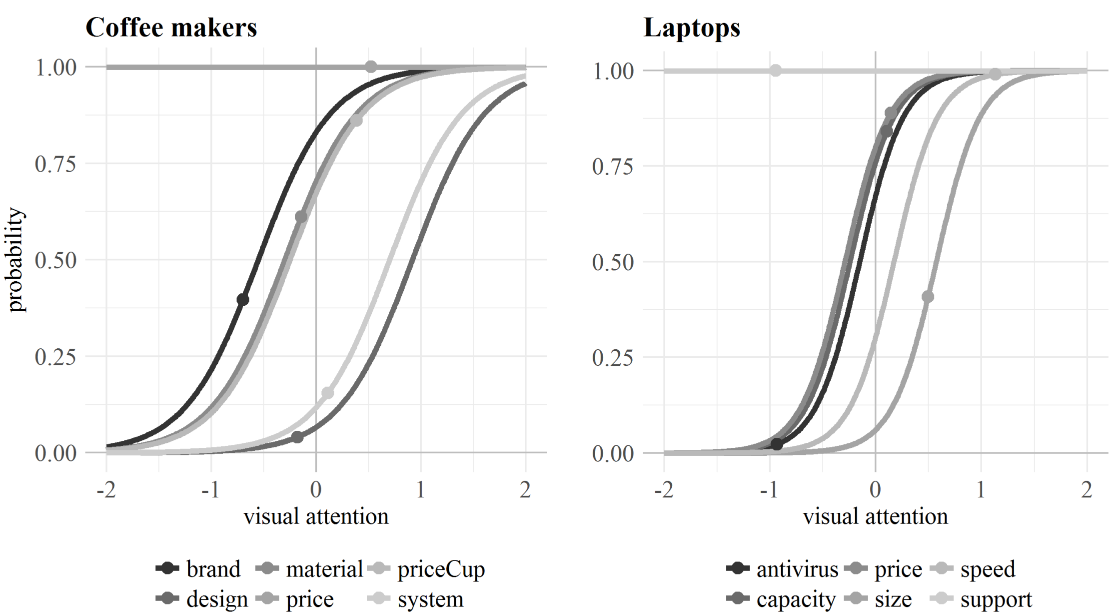


Figure 3: Probability of attending a certain number of attributes



Note: points represent the mean visual attention for a given attribute.

Figure 4: Effect of visual attention on attribute attendance probability

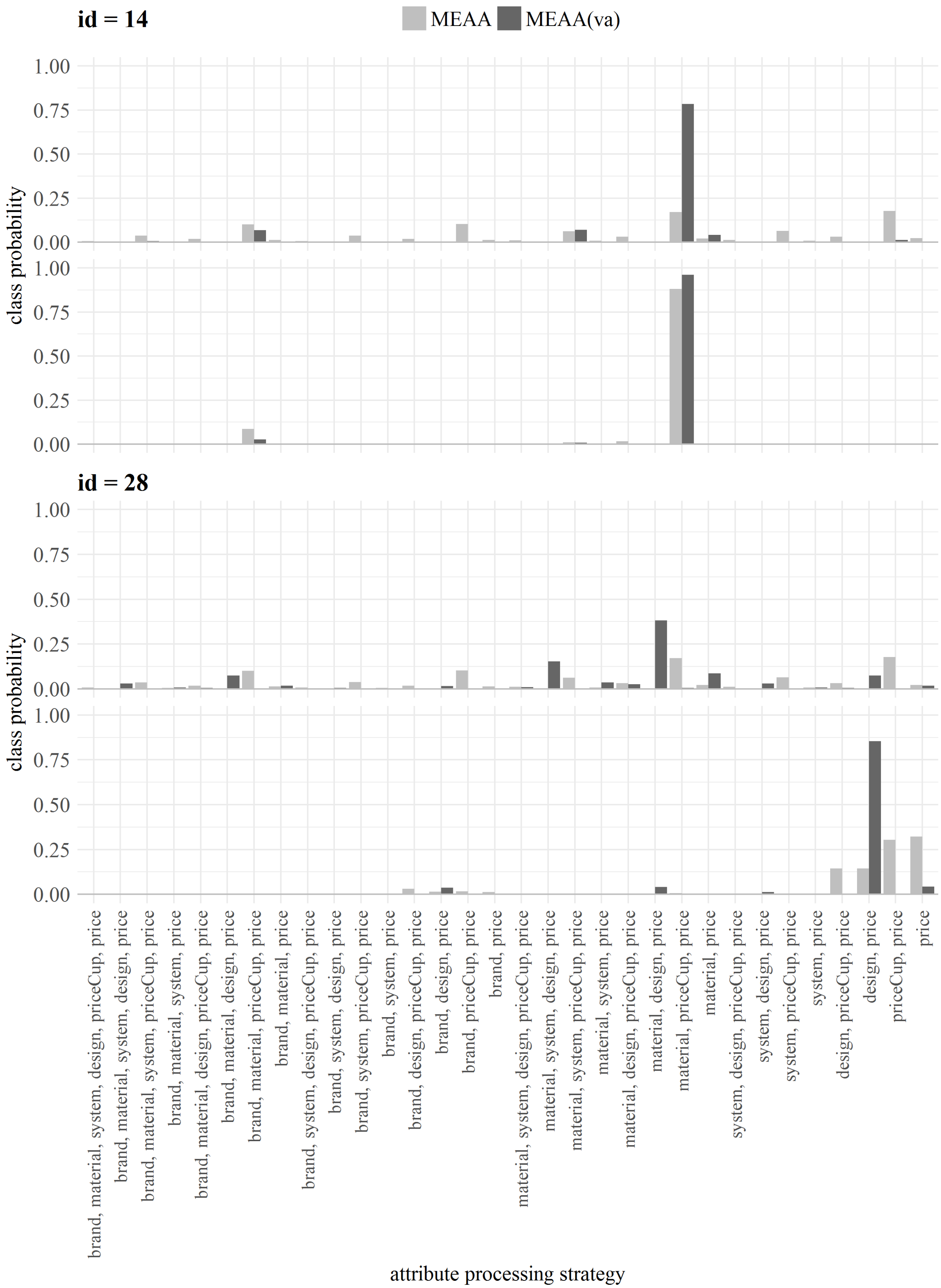


Figure 5: Class allocation based on MEAA models for selected individuals in the coffee makers study

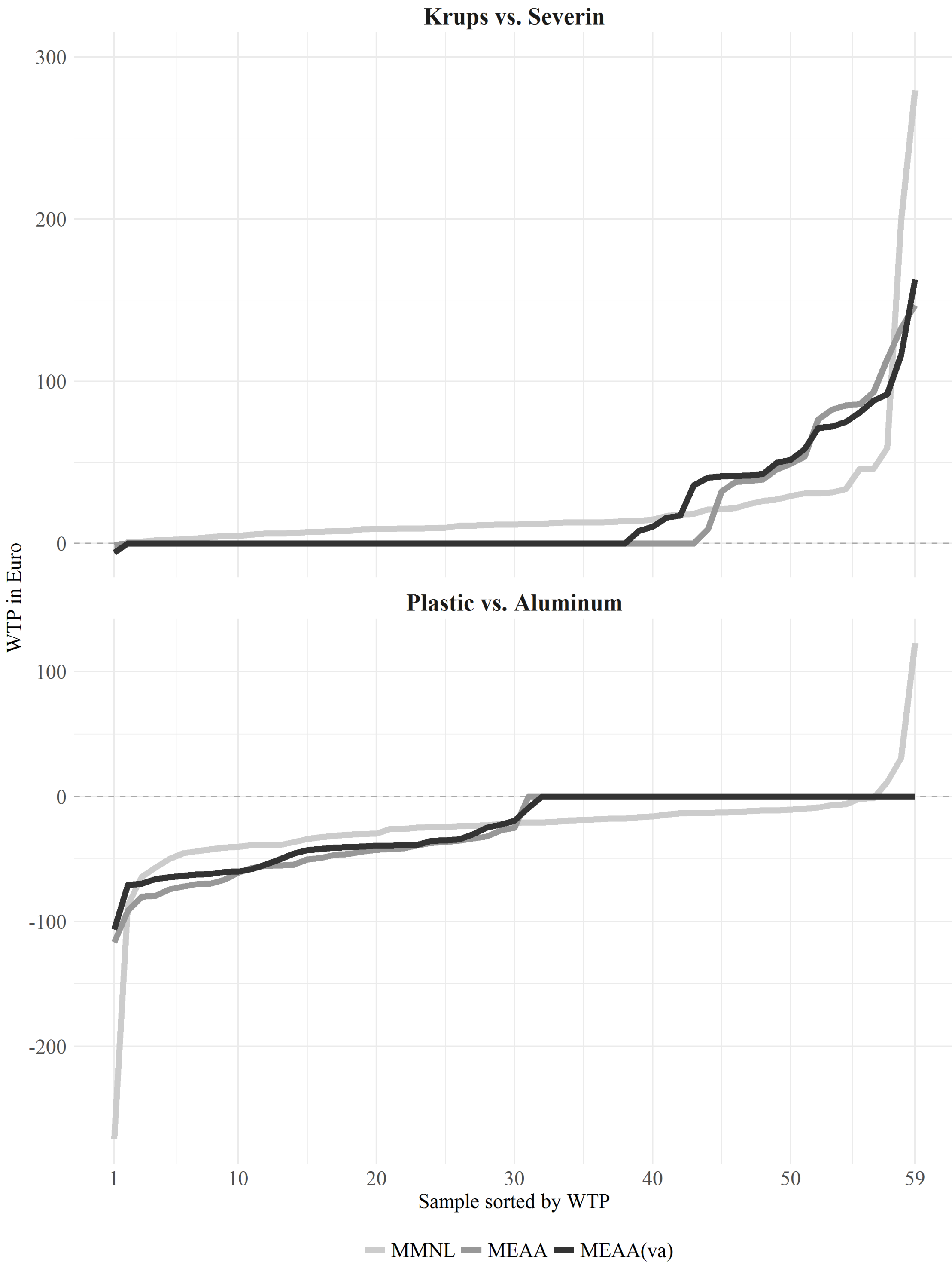


Figure 6: Cumulative willingness-to-pay distribution of selected attribute level comparisons for coffee makers