Attribute Non-Attendance as an Information Processing Strategy in Stated Preference Choice Experiments: Origins, Current Practices, and Future Directions

Daniel K. Lew[♦]
Alaska Fisheries Science Center, NOAA, National Marine Fisheries Service

John C. Whitehead[♣]
Department of Economics, Appalachian State University

ABSTRACT

Stated preference discrete choice experiments (CE) are increasingly being used by researchers seeking to understand people's preferences and values in environmental economics, transportation, health, and marketing. An active CE research area relates to behaviors that break from the assumptions of full rationality assumed in standard discrete choice models. In particular, considerable attention in recent years has been on attribute non-attendance (ANA), a type of choice behavior where individuals ignore one or more attributes in CE questions. In this article, we delve into the origins and motivations for the study of ANA as an information processing strategy, delineate the variety of approaches that have developed in the growing literature to identify and account for ANA behavior, and discuss several promising directions for this literature that could enhance our understanding of decision-making in CE studies.

Key words: Attribute non-attendance, attribute processing strategies, discrete choice experiments, stated preference, bounded rationality.

JEL Classification Q5, C5, D8, D9

Short title: Attribute Non-Attendance in Choice Experiments

Acknowledgments: This article and its findings are those of the authors and do not necessarily reflect the views of the National Marine Fisheries Service or the U.S. Department of Commerce. We thank Steve Kasperski, Rosemary Kosaka, and an anonymous reviewer for useful suggestions that have improved this article. Any remaining errors are our own.

[♦]**Daniel K. Lew** is an economist, NOAA, National Marine Fisheries Service, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115 USA (e-mail: Dan.Lew@noaa.gov).
[♣]**John C. Whitehead** is a professor, Department of Economics, Appalachian State University, Boone, North Carolina 28607 (e-mail: whiteheadjc@appstate.edu)

Introduction

Discrete choice experiments (CE) are now a common tool for measuring economic preferences and values for both market and non-market goods in transportation (Louviere 1988), marketing (Louviere and Woodworth 1983), health (Louviere and Lancsar 2009; de Bekker-Grob, Ryan, and Gerard 2012), and environmental economics (Hanley, Wright, and Adamowicz 1998; Johnston et al. 2017). In a typical CE question, an individual is presented with a choice between two or more alternatives that are each described in terms of multiple attributes. Most commonly, the individual is asked to choose the best, or most-preferred, alternative from among the choices. Under the rational choice model of behavior, the individual in this choice setting is assumed to have well-formed preferences, full information, and be utility maximizing and will therefore choose the alternative that yields the highest utility.

However, research in diverse fields, most notably in the psychology literature, has long documented behavior inconsistent with rational choice theory that leads to "irrational choices" (Gilovich et al. 2002). In behavioral economics, the concept of bounded rationality (Simon 1955, 1956, 1959), which recognizes the cognitive and environmental constraints on an individual's ability to act in a rational manner, has provided a platform from which to explore issues like complexity and information processing strategies and how they affect choice. In the CE literature, recognition of the limitations of the rational choice model and efforts to better address the potential for deviations from it have grown tremendously over the last two decades. While most of the early work focused on the role of complexity on choice behavior (Mazzota and Opaluch 1995; DeShazo and Fermo 2002; Dellaert et al. 1999), a focus on choice heuristics—simplifying information processing strategies—emerged with studies by David Hensher and associates (Hensher 2004, 2006a, 2006b; Hensher, Rose, and Greene 2005).

Much of the work in this area has focused on a specific type of information processing strategy, one focused on how people process attribute information (attribute-processing strategies). A number of attribute processing strategies have been surmised in the literature, including common-metric aggregation (Layton and Hensher 2010), common-metric parameter transfer assignment (Hensher and Layton 2010), attribute thresholds (Hensher and Rose 2012; Campbell, Hensher, and Scarpa 2012), reference point revision and value learning (DeShazo 2002), and attribute non-attendance (ANA). Our focus here is primarily on ANA, which occurs when individuals ignore one or more attributes, though some discussion of the other attribute-processing strategies are discussed in the context of studies employing ANA.

Why should we pay particular attention to the ANA literature? A focus is warranted on ANA for several reasons. First, there has been considerable growth in the number of studies focusing on this attribute processing strategy relative to others such as common-metric aggregation (Hensher and Layton 2010) that require specific types of choice designs (Lew and Whitehead, this issue). Second, the recent best practices guidance for stated preference valuation by Johnston et al. (2017) mentions ANA explicitly in terms of behavioral response anomalies that should be considered in the study design (experimental design), pretesting, and analysis (construct validity testing) stages. Third, ANA behavior can have serious implications on welfare measurement, particularly as it relates to the cost attribute. For individuals who ignore the cost attribute, there is no trading off of money for changes in other attribute levels. As a result, welfare estimation is precluded for these individuals. For individuals ignoring other attributes, welfare values associated with those attributes are precluded. Thus, if ANA is present but not accounted for, welfare estimates will be biased (Lew and Whitehead, this issue). Finally, even in studies that account for other attribute processing strategies (concurrently with ANA), some amount of attribute non-attendance behavior is typically found.

This article differs from another recent review by Hensher (2014) in a few important ways. First, Hensher (2014) is broader in scope, covering ANA, attribute thresholds, and reference point revision and learning. Second, more narrowly focusing on ANA allows us to provide depth of coverage to the methods used in the current literature and the general findings from the dynamic and growing empirical literature on the subject. Third, the empirical ANA literature has expanded in the years since Hensher's review.

The next section provides more background about the origins of ANA and other attribute processing strategies. This is followed by an exposition of the different approaches that have emerged in the literature to identify and account for ANA behavior in choice experiments. In the subsequent section, we examine welfare estimation issues associated with the different types of ANA strategies. We conclude with a discussion of the implications of this review for past stated preference research and future directions for ANA-related research.

Origins of Attribute Non-Attendance and other Attribute-Based Processing Strategies

ANA and other information processing strategies are often interpreted as rational behavioral responses to complex choices. These strategies are viewed through the lens of bounded rationality. In this section, we discuss the origins of ANA by first looking at its roots in the concept of bounded rationality. Then, we discuss early explorations into ANA behavior in choice experiments.

Roots in Bounded Rationality

In the neo-classical economic model, individuals are assumed to be self-interested, utility maximizing agents with perfect knowledge—a view of rational decision makers as "Economic Man" (Simon 1959) or *Homo economicus*. Perfect knowledge assumes not only that the individual has full information about the choices she makes—including a full understanding of the benefits and costs of choice outcomes--but also limitless resources to cognitively gather and process this information. This standard rationality assumption began to be challenged from those inside and outside economics as empirical evidence mounted of "irrational" behavior (Gilovich, Griffin, and Kahneman 2002; Meehl 1954, 1986; Dawes, Faust, and Meehl 1991; Conlisk 1996). An influential empirical finding that raised the question of whether people made rational decisions was a study by Meehl (1954). He documented how judgments based on clinical (subjective) assessments differed from actuarial methods (ones based on a formula) using the same data. In fact, reflecting back on this work decades later, Meehl noted that "There are no strong arguments, from the armchair or from empirical studies of cognitive psychology, for believing that human beings can assign optimal weights in equations subjectively or that they apply their own weights consistently" (Meehl 1986, p. 372). Conlisk (1996) provides a useful summary of the literature that directly tests the rationality in individuals, noting that even in experiments with objectively correct answers that are often simple in nature, responses are "frequently way off" (p. 670). Observations such as these suggest limitations of rational choice theory to explain decision-making in a number of contexts.

In economics, a theoretical explanation for deviations from the rational choice model was offered by Simon (1955, 1956, 1959). Simon proposed the idea that limited processing capabilities (limits to time, attention, resources, information, mental processing abilities, and environmental factors affecting the choice setting) effectively constrain a person's ability to act in a fully rational manner. But within these constraints, individuals will operate rationally. This "bounded rationality" concept maintained the primary idea within the rational choice model—that individuals are optimizers—but allowed for factors that constrain their information search and computational capacities and, thus, their optimization behavior. This brought to light a fundamental insight largely absent in economics (but known in other fields) up to that point—that cognitive processes, and not just outcomes, matter in decision-making (Simon 1978).

This insight helped fuel research into alternative decision-making strategies that individuals may employ in response to their limitations in choice settings. Some notable decision-making strategies developed as alternatives to the standard rational choice model include satisficing (Simon 1955), elimination-by-aspects (Tversky 1972), lexicographic choices (Fishburn 1974), and the majority of confirming dimensions strategy (Russo and Dosher 1983). These alternative decision strategies represent departures from the standard rational choice decision strategy that Bettman, Luce, and Payne (1998) described as a "weighted adding" strategy that assumes individuals (1) fully assess the importance of each attribute and assign subjective values for each attribute level; (2) evaluate each alternative separately by evaluating every attribute; (3) sum up the subjective values for each attribute weighted by their importance into a single score for each alternative; and (4) then choose the alternative with the highest score.

A parallel research area has focused on explaining the effects of there being bounds on choice. Much of this literature has focused on examining the role of complexity and its implications on choices within a rational choice, or weighted adding, decision framework. Complexity of choices can arise due to a number of factors—the number of choice alternatives, the amount of information readily available, the quality of that information, and the differences between choice alternatives. In effect, this research has sought to explain the implications of added complexity of choices within a decision framework that acknowledges the effects of limitations of one's cognitive ability in the face of complex choices. Heiner (1983), for example, argued that the standard theory suggests no difference between an individual's "competence" and the "difficulty" of the decision task. The presence of a gap between the two, which he called the "C-D gap," will introduce uncertainty that increases the likelihood of individuals making errors in their decisions. These errors, in turn, result in "irrational behavior" and can be represented in econometric modeling in the stochastic component of utility. Specifically, this suggests the presence of a C-D gap will lead to increases in utility variance. This is frequently examined empirically in CE studies by allowing the scale parameter in random utility models to be a function of the complexity of the choice, as reflected in the attribute level variation within a choice set (Dellaert et al. 1999; DeShazo and Fermo 2002). Using this scale-heteroscedastic modeling approach, numerous empirical studies have found evidence that increased choice complexity increases utility error, as reflected in scale parameter differences (Mazzotta and Opaluch 1995; Dellaert et al. 1999; Swait and Adamowicz 2001; Foster and Mourato 2002; Saelensminde 2002; DeShazo and Fermo 2002). An implicit and common assumption of the complexity literature is that individuals will use all information in the decision-making process even in the face of increased complexity or difficulty (or conversely diminished competence) that results in increased errors. This is consistent with what DeShazo and Fermo (2004) term the passive-bounded rationality model (e.g., de Palma, Myers, and Papageorgiou 1994).

A competing behavioral model to the passive bounded rationality model provides a link to ANA and other attribute processing behavior. This competing model, which DeShazo and Fermo (2004) term the rationally-adaptive rational choice model builds off of insights from Bettman, Luce, and Payne (1988) and Gabaix and Laibson (2000). Bettman, Luce, and Payne

(1988) argue that due to cognitive limitations individuals often do not have well-defined preferences and thus need to construct them using a variety of strategies. However, information collection and evaluation is costly, so the individual must evaluate the benefits and costs of doing so. Gabaix and Laibson (2000) suggest that individuals may be rationally-adaptive in the attention they allocate across information in the choice setting in a way that maximizes the benefits and minimizes the costs of information evaluation.

In their paper, DeShazo and Fermo (2004) use the rationally-adaptive rational choice model as the basis from which to challenge the notion that, empirically, bounded rationality should be expected to manifest solely as increased utility variance (as measured in the scale parameter in choice models)—the common treatment up to that point in the literature. They argue instead that individuals may sequentially evaluate the (costly) information up to the point where the marginal benefits and marginal costs of further information evaluation are equal (thus maximizing the net benefits of the information to the choice process) in a rationally-adaptive process. As a result, this process would likely result in the adoption of choice heuristics to simplify decision-making but would not be captured in the scale-heteroscedastic logit-based analyses done to that point within the complexity literature. Attribute non-attendance is one such outcome of the information evaluation done by the individual at the "process" stage of the overall process-outcome decision-making process. Their empirical application involved a stated preference conjoint survey with an experimental design that varied the complexity of choices in terms of the number of alternatives in each choice set, number of attributes in each alternative, and the difficulty of choices—as reflected by the similarities and differences in the attributes and attribute levels within choice sets. They found strong evidence in favor of the rationally-adaptive model that suggests individuals allocate their attention to choice information and likely often use less than the full information available to them as a result. Their empirical findings also suggest that in the presence of less than full attendance to choice information, like to attribute information, standard models will lead to biased parameter estimates and those that account for scale heterogeneity may also be overestimating utility variance. In short, they find empirical evidence that in the face of complexity individuals may use less than the full information available to them, which has implications on both preference parameters and scale.

DeShazo and Fermo (2004) cogently made the argument for the importance of considering the "process" stage of the process-outcome decision-making framework in choice experiments, and they importantly identified attribute non-attendance as a type of choice heuristic that could arise in the information processing stage that researchers should be wary of when modeling preferences. However, their empirical application and estimation models corresponding to the adaptively-rational model were specific to the unique nature of their study and were not designed for analyzing CE data when the number of alternatives and attributes are fixed in the study design, which is more common.

Evolution of the study of ANA in choice experiments

The first directed investigations of ANA behavior in a CE study to appear in the published literature are a pair of studies done by Hensher, Rose, and Greene (2005) and Rose, Hensher, and Greene (2005).[1] In the Hensher, Rose, and Greene (2005) study, data were analyzed from a face-to-face computer-assisted personal interview (CAPI) survey that asked both CE questions about commuting trip alternatives and a follow-up question asking respondents which attributes describing the alternatives they ignored when answering the CE

[1] Hensher (2004) appears to be an early draft of this work.

questions. The latter question yields self-reported information on ANA behavior and is referred to as *stated ANA* information. The stated ANA data on which attributes each respondent ignored were used to condition the utility function, which took the form of setting the parameters associated with ignored attributes to zero for those individuals. The authors termed this utility parameter conditioning approach for bounding the attribute processing task "attribute elimination." Accounting for this stated ANA behavior resulted in a downward shift in willingness-to-pay (WTP) relative to the model that did not account for ignored attributes. The second study, by Rose, Hensher, and Greene (2005), used data from a survey of university marketing students who were asked CE questions related to choice of airline carrier for interstate holiday travel in Australia. It utilized the same binary response stated ANA questions and the same attribute elimination estimation approach for accounting for ANA behavior.

Hensher (2006b) and Hensher, Rose, and Bertoia (2007) proposed an alternative to the attribute elimination approach that tries to account for the uncertainty surrounding whether attributes respondents say they ignore are actually ignored. Using the same data as Hensher, Rose, and Greene (2005), a two-stage model was proposed that allows for stochastic, rather than deterministic, attribute exclusion. In the first stage, choice models assuming each of the observed patterns of stated ANA (i.e., full attendance to all attributes and attendance to partial sets of attributes) in the data are separately estimated assuming conditional indirect utility is a function of the attributes and demographics (age and income). In the second-stage choice model, the log-sum of expected utility across the separate stated ANA models are generated from the first-stage choice models for each respondent (a measure of the expected utility across the possible ANA behaviors) and interacted with the attributes in the second-stage utility specification to condition the marginal effects of the attributes on the ANA behavior.

In a different study, Hensher (2006a) examines the role of study design on stated ANA, specifically the effect of design features on the propensity of respondents to ignore attributes. A series of ordered scale-heterogeneity logit models are estimated for different subsets of a design-of-designs survey dataset, where the dependent variable is the number of attributes ignored as reported by respondents in the survey. He found that more attributes were ignored as the design becomes more complex. Complexity in this case was defined by the number of attributes describing each alternative, deviation of attribute levels from a familiar or established baseline level for the individual, the use of multiple common-metric attributes (that could be "added up"), the number of choice sets evaluated, and personal income of the respondent.

In contrast to these early studies that relied on stated ANA information, a study by Gilbride, Allenby, and Brazell (2006) in the marketing literature introduced a Bayesian approach for inferring ANA behavior from CE data without stated ANA information. Building off of the stochastic search variable selection model of George and McCulloch (1993), they developed a model that captures ANA behavior by modeling individual-level heterogeneity in the utility parameters with distributions with mass concentrated at zero (if ANA is present) or away from zero (if ANA is not present). Scarpa et al. (2009) termed this the "stochastic attribute selection" (STAS) model and used it to compare to a non-Bayesian finite mixture modeling approach in a CE study of rural landscape values.

The latter model, introduced by Scarpa et al. (2009) and called the equality-constrained latent class (ECLC) model, is based on the latent class logit model (Boxall and Adamowicz 2002). The model assumes individuals fall into one of several discrete and latent classes, where each latent class is defined by which attributes are attended to and which are not. Further, the model assumes that across classes, utility parameters for the attributes that are attended to are the same—i.e., the utility parameters are equality-constrained (hence the name). This contrasts with the standard formulation of latent class logit models that typically allow the preference parameters to differ across classes. The STAS and ECLC models are referred to as *inferred ANA* models since they infer ANA behavior directly from the choice data and without the aid of stated ANA information. Together with the attribute elimination stated ANA modeling approach discussed above, these inferred ANA models represent the early foundational models used and then built upon in subsequent studies of ANA.

Recently, there has been the emergence of a third type of data used to identify ANA behavior in choice experiments. *Visual ANA* data are observed behavioral information on CE respondents' attention to individual attributes obtained through biometric data on eye movements (Balcombe, Fraser, and McSorley 2015; van Loo et al. 2018). These studies typically are conducted in lab settings since they utilize specialized eye-tracking equipment (e.g., head-mounted video-based eye trackers and large high-resolution video monitors) and require participants to minimize head movement to get accurate measurements. Eye movement information is recorded while respondents answer CE questions. Attention to specific attributes in these studies is measured by analyzing the frequency and duration of time spent looking at each attribute. Visual non-attention to an attribute is presumed to be indicative of ANA. As discussed later, studies employing this recently-developed approach generally use similar analytic approaches developed to analyze stated and inferred ANA data. In the next sections, after discussing the standard models used to model CE data, we discuss the early models used to analyze ANA data and subsequent variants more formally.

Standard approaches to modeling choice experiment data

In a typical CE study, an individual is presented with T ($t = 1,\ldots,T$) CE questions. Each CE question involves a choice between J ($j = 1,2,\ldots,J$) alternatives each of which is described in terms of K ($k = 1,2,...,K$) attributes. For a given CE question, the individual is asked to choose the best or most-preferred alternative from among the J alternatives. Under the rational choice model of behavior, the individual is assumed to have well-formed preferences, full information, and be utility maximizing. Therefore, the individual will choose the alternative that yields the highest utility from among the J alternatives in the specific CE question's choice set.

Random utility maximization (RUM) models (or more commonly "random utility models"), are the workhorse models for analyzing CE data (Luce 1959; McFadden 1974; Manski 1977). In RUM model formulations, utility for the jth alternative in the tth question (U_{jt}) is the sum of a deterministic (V_{jt}) and stochastic (ε_{jt}) component, such that $U_{jt} = V_{jt} + \varepsilon_{jt}$. The deterministic component of utility is modeled with observable variables by the researcher while the idiosyncratic component of utility is known to the individual but not the researcher. The probability that the individual chooses the jth alternative is denoted $P_{jt} = \Pr[U_{jt} \geq \max\{U_{1t}, U_{2t},\ldots U_{Jt}\}]$. It is common to assume that the idiosyncratic error associated with the jth alternative, $\varepsilon_{jt,}$ is independent and identically Gumbel distributed across alternatives and CE questions with a scale parameter $\lambda > 0$ (such that the variance equals $\pi^2/6\lambda$), which leads to the conditional logit (CL) model and probabilities of the form:[2]

$$P_{jt} = \exp(\lambda \cdot V_{jt})/\Sigma_k [\exp(\lambda \cdot V_{kt})] \qquad j,k \in \{1,\ldots,J\}. \tag{1}$$

The panel CL model is based on the joint probabilities of observing the set of CE responses across the T questions, which are calculated as the product of the individual probabilities.

[2] Logit-based RUM models of the form presented here are the most common ones used to analyze CE data, but others exist.

Note that the scale parameter is inversely proportional to the variance of the idiosyncratic error (Ben-Akiva and Lerman 1985; Swait and Louviere 1993). While it is commonly normalized to be one (λ = 1) in many applications, there are notable cases in which the scale is allowed to vary to capture differences in utility variance. For example, different scale parameters are often estimated in pooled data models, such as in studies combining revealed and stated preference (RP-SP) data (Whitehead, Haab, and Huang 2011). More generally, scale heteroskedastic logit models, models in which λ is different over a subset of individuals, have been used to analyze the effects of complexity on utility variance discussed above (e.g., Dellaert et al. 1999; Swait and Adamowicz 2001; DeShazo and Fermo 2002), as well as to control for scale heterogeneity associated with different types of stated ANA behavior (Campbell, Hutchinson, and Scarpa 2008). In the complexity studies, scale was typically a function of choice design characteristics and individual-level characteristics. Note that in studies using the scale heteroskedastic logit model generally, scale is only identifiable relative to a constant, such that the scale parameter associated with one subset of individuals is typically normalized to 1 (i.e., λ = 1). [3]

The deterministic portion of conditional indirect utility (V_{jt}) specifications in most CE studies takes a linear-in-parameters form,

$$V_{jt} = \boldsymbol{\beta}' \cdot \boldsymbol{x}_{jt} \tag{2}$$

where $\boldsymbol{x}_{jt}$ includes variables that vary across the choice alternatives and CE questions and $\boldsymbol{\beta}$ is a vector of parameters to be estimated. The $\boldsymbol{x}_{jt}$ variables are constructed from the attributes (typically attribute levels or dummy variables representing specific levels of the attributes) and

[3] Note also that scale heterogeneity can be interpreted as a specific form of correlation among utility coefficients, and as such is a specific form of preference heterogeneity (Hess and Train 2017).

may also contain alternative specific constants (ASCs) and interaction terms with individual-specific variables interacted with the attribute-related variables or with the ASCs.

A common variant of the CL, the mixed logit (MXL) model, is frequently turned to as a way to account for preference heterogeneity and allow more flexibility in the estimated probabilities, thus avoiding the well-known independence from irrelevant alternatives (IIA) property of CL models (Train 2009). The MXL model recognizes that preference parameters may vary across individuals in the sample. To this end, the deterministic part of the conditional indirect utility function in the MXL is modeled for the *n*th person as

$$V_{njt} = \boldsymbol{\beta_n}' \cdot \boldsymbol{x}_{njt}. \tag{3}$$

The individual's preference parameters ($\boldsymbol{\beta_n}$) are assumed to follow a distribution, $f(\boldsymbol{\beta_n})$ and are typically based on the normal, lognormal, or triangular distributions. For any individual (dropping the individual index n for convenience), and assuming scale is unity ($\lambda = 1$), the probability of choosing the jth alternative in the tth choice question in the mixed logit model is evaluated over the distribution of the preference parameter distribution:

$$\mathrm{P}_{jt} = \int \exp(V_{jt}(\boldsymbol{\beta})/\Sigma_k\, [\exp(V_{kt}(\boldsymbol{\beta}))]\, f(\boldsymbol{\beta})\, d\boldsymbol{\beta}. \tag{4}$$

This is approximated by the simulated probability (P_j^s)

$$\mathrm{P}_{jt}^s = (1/R) \cdot \Sigma_r \exp(V_{jt}(\boldsymbol{\beta}^r)/\Sigma_k\, [\exp(V_{kt}(\boldsymbol{\beta}^r))] \quad \forall r \in \{1,\ldots,R\}, \tag{5}$$

where R is the number of draws and $\boldsymbol{\beta}^r$ are from the distribution $f(\boldsymbol{\beta})$. Panel MXL models can be constructed by assuming errors are independent across choice questions, resulting in joint probabilities formed by the product of the individual probabilities (equation 5) over the T questions.

Stated, Inferred, and Visual ANA Approaches

As noted earlier, the literature includes three basic approaches for identifying and accounting for ANA behavior in choice experiments. The first, the stated ANA approach, employs auxiliary questions asked in conjunction with the CE questions that provide self-reported information about ANA behavior while answering the CE questions. The second, the inferred ANA approach, relies on the use of more flexible econometric models to allow identification of ANA behavior. The third, the visual ANA approach, uses information about respondents' attention to individual attributes inferred from eye movement data to identify ANA behavior. This section provides more details about these approaches and how the data and methods have evolved in the literature with the goal of providing the reader a solid overview and understanding of them. Figures 1 and 2 provide visual aids to our stated and inferred ANA model typology, respectively (italicized terms in the text refer to the Figures).

Stated ANA Approaches

Stated ANA information can be measured in a variety of ways. Stated ANA questions ask respondents to indicate whether or not they considered or did not consider each attribute in the CE questions. Sometimes, responses to these questions are posed as binary (ignore/did not ignore or considered/did not consider) (e.g., Hensher 2006; Hensher and Rose 2009). Other times, respondents are able to indicate the degree to which they ignored each attribute over the sequence of CE questions, such as on an ordinal scale from "never" to "always" (e.g., Scarpa et al. 2013; Weller et al. 2014). Regardless of the form, these questions are generally asked after either every CE question (choice task ANA) or only after the last CE question (serial ANA). While the early studies (e.g., Rose, Hensher, and Greene 2005; Hensher, Rose, and Greene 2005) relied on serial ANA questions, it was recognized fairly early on that individuals may apply different ANA choice heuristics across different questions (Puckett and Hensher 2008; Scarpa, Thiene, and Hensher 2010).

The *attribute elimination model* introduced by Hensher, Rose, and Greene (2005)--and described earlier--assumes that the stated ANA information is exogenous and conditions the indirect utility function for the jth choice alternative (Equation 2) by each respondent's responses to the stated ANA questions. Suppose for any individual that $\mathbf{a}_t$ is a K×1 vector of dummy variables with elements that take the value of 1 when the respondent, for the tth choice question, indicates the kth attribute ($k = 1,\ldots,K$) is attended to and 0 when it is ignored. Then, the deterministic component of conditional indirect utility associated with the jth choice in the tth choice question that enters the choice model is

$$V_{jt} = (\mathbf{a}_t \cdot \boldsymbol{\beta})' \cdot \boldsymbol{x}_{jt}. \tag{6}$$

The vector $\mathbf{a}_t$ can be determined directly from the stated ANA responses if those questions are posed as binary (ignore/did not ignore attribute), but for ordinal responses (e.g., Colombo, Christie, and Hanley 2013) the researcher must determine a threshold level that would indicate ANA behavior and construct $\mathbf{a}_t$. Note that equation 6 allows for choice task stated ANA information, but serial stated ANA information means respondents ignored or attended to each attribute in the same way across the choice questions; i.e., $\mathbf{a}_1 = \mathbf{a}_2 = \ldots = \mathbf{a}_T$.

The attribute elimination model assumes a zero marginal utility for the attributes that a respondent indicates she ignores. Hess and Hensher (2010) challenged this notion, suggesting that respondents may in fact have a non-zero marginal utility for the attributes they stated they ignored. This could be because the stated ANA was derived from a serial ANA question, and the respondent did pay attention to the attribute in at least some of the choice questions.

Several issues have been raised with the attribute elimination model. There is a potential for endogeneity to be introduced by conditioning utility parameters on stated ANA information that itself is a realization of a choice process—the information processing stage.[4] Additionally, the accuracy of the self-reported assessment of ANA behavior itself may be questionable due to concerns about recall bias, cognitive constraints, and potential strategic or other behavioral biases often levied against stated preference information. Hess and Hensher (2010) note that it is possible that individuals saying they ignored an attribute may have just placed less importance on it and therefore its marginal utility really should be non-zero. Their solution is to estimate separate preference parameters for those stating they ignore the attribute from those who stated they attend to it. Therefore, the deterministic portion of conditional indirect utility takes the following form:

$$V_{jt} = (\mathbf{a}_t \cdot \boldsymbol{\beta}_a + (\mathbf{I}_K - \mathbf{a}_t) \cdot \boldsymbol{\beta}_{\sim a})' \cdot \boldsymbol{x}_{jt}, \tag{7}$$

where $\mathbf{I}_K$ is a K×1 vector of ones, $\boldsymbol{\beta}_a$ is the K×1 parameter vector for those who attended to the attributes, and $\boldsymbol{\beta}_{\sim a}$ is the parameter vector for those who ignored the attributes. Using CE data from a CAPI survey about transportation choices, Hess and Hensher (2010) found that parameters associated with those ignoring the attributes (i.e., $\boldsymbol{\beta}_{\sim a}$) were statistically non-zero instead of zero as is assumed under the attribute elimination model and generally of a lower magnitude than the parameters associated with those attending to the attributes (i.e., $\boldsymbol{\beta}_a$). Moreover, they found that in a mixed logit model with this utility specification there was significant preference heterogeneity for the "ignore" standard deviation parameters. Hess and Hensher's model has been termed the *ANA validation model* and used in studies by Scarpa et al. (2013), Alemu et al. (2013), and Caputo et al. (2018), among others. Extensions of the ANA

[4] Hensher (2008) addresses this issue by modeling the processing decision with the related CE choice stage.

validation modeling approach include models allowing for different "ignore" parameters depending upon responses to supplemental questions about the reasons why the respondent ignored each attribute that was ignored (Alemu et al. 2013) and for different ordinal responses to stated ANA questions with possible responses of never, sometimes, and always considered the attribute (Colombo, Christie, and Hanley 2013).

Balcombe, Burton, and Rigby (2011) introduced a Bayesian mixed logit model that accounted for stated ANA information. Their model allowed for incorporation of stated ANA information through monotonic *transformations of the random parameters* such that the *n*th person's marginal utility for the *k*th attribute depends on her response to the stated ANA question (a_{nk}), such that $\beta_{nk} = \alpha_0 + \alpha_1 \cdot a_{nk} + u_{nk}$, where α_0 and α_1 are parameters and u_{nk} is a disturbance term.[5] They estimated Bayesian mixed logit models that imposed $\alpha_1 = 0$ (full attribute attendance), α_0 and α_1 are freely estimated (analogous to the ANA validation model), and $\alpha_0 = -\alpha_1$ (analogous to the attribute elimination model but allowing for random error).[6]

Kelbacher, Balcombe, and Bennett (2015) apply a different Bayesian MXL model specification that conditions utility by applying a *shrinkage factor*, ρ, to the transformation of the random parameters, such that $\beta_{nk} = \rho \cdot g(\alpha_n)$ when $a_{nk} = 1$ (respondent *n* indicates ignoring *k*th attribute) and $\beta_{nk} = g(\alpha_n)$ when $a_{nk} = 0$ (respondent *n* indicates attending to the *k*th attribute), where $g(\cdot)$ is a transformation function and ρ is bounded to the unit interval, [0,1]. In this specification, the shrinkage factor reduces the magnitude of the marginal attribute utility for attribute non-attenders and is assumed to apply to every attribute equally. It does not impose zero marginal attribute utility. This specification is analogous to the ANA validation model and

[5] These types of utility parameter transformations are not uncommon (e.g., Train and Sonnier 2005).
[6] They also explored a transformation that accounted for the potential skewness of random parameter distributions that could be caused, at least in part, by the presence of ANA.

imposes that attribute non-attenders have a smaller marginal utility than those who attend to the attributes. It was applied in a U.K. study of consumer meat and animal welfare preferences with serial stated ANA data. In contrast to the Balcombe, Burton, and Rigby (2011) model in which the stated ANA enters the parameter transformation function, the shrinkage factor approach conditions the entire transformation function, thus affecting both mean and standard deviation (variance) parameter estimates. Balcombe, Fraser, and McSorley (2015) also use this specification in a study involving visual ANA data. Mohanty et al. (2019) further extend this approach by allowing the shrinkage factor to differ across attributes and found variation in shrinkage factors across attributes in a study of consumer preferences for aromatic rice in India.

The stated ANA models discussed thus far condition the utility specification on the responses to the stated ANA questions while assuming that they can be treated as exogenous variables. As touched on earlier, however, stated ANA responses are the outcomes of a decision-making process and therefore are endogenous. Hess and Hensher (2013) proposed a *hybrid stated ANA model* that treats responses to stated ANA questions as dependent variables and not as explanatory ones to avoid the endogeneity bias caused by the correlation between the stated ANA responses ($\mathbf{a}_n$) and the CE utility error (ε_{nj}). This hybrid stated ANA model takes the form of a joint process-outcome model where the attribute processing stage results in stated ANA information ($\mathbf{a}_n$) and the outcome stage is the CE choices made conditional on $\mathbf{a}_n$. Hess and Hensher (2013) proposed modeling the first stage model as a binary logit, where the latent variable—which they associate with attribute importance—for the nth individual and kth attribute that is indicated in the data by a_{nk} is denoted a_{nk}^* and is a function of characteristics of the individual ($\mathbf{z}_n$) and a random disturbance term e_{nk} that is standard normally distributed; i.e.,

$$\mathrm{a}_{nk}^* = \zeta_k \cdot \mathbf{z}_n + e_{nk}, \qquad (8)$$

where ζ_k are parameters that relate how the individual characteristics affect the kth latent variable.

Letting γ_{0k} and γ_{1k} be parameters to be estimated, the probability of stated non-attendance to the kth attribute by the nth individual, $a_{nk} = 0$, is

$$\Pr(a_{nk} = 0) = \exp(\gamma_{0k} + \gamma_{1k} \cdot a_{nk}^*)/[1 + \exp(\gamma_{0k} + \gamma_{1k} \cdot a_{nk}^*)]. \tag{9}$$

The probability of stated attendance, $a_{nk} = 1$, is

$$\Pr(a_{nk} = 1) = [1 + \exp(\gamma_{0k} + \gamma_{1k} \cdot a_{nk}^*)]^{-1}. \tag{10}$$

Assuming independent errors, the joint probability of observing the nth individual's stated ANA responses across the K attributes can be constructed from equations 9 and 10 and is denoted L_{n1}:

$$L_{n1} = \prod_{k=1}^{K} \Pr(a_{nk} = 0)^{1-a_{nk}} \cdot \Pr(a_{nk} = 1)^{a_{nk}}. \tag{11}$$

The outcome stage of the hybrid model involves modeling the CE responses. The latent variables are included in this stage as shrinkage factors inside the choice model (assumed to be a MXL model) by replacing the utility parameters $\boldsymbol{\beta}_n$ with $\exp(\boldsymbol{\lambda} \cdot \mathbf{a}_n^*) \cdot \boldsymbol{\beta}_n$. Assuming a linear-in-parameters utility specification, they note that the probability of observing the jth alternative in the tth CE question conditional on the latent variables ($\mathbf{a}_n^*$) is

$$\Pr(\text{choose } j \text{ in question } t | \mathbf{a}_n^*) = P_{jnt|a_n^*} = \frac{\exp(\sum_{k=1}^{K} \exp(\lambda_k \cdot a_{nk}^*) \cdot \beta_{nk} \cdot x_{jntk})}{\sum_{i=1}^{J} \exp(\sum_{k=1}^{K} \exp(\lambda_k \cdot a_{nk}^*) \cdot \ _{nk} \cdot x_{intk})}. \tag{12}$$

This model has two random components, arising from the latent variable ($\mathbf{a}_n^*$) and the preference parameters ($\boldsymbol{\beta}$). Thus, the joint probability is integrated over the distributions of $\mathbf{a}_n^*$ and $\boldsymbol{\beta}$, such that the individual's likelihood function can be written as

$$L_n = \int_{a_n^*} \int_{\ n} \left[\prod_{t=1}^{T} P_{jnt|a_n^*}\right] \cdot L_{n1} \cdot h(\beta | b, \Omega) \cdot g(a_n^* | \zeta, \mathbf{z}_n) d\beta d a_n^*, \tag{13}$$

where $g(\cdot)$ is the standard normal distribution and $h(\cdot)$ is the density function appropriate for the assumed random parameter distributions. Hess and Hensher (2013) applied this model to a CE study of car-driving non-commuters and their preferences for different transit options. While their model also jointly estimated responses to other survey questions informing the attribute importance (in the form of attribute rankings), the hybrid model CE performed reasonably well statistically relative to a MXL model that did not account for the stated ANA information and relative to a stated ANA validation model. Bello and Abdulai (2016) apply the hybrid model approach in a CE study of consumer preferences for organic products in Nigeria. Chalak, Abiad, and Balcombe (2016) also apply the hybrid approach but in a Bayesian estimation model that, like Hess and Hensher (2013), also accounts for attribute ranking data as a second source of information about the latent attribute importance.

Inferred ANA Approaches

Inferred ANA approaches generally involve applying flexible econometric models that allow for ANA to be identified directly from the CE data. In this way, they do not require the use of stated ANA information or other supplemental information provided by individuals. Rather, they attempt to let the CE data speak for itself about whether or not ANA behavior is present. Two of the inferred ANA modeling approaches discussed below, the *equality-constrained latent class ANA model* and the *endogenous attribute attendance model*, are based on joint process-outcome model frameworks.[7] A third approach is the *Bayesian stochastic attribute selection model* and its variants. Finally, a fourth approach utilizes the fact that the researcher can generate individual-level conditional distributions from MXL choice models and

[7] Another joint process-outcome model suggested by Hensher and Rose (2009) estimates an additional parameter for each attribute to capture sample-level ANA behavior. These parameters determine the probability of ANA, are assumed exponentially distributed, and act analogously to a shrinkage factor at the sample-level. In their transportation choice application, they were estimated jointly with the preference parameters in a conditional logit-based model.

then use them to infer ANA behavior through an examination of the mean and variance of those conditional distributions.[8]

Latent class logit (LCL) models are often employed when preference heterogeneity across the sample is believed to be best captured by discrete classes of individuals in the sample (segments) who share the same preference structure. As noted above, Scarpa et al. (2009) were first to propose using a special form of the LCL modeling approach to identify segments of the sample that pay attention to the same set of attributes (and thus ignore the same ones too).[9] A single parameter vector $\boldsymbol{\beta}$ is estimated in the model,[10] and each latent class is differentiated by which parameters in $\boldsymbol{\beta}$, if any, are ignored and thus assumed to be zero. Since the parameters that are attended to are assumed to be the same across classes, these models are generally referred to as equality-constrained latent class logit ANA models.

Let $\boldsymbol{\beta_q}$ be the K×1 parameter vector associated with class q (q = 1, 2,…,Q), then the unconditional probability of observing the individual choosing alternative j is

$$\Pr[\text{choose } j] = P_j = \sum_{q=1}^{Q}\left[\left(\frac{\exp(\theta_q)}{\sum_{l=1}^{Q}\exp(\theta_l)}\right)\times\left(\frac{\exp(V_j(\boldsymbol{\beta_q}))}{\sum_k exp(V_k(\boldsymbol{\beta_q}))}\right)\right], \tag{14}$$

$$\text{where } \sum_{q=1}^{Q}\frac{\exp(\theta_q)}{\sum_{l=1}^{Q}\exp(\theta_l)} = 1,$$

for all $q \in Q$ and for all k and $j \in \{1,2,\ldots,J\}$. The left-most term in the right-hand side equation is the probability of membership in latent class q, where θ_q ($q = 1,2,\ldots,Q$) is a class-specific constant parameter to be estimated.[11] θ_q can also be specified as a function that varies over

[8] We should acknowledge that there are some studies that attempt to infer ANA behavior by examining patterns of responses to particular CE questions constructed with specific choice alternative and attribute combinations (e.g., Hensher 2006b; Saelensminde 2009; Hensher and Collins 2011).
[9] A conference presentation by Hess and Rose (2007) may have served as motivation for this approach. In it, latent class logit models were suggested as a potentially useful approach for identifying ANA behavior.
[10] Hensher and Greene (2010) and Campbell, Hensher, and Scarpa (2011) use a LCL model that assigns zeros to the attribute parameters that are ignored within class but do not impose equality of the parameter vector across classes.
[11] For identification purposes in estimation, it is typical to set one class-specific constant equal to zero.

individual- or choice-level characteristics (e.g., Hole, Norman, and Viney 2016). The other term in the right-hand side of the equation is the probability of observing alternative j being selected conditional on being in class q, which is associated with the parameter vector $\boldsymbol{\beta_q}$. It is typical to include a full attendance class for which no parameters are set equal to zero.

Note that the selection of latent classes is up to the researcher. There are 2^K possible combinations of attribute non-attendance patterns, and therefore 2^K possible latent classes. For large experimental designs with many attributes, the researcher may likely run into a dimensionality issue due to the exponential growth in the number of latent classes as K increases that can result in convergence challenges in estimation.[12] Due to this problem, it is rare for researchers to choose a full 2^K ECLC model. Instead, researchers typically choose $Q < 2^K$. An extreme example is Koetse (2017) who estimates ECLC models with ANA solely on the cost parameter.

Somewhat surprisingly, little attention has been paid to the process of selecting the ANA latent classes in the ECLC model when a full 2^k model is not used, with most studies providing scant insight into the specification process they employ. An exception is Lagarde (2013), who proposed a stepwise process for selecting the ANA classes when a full 2^K model is infeasible. The process begins by specifying and estimating an ECLC model with K+1 classes—a full attendance class (no ANA behavior) and K classes that represent ANA behavior related to a single attribute. Using the model results from the K+1 ECLC model, latent classes are retained if they were found to be statistically significant (i.e., its class membership is statistically different from zero), the statistically insignificant ones are dropped, and a set of new classes are added that capture other ANA behaviors. The new ECLC model is then estimated, and the process is repeated until all

[12] For example, the econometric software NLOGIT (version 6) can only handle up to K=4.

possible ANA patterns have been included in at least one model run. The final ECLC model will include all statistically significant latent classes supported by the data.

The ECLC ANA model is the most commonly applied inferred ANA modeling approach used in the empirical literature (Hensher and Greene 2010; Hensher 2010; Hensher, Rose, and Greene 2012; Collins, Rose, and Hensher 2013; Kragt 2013; Lagarde 2013; Scarpa et al. 2013; Weller et al. 2014; Yao et al. 2015; Glenk et al. 2015; Koetse 2017; Jourdain and Vivithkeyoonvong 2017; Caputo et al. 2018; Glenk et al. 2019; Lew 2019; Petrolia and Hwang this issue). For tractability, these models are commonly estimated without preference heterogeneity, but there are a few exceptions that introduce heterogeneity through additional latent classes that differ in preferences (e.g., Caputo et al. 2018) or through a random parameters (mixed logit) specification (e.g., Lew 2019).

An alternative to the ECLC model that avoids the dimensionality issue is the endogenous attribute attendance (EAA) model (Hole 2011). The EAA model assumes a two-step process. In the first step, the individual chooses which attributes will be paid attention to (or ignored) in the choice. The second step involves choosing the alternative that yields the most utility given the underlying preferences conditional on the first step. The dimensionality is reduced by modeling the probabilities of attending to each of the attributes and assuming independence of those probabilities.

The EAA model proceeds as follows: For K attributes there are $Q = 2^K$ possible combinations of attributes, including the case with all K attributes and one where there are no attributes included. Let A_q be the set of attributes to which an individual pays attention in the choice occasion (for $q = 1, 2, \ldots, Q$). The probability of observing an attribute k ($k \in K$) being paid attention to is specified as a logit probability:

$$\Pr[k] = \exp(\theta_k)/[1+\exp(\theta_k)]. \tag{15}$$

Note that the scalar parameter θ_k is an attribute-specific constant that could be replaced by a function of individual characteristics to allow for heterogeneity in attribute attendance behavior.[13]

The probability of observing the specific set of attributes in A_q being used by the individual in the choice setting is the product of the binomial logit probabilities in equation 15:

$$\Pr[A_q] = H_{A_q} = \prod_{k \in A_q} \left(\frac{\exp(\theta_k)}{1+\exp(\theta_k)}\right) \times \prod_{k \notin A_q} \left(\frac{1}{1+\exp(\theta_k)}\right). \tag{16}$$

Note that this assumes independence of the individual attribute attendance probabilities.

Let the deterministic indirect utility of alternative j conditional on the individual paying attention to the A_q set of attributes be $V_j(\boldsymbol{\beta}^{Aq}) = \boldsymbol{\beta}^{Aq} \cdot \boldsymbol{x}_j^{Aq}$, where $\boldsymbol{x}_j^{Aq}$ consists of variables associated with only the attributes included in A_q and $\boldsymbol{\beta}^{Aq}$ be parameters to estimate. Assuming TEV errors and common $\boldsymbol{\beta}^{Aq}$ across respondents, the joint probability of observing the selection of j conditional on A_q is then

$$\Pr[\text{choose } j|A_q] = P_{j|Aq} = \{\exp(V_j(\boldsymbol{\beta}^{Aq}))/\Sigma_m \exp(V_m(\boldsymbol{\beta}^{Aq}))\} \tag{17}$$

for all m and $j \in \{\text{A, B, C}\}$. The unconditional probability is

$$\Pr[\text{choose } j] = P_j = \sum_{q=1}^{Q} H_{A_q} \times P_{j|Aq}. \tag{18}$$

Note that when $K = 1$ (ANA behavior is limited to a single attribute), then the EAA and ECLC models are equivalent. Hole (2011) applies this model to analyze CE data in a healthcare context on preferences for features of primary care consultation.

Note that the EAA model involves estimation of K ANA-related parameters compared to up to $(2^K - 1)$ ANA-related parameters in the full 2^K ECLC model and avoids the potential

[13] For example, Sandorf, Campbell, and Hanley (2017) parameterize equation 15 as a function of scores to a quiz included in the survey and whether or not respondents were informed of their quiz score.

specification bias involved with selection of the ANA behaviors to include in the model. This advantage of the EAA model over the ECLC model arises from the assumption of the ANA probabilities being independent across attributes. In some datasets, this assumption may not hold, though in at least one study comparing the two models, similar ANA patterns were estimated in both models (Hole, Norman, and Viney 2016).[14]

One common feature of the ECLC and EAA models is that they both involve imposing a zero marginal utility on ignored attributes. Hess et al. (2013) refer to these types of models, which have an estimated parameter for those attending to an attribute and another (set to zero by the analyst) for those ignoring the attribute, as *confirmatory* ANA models. They differentiate these models from *exploratory* ANA models that allow for non-zero marginal utilities for attributes that are not attended to by respondents. They argue that confirmatory models may be biased in the presence of preference heterogeneity since individuals who attend to an attribute but who have near-zero marginal utility for it (i.e., who don't care much) being lumped in with true non-attenders. In other words, confirmatory ANA models like the ECLC and EAA models may be confounding preference heterogeneity—specifically from those with low marginal utility for attributes—with actual ANA behavior.

Hess et al. (2013) suggested two approaches for addressing this issue. The first was an analogous model to the ANA validation model applied to the EAA model where instead of assuming non-attenders have zero marginal utility, separate (non-zero) parameters are estimated for attenders and non-attenders. In the EAA model, this means specifying utility as $V_j(\boldsymbol{\beta}^{Aq},\boldsymbol{\beta}^{\sim Aq}|A_q, \sim A_q) = \boldsymbol{\beta}^{Aq} \cdot \boldsymbol{x}_j^{Aq} + \boldsymbol{\beta}^{\sim Aq} \cdot \boldsymbol{x}_j^{\sim Aq}$, where $\sim A_q$ denotes the attributes not in A_q, $\boldsymbol{x}_j^{\sim Aq}$

[14] Collins, Rose, and Hensher (2013) propose a hybrid between the ECLC and EAA models that assumes independence across latent classes, but allows correlation between attribute attendance within classes. They use stated ANA responses as covariates in the class membership determination, similarly to Hole, Kolstad, and Gyrd-Hansen (2013).

consists of variables associated with only the attributes included in $\sim A_q$ (those ignored) and $\sim\boldsymbol{\beta}^{Aq}$ are parameters to estimate. The second approach they suggest is to estimate mixed logit versions of the confirmatory ANA models to capture the preference heterogeneity of attribute attenders and thus reduce the chances that individuals that place less importance--and thus less value--on the attribute may be erroneously identified as a non-attender. They apply these two approaches in three case studies (Hess et al. 2013). In each, they compare the proposed models against the fixed parameter version of the EAA and a MXL model that ignores ANA, finding that both proposed models outperform the confirmatory EAA model and suggest the confirmatory model results may be misleading in terms of measured ANA behavior, given the exploratory models yielded lower levels of ANA behavior than the confirmatory models. MXL-based versions of the EAA model (i.e., the mixed EAA model) were also used by Hole et al. (2013) in a study of doctors' preferences for characteristics of medicine and by Lew (2018) and Lewis et al. (2019) in studies of public preferences for threatened and endangered marine species conservation goals.

An inferred ANA model based on the Bayesian mixed logit model, the STAS model, was developed by Gilbride, Allenby, and Brazell (2006). The STAS model introduces a vector $\boldsymbol{\tau}_n$ to a Bayesian MXL model that has the same dimensions as the parameter vector, $\boldsymbol{\beta}_n$. In the vector, $\tau_{nk} = 1$ if the kth attribute is attended to by the nth individual and a small constant (near zero) denoted by c otherwise. Similar to many of the earlier models discussed above, $\boldsymbol{\tau}_n$ conditions the utility specification to make the marginal utility of non-attended attributes essentially zero. That is, the deterministic portion of conditional indirect utility of the jth alternative for the nth individual is written as $V_{nj} = \boldsymbol{\tau}_n \cdot \boldsymbol{\beta}_n' \cdot \boldsymbol{x}_{nj}$. $\boldsymbol{\tau}_n$ is assumed to follow a binomial distribution. The constant c is selected by the researcher to be sufficiently close to zero with a very small variance to indicate the attribute does not affect choice. In Scarpa et al. (2009), for instance, c was set to

0.01. If $C_{\tau n} = \text{diag}(\boldsymbol{\tau}_n)$, a K×K square matrix with $\boldsymbol{\tau}_n$ on the diagonal, then the parameter heterogeneity for the kth attribute assuming a multivariate normal distribution is $\beta_{nk} \sim \text{N}(C_{\tau nk} \cdot \text{b}_k, C_{\tau nk} \cdot \Omega\, C_{\tau nk})$, where **b** and Ω are the mean vector and variance-covariance matrix, respectively. Heterogeneity in τ_{nk} is introduced by assuming $\tau_{nk} = 1$ (attribute attendance) with probability θ_k and $\tau_{nk} = c$ (ANA) with probability $1\text{-}\theta_k$. The prior distributions of **b** and Ω are assumed normal and inverted Wishart, following the standard Bayesian MXL model (Train and Sonnier 2005). This set-up allows the use of standard Bayesian methods for updating parameters and drawing from posteriors while enabling inferences about ANA behavior. The few applications in which the STAS model has been applied include a study of public preferences for rural environmental land improvements in Ireland (Scarpa et al. 2009) and a Dutch study to understand the general public's preferences for their future health (Jonker et al. 2018).

The three inferred ANA approaches discussed thus far—the ECLC, EAA, and STAS models—either explicitly or implicitly use a joint process-outcome modeling approach where an individual's choices are decomposed into an attribute processing strategy stage that determines the set of attributes that are paid attention to and the choice stage that is dependent upon the attribute processing stage. Note that in each, however, the information processing stage is represented in simple ways. For example, the ECLC and EAA models account for the probabilities of different types of ANA behavior, but do not take a structural perspective on the decision to attend or not attend to each attribute that underlies these ANA behaviors.

To our knowledge, the only study to attempt a more structural approach is one by Cameron and DeShazo (2010). Drawing inspiration from studies by Gabaix and Laibson (2000) and DellaVigna (2009) that form the early foundations of the rational inattention literature (e.g., Gabaix 2017), they develop a "propensity to attend" model that explains how individuals choose how much attention to allocate among attributes. Their theoretical model suggests the factors affecting an individual's decision to pay attention to an attribute are the marginal benefits of paying attention to other attributes (cross-attribute utility), the marginal importance of the attribute relative to the total utility (own-attribute utility), and the marginal opportunity costs of attention to each attribute (which depends upon the cognitive resources available to the individual) (Cameron and DeShazo 2010). They suggest an empirical process-outcome model specification that takes the form of attribute-specific shrinkage factors that are functions of the above factors. However, their exploratory application proved difficult to operationalize given the information needs of the specification, even with a large sample.

The final inferred ANA modeling approach we discuss departs from the joint process-outcome modeling framework. Hess and Hensher (2010) propose a *coefficient of variation* (COV) *threshold* approach that seeks to identify individual-level (serial) ANA behavior through an examination of individual-level conditional parameter distributions. Huber and Train (2001) have shown how MXL model results and an individual's choice patterns over a series of CE questions can be used to determine individual-level conditional parameter distributions. Suppose this is done for the nth individual using results from a MXL model assuming normally-distributed random parameters. Then individual n has a utility parameter for attribute k, β_{nk}, that is distributed $N(b_{nk},\sigma_{nk}^2)$, where b_{nk} is the conditional mean and σ_{nk}^2 is the conditional variance of the distribution. The ratio σ_{nk}/b_{nk} is the COV, a measure of the signal-to-noise ratio on the variability of strength of preference for the kth attribute for the nth individual based on that person's observed behavior. Hess and Hensher (2010) found that in their study that individuals who stated that they ignored an attribute had a lower mean parameter and smaller range (variance) for β_{nk}, and a higher COV, compared to those who did not say they ignored the attribute. They suggested that a large COV—one exceeding some threshold level—indicates the individual-level parameter distribution is overdispersed and that the choice behavior would be consistent with the respondent ignoring the attribute.

The COV threshold ANA approach involves two estimation steps that are done sequentially rather than jointly. In the first, the CE data are estimated with a MXL model. The model results, together with the actual choices made by individuals, are then used to generate individual-level conditional parameter distributions following Huber and Train (2001). The conditional parameter distributions are used for identifying—at an individual level—which attributes are ignored or not by comparing the COV for each conditional parameter distribution with a threshold value. If the COV exceeds the threshold value, then the individual is assumed to ignore that attribute. In the second estimation step, the individual's utility specification is conditioned to reflect the implied ANA behavior and then the choice model is estimated again. Note that the second step estimation model could condition the utility similarly to the attribute elimination model or ANA validation model approaches (as done by Hess and Hensher 2010).

For their study, Hess and Hensher (2010) used a threshold value of two, but admitted that it was somewhat arbitrary. Most studies that have employed this approach use the two-threshold value (e.g., Scarpa et al. 2013), but Mariel, Hoyos, and Meyerhoff (2013) use a Monte Carlo approach to examine the concurrence between true (simulated) serial ANA and different threshold values in the COV threshold inferred ANA modeling approach, finding evidence that the ability to correctly identify ANA with a specific threshold value using the COV threshold model approach is sensitive to the amount of ANA present.

A key limitation of this approach, however, is that it does not distinguish ANA behavior from low importance or value for an attribute. Both would be represented in the first step MXL model by a mean utility parameter that is close to zero. Thus, this approach can be viewed as a "partial detection" approach to identifying ANA behavior.

Visual ANA approaches

Eye tracking (ET), or eye movement, research has a long history as a tool for studying decision-making in marketing (Wedel and Pieters 2008) and psychology (Glaholt and Reingold 2011) and has recently gained attention in economics (Lahey and Oxley 2016). In a review of ET studies of decision making, Orquin and Muller Loose (2013) find evidence that visual attention processes play an active role in constructing decisions. Subsequently, several studies began exploring the use of eye movement data collected from respondents while answering CE questions to gain insights about the role attention, specifically visual attention to attributes, plays on choices. Eye movement CE studies explicitly focused on visual attention as an indicator of ANA behavior include Balcombe, Fraser, and McSorley (2015), Krucien, Ryan, and Hermes (2017), Chavez, Palma, and Collart (2018), and van Loo et al. (2018). Other ET-based CE studies not focusing on ANA have examined the influence of visual attention on marginal attribute values (van Loo et al. 2015; Balcombe et al. 2017), the relationship between visual attention and choice certainty (Uggeldahl et al. 2016), and the identification of visual processing biases attributable to CE survey design features (Ryan, Krucien, and Hermens 2017).

In these studies, ET equipment is used in controlled lab settings to record eye movements from an individual while she answers CE questions on a video monitor. Specifically, two types of eye movements are of interest—fixations (moments when the eye is relatively still) and saccades (rapid eye movements that serve to project specific objects within view into the individual's fovea, the central and most sensitive part of the retina, opposite the eye's lens)

(Wedel and Pieters 2008). Each CE question is divided into specific visual areas, called regions of interest (ROIs), and saccades and fixations are spatially and temporally recorded and mapped over the ROIs while the individual answers the survey.

Visual attention in these studies is measured through metrics related to whether and how much an individual looks at the ROIs in the CE questions. For visual ANA studies, the ROIs are defined over the attribute information in each CE question. The number and duration of eye fixations—fixation counts (FC) and fixation time (FT)—are common measures of visual attention, given the assumption that attribute processing occurs during fixations. Measures of visual attention based on FT were used in Krucien, Ryan, and Hermens (2017), Balcombe et al. (2017), and Chavez, Palma, and Collart (2018), while ones based on FC were used by Balcombe, Fraser, and McSorley (2015) and van Loo et al. (2018).

How visual attention is used in an individual's attribute processing strategy, and thus whether and how visual ANA translates to ANA in choice behavior, is an unsettled question. Visual ANA was defined by Balcombe, Fraser, and McSorley (2015) as some or all attribute information being visually ignored. If there is full visual ANA to an attribute—that is, the individual does not look at the attribute information at all—it is unlikely she will use the information in deciding between choice alternatives in a given CE question. In this case, no visual attention to an attribute is likely a good indicator of ANA. Visual ANA studies have tended to differ in what they define as no visual attention.

Balcombe, Fraser, and McSorley (2015) assumed a threshold of at least two fixations (lasting at least 200 milliseconds) on an attribute's information in each CE question to qualify as visually attending to the attribute. Further, they take a serial approach to identification of visual ANA by classifying individuals who do not visually attend to the majority of CE questions as visual non-attenders, while those who visually attend to the attribute in a majority of the choice tasks are classified as visual attenders. They incorporate the information on visual attribute attendance into a Bayesian MXL choice model using a shrinkage factor approach similar to the stated ANA approach outlined in Kehlbacher, Balcombe, and Bennett (2013) but allowing for both visual and stated ANA to contribute to the shrinkage factor applied to each attribute's utility parameter. In the study, they only found a weak association between the visual ANA metrics (total eye fixations) and stated ANA responses to a serial ANA question included in the survey.

In contrast to the serial visual ANA approach taken by Balcombe, Fraser, and McSorley (2015), van Loo et al. (2018) proposed identifying visual ANA at the choice task (CE question) level. They go one step further by determining visual ANA separately for each choice alternative (rather than at the choice task level) and varying the FC threshold, estimating models using either one or two fixations as the threshold. Visual non-attendance under the various definitions (serial or choice task visual ANA, one or two fixation counts to identify visual ANA, and evaluation at the choice alternative or choice task level) was incorporated into several models. The estimation models included an error-components MXL model assuming full attribute attendance and two stated ANA models—a MXL version of the attribute elimination model (Hensher, Rose, and Greene 2005) and a CL version of the ANA validation model (Hess and Hensher 2010). In these ANA models, the visual ANA information is used in place of stated ANA data. Additionally, they compare the visual ANA behavior of each individual against the inferred ANA behavior resulting from the COV threshold inferred ANA model (Hess and Hensher 2010) using a threshold value for the COV of two.

Unsurprisingly, van Loo et al. (2018) found non-trivial differences in model outcomes under different visual ANA definitions. They found better-fitting models when using definitions of visual ANA with two fixations as the threshold, evaluation at the whole choice task level rather than at the choice alternative level, and assuming serial visual ANA rather than choice task visual ANA. In their application, the ANA validation model was preferred over the attribute elimination model and indicated that utility parameters for visually ignored attributes were smaller in magnitude, but generally not zero, compared to those that were visually attended. Further, a comparison of visually ignored attributes with the results from the COV threshold model suggest that not all visually ignored attributes were ignored in choice behavior. Their results also suggest visual ANA may differ across attributes, with the cost attribute in particular being one that may not need to be visually attended to for it to be used in choice behavior.

Other studies have used fixation time as an indicator of visual ANA. Chavez, Palma, and Collart (2018) suggest using a FT measure that is weighted over the total time spent on each CE question to indicate visual attention and to control for potential learning and fatigue effects over the repeated choice tasks. An attribute elimination model is applied to the data assuming visual non-attenders to an attribute spend less than 10% of the total time spent on a choice set visually looking at the attribute information. They also compared their classification of visual non-attenders to those resulting from an EAA model (ignoring visual attendance) and find evidence of both inferred and visual ANA across the three attributes in their study, though the amount of ANA differed across the approaches considerably. Krucien, Ryan, and Hermens (2017) also use a FT measure to identify visual ANA—an individual is a visual non-attender if her fixation time on the attribute is less than some arbitrary threshold duration. They used threshold durations ranging from 0 to 2,000 milliseconds to define visual non-attenders in separate attribute elimination models that account for heteroskedastic scale across survey designs. Model fits and the proportion of visual non-attenders were found to be sensitive to the choice of threshold.

The sensitivity of model results to the thresholds chosen to identify visual non-attenders highlights a difficulty of visual ANA studies (Krucien, Ryan, and Hermens 2017; van Loo et al. 2018). Any threshold approach for distinguishing visual ANA behavior could potentially misidentify the true proportion of attributes considered (or ignored) by a given individual. Further, the available evidence does not support a strong concordance of visual ANA behavior with stated ANA (Balcombe, Fraser, and McSorley 2015) or inferred ANA (Chavez, Palma, and Collart 2018; van Loo et al. 2018) behavior. At the same time, other studies have shown that for at least some attributes, the amount of time spent visually paying attention to an attribute's information can explain its importance or utility (Balcombe et al. 2017; van Loo et al. 2015).

Krucien, Ryan, and Hermens (2017) question the use of any particular ET-based definition for defining visual ANA. Instead, they argue that visual attribute attention information should be treated as an indicator of a latent variable that enters the utility specification directly. The amount of visual attention given to an attribute likely contributes to attribute attendance but is not fully deterministic. There could be idiosyncratic errors or measurement errors or both. Due to potential individual-level heterogeneity in the effect visual attendance has on ANA, any visual threshold chosen—based on number of fixations or total duration—could potentially misidentify the true proportion of attributes considered or ignored by an individual. Recognizing this, Krucien, Ryan, and Hermens (2017) propose an approach paralleling the stated ANA hybrid choice model of Hess and Hensher (2013). Their model, called the latent information processing (LIP) model, uses the individual's total fixation times to each attribute as factors affecting the latent information (a stochastic function) that enters the utility function. However, in their application to consumer preferences of health and lifestyle programs in the U.K., the LIP model provided a poorer model fit than the attribute elimination models that used threshold-determined classifications of visual attribute attenders and non-attenders. Of course, this does not invalidate their concerns about the treatment of visual attention data as deterministically signaling ANA behavior, but is indicative that additional work in this area is needed.

Welfare Estimation Issues

In CE models, two types of welfare estimates are commonly calculated—marginal willingness to pay (MWTP) and WTP associated with non-marginal changes in attributes. We begin by examining welfare estimation in the standard (full attendance) CE model.[15] Let $V(\boldsymbol{\beta} \mid \boldsymbol{x})$ be the deterministic component of utility with estimated parameters $\boldsymbol{\beta}$ conditional on attributes $\boldsymbol{x}$. In the CL model, assuming $V(\boldsymbol{\beta} \mid \boldsymbol{x})$ is a linear function in attributes (as in equation 2), the marginal utility of the kth attribute equals the parameter on the attribute, β_k. Thus,

MWTP for the kth attribute $(MWTP_k) = -\partial V(\boldsymbol{\beta} \mid \boldsymbol{x})/\partial x_k / \partial V(\boldsymbol{\beta} \mid \boldsymbol{x})/\partial cost = -\beta_k/\beta_{cost}$, (19)

where β_{cost} is the parameter on the cost attribute that represents the marginal utility of money. In the CL model, this is the average MWTP for the kth attribute over the sample. In the MXL model, there are individual-level parameters (β_{nk}) suggesting MWTP over the sample must be evaluated over the distribution of the parameters.

In CE models, WTP for non-marginal changes in attributes are often of interest. Suppose we are interested in a change in attributes from $\boldsymbol{x}^0$ to $\boldsymbol{x}^1$, where $\boldsymbol{x}^0$ and $\boldsymbol{x}^1$ are K×1 attribute vectors. $\boldsymbol{x}^0$ represents the original (status quo) attribute levels, while $\boldsymbol{x}^1$ represents the attribute levels in the changed state. Calculating WTP involves computing the change in utility resulting

[15] See Sonnier and Train (2005) or Train (2009) for a derivation of welfare estimates in the Bayesian mixed logit model, which involve drawing from the estimated conditional posterior distribution.

from a change in the attribute levels. Assuming a linear deterministic component of utility as before, WTP associated with the change from $\boldsymbol{x}^0$ to $\boldsymbol{x}^1$ is:

$$\begin{aligned}\mathrm{WTP}(\boldsymbol{x}^0 \rightarrow \boldsymbol{x}^1) \equiv \mathrm{WTP}_{\mathbf{x}0,\mathbf{x}1} &= (\partial V(\boldsymbol{\beta} \,|\, \boldsymbol{x}_{\mathrm{j}})/\partial cost)\cdot\{V(\boldsymbol{\beta} \,|\, \boldsymbol{x}^1) - V(\boldsymbol{\beta} \,|\, \boldsymbol{x}^0)\} \\ &= (-1/\beta_{\mathrm{cost}})\cdot\{\boldsymbol{\beta}'\cdot(\boldsymbol{x}^1 - \boldsymbol{x}^0)\},\end{aligned} \tag{20}$$

for the CL model. Similarly to the MWTP calculation for MXL models, we can calculate the WTP over the parameter distributions in the MXL model. These standard model welfare estimates are the means over the sample of respondents.

In the presence of ANA behavior, welfare estimation becomes more complex. For individuals who ignore the cost attribute, the cost parameter is set to zero and no welfare estimates can be calculated. Likewise, for individuals who ignore a non-cost attribute, MWTP calculations cannot be done for that attribute. Below we discuss some welfare estimation issues in the most commonly used stated ANA models (attribute elimination model and ANA validation model) and inferred ANA models (ECLC and EAA models).

A key implicit assumption underlying the attribute elimination model is that there is a common set of utility parameters for all respondents in the sample. While the parameter associated with an ignored attribute is set to zero, the other non-ignored parameters are assumed the same across respondents. Three basic approaches have been used in the stated ANA literature to calculate welfare change associated with a single attribute. Most applications use the sample-level model estimates from the attribute elimination ANA models to derive welfare estimates (e.g., Hensher, Rose, and Greene 2005; Puckett and Hensher 2008; Campbell, Hutchinson, and Scarpa 2008), thus implicitly ignoring those individuals with welfare estimates that cannot be estimated due to ANA and assuming everyone has the same underlying preference parameters, and hence WTP.

Carlsson, Kataria, and Lampi (2010) offered an alternative approach that weights WTP associated with a change in a single attribute by the proportion of respondents for which valid welfare estimates can be calculated for the attribute as determined by binary (ignored/didn't ignore) stated ANA responses. They assessed weighting WTP for two definitions of valid respondents: (i) those who paid attention to the non-cost attribute (but who may have ignored the cost attribute) and (ii) those who indicated paying attention to both the non-cost attribute and the cost attribute. In other words, these weighted average (or conditional) WTP calculations only count the WTP of the subsample of respondents who paid attention to the attribute (the conditional sample). Since there are the fewest individuals in the latter class, WTP calculated using that definition is smaller than for the former class, which is also smaller than the full sample by definition. Colombo, Christie, and Hanley (2013) extend this weighted WTP approach to serial stated ANA data with possible responses to the attendance of attributes across the choice questions of never, sometimes, and always.

A third approach used to generate welfare estimates from the attribute elimination model is to calculate WTP at the individual level using the Bayesian conditional parameter distributions that can be derived using mixed logit-based model results, the observed choices, and stated ANA responses (e.g., Rose et al. 2012).

Similar options for welfare estimation with the ANA validation model exist. Since the model results in two sets of parameter vectors—one for those attending to the attribute and one of those ignoring it ($\boldsymbol{\beta}_a$ and $\boldsymbol{\beta}_{\sim a}$, respectively, in equation 7)--the unconditional welfare for an attribute change can be calculated as a weighted average over the two types of individuals (those who attend and those who don't attend to the attribute) (e.g., Colombo, Christie, and Hanley 2013). Alternatively, one can base WTP only on the estimated parameters for those attending to them, $\boldsymbol{\beta}_a$ (e.g., van Loo et al. 2018), which assumes either that (i) all respondents share the same preference parameters if WTP is not weighted to reflect only those that attend to attributes, or (ii) that WTP is zero for those ignoring attributes if the WTP is weighted by the proportion of attribute attenders. WTP estimation based on individual-level conditional parameter distributions are also possible, but to our knowledge has yet to be implemented.

Note that for both the attribute elimination and ANA validation models, welfare estimates associated with a simultaneous change in multiple attributes (equation 20) complicate unconditional WTP (weighting) calculations. The weighted WTP approach described above can be extended to the case of multiple attribute changes as follows: First, identify all combinations of attributes that respondents in the sample indicated they ignored in the stated ANA questions. Second, calculate the change in an individual's WTP implied for each of these ANA combinations. For respondents who ignore the cost attribute, no welfare estimates can be calculated in the attribute elimination model. For respondents ignoring one or more non-cost attributes, these calculations assume their MWTP for the ignored attributes are zero but the welfare contributions of the other attributes are non-zero. Third, weight the conditional WTP calculated in the second step by the proportions of the sample displaying the ANA behavior and sum over the weighted conditional WTP estimates.

Analogous to the attribute elimination model, the ECLC and EAA models assume a common set of utility parameters. For the ECLC, the two most common welfare estimation strategies are to (1) calculate WTP from the common set of utility parameter estimates without adjusting for individual-level ANA behavior that is represented by sorting into the latent classes (e.g., Scarpa et al. 2009; Collins, Rose, and Hensher 2013; Lagarde 2013; Jourdain and Vivithkeyoonvong 2017) or to (2) weight WTP by the proportion of the sample attending to the attributes (latent class membership probabilities) following a procedure similar to the 3-step approach described above (e.g., Kragt 2013; Scarpa et al. 2013; Glenk et al. 2015). The former strategy ignores the fact that some proportion of the sample ignores attributes, while the latter strategy accounts for ANA behavior assuming zero WTP for attribute non-attenders. For the EAA model, these two strategies translate to the following: (1) calculating welfare from the EAA model estimates without any adjustments (e.g., Hole 2011; Lew 2018) and (2) scale the WTP calculated directly from the model estimates downward by the probability of cost ANA (e.g., Hess et al. 2013), thus only counting the proportion of the sample who (probabilistically) pay attention to cost while assuming those (probabilistically) ignoring cost have a zero WTP. Sandorf, Campbell, and Hanley (2017) calculate individual-specific WTP from the conditional parameter distributions, excluding welfare estimates for attributes that were ignored.

Future Directions and Closing Thoughts

Johnston et al. (2017) include ANA in a list of "response anomalies" that should be evaluated and, if possible, controlled for in the design of CE surveys and analysis of CE data. In this review, we have argued that ANA behavior is consistent with a rationally-adaptive choice model where an individual's decisions are affected by the choice setting, her cognitive abilities and goals, and her evaluation of the opportunity costs and benefits of information search and evaluation. Both self-reported and biometric information on ANA behavior—and that inferred from models that are flexible enough to identify this behavior—were reviewed, as were the methods used to identify and account for ANA behavior. The literature reviewed consistently finds at least some level of ANA behavior in CE applications across a diversity of ANA approaches and empirical applications, reinforcing the recommendation by Johnston et al. (2017)

that "when prior research or pretesting indicates that undesirable response anomalies may be influential, data analysis should investigate these anomalies to determine whether they significantly affect SP responses" (p. 362). Given the weight of evidence of the prevalence of ANA behavior in CE studies and the potentially large impact it can have on model estimates and welfare, assessment of ANA in all CE studies seems prudent as a standard part of any CE study.

Unlike stated and visual ANA approaches, inferred ANA approaches do not require any supplemental information be collected by researchers and therefore offer a relatively low-cost way of evaluating the presence of ANA behavior. As a result, inferred ANA approaches can be used in any CE study. Our discussion of the inferred ANA modeling approaches suggest that confirmatory ANA models that impose the marginal utility of ignored attributes is zero tend to be overly restrictive given exploratory ANA models generally indicate a lower but non-zero marginal utility for ignored attributes. It is for this reason that we recommend the use of exploratory ANA models as a general rule for estimating ANA in CE experiments.

Our review also points to several potentially fruitful directions for research on ANA in choice experiments. For stated ANA, additional work is needed to understand the extent to which the endogeneity of stated ANA information should be controlled for to ensure unbiased model estimates. The hybrid choice model approach proposed by Hess and Hensher (2013) provides a useful framework to mitigate endogeneity concerns, and further efforts to use hybrid choice models to better account for the latent variable nature of stated ANA responses and to assess the bias associated with ignoring the endogeneity seems warranted. Further, there remain questions about how well ANA behavior indicated through stated and inferred methods correspond, with numerous studies providing evidence that stated ANA information may not be a good indicator of ANA behavior (Kragt 2013; Scarpa et al. 2013; Caputo et al. 2018). The inclusion of stated ANA information in ECLC and EAA models as factors affecting latent class membership seems to be a promising approach for utilizing this information without relying solely upon it to determine ANA behavior (Hole, Kolstad, and Gyrd-Hansen 2013).

Moreover, there remain a number of topics deserving of additional attention. The aggregation of welfare estimates from ANA models remains an unexplored topic. As was apparent in our discussion of welfare estimation, WTP is sensitive to the assumptions made about whose values should count in the presence of ANA behavior. The unconditional WTP estimates obtained by weighting the WTP for different individuals or classes of individuals displaying a common ANA behavior is specific to the sample. It is unclear the appropriate way to extrapolate the sample results to the population—what assumption does one make about the extent of ANA behavior that exists out-of-sample in the population?

This review has focused largely on discussing the evolution of the ANA literature and the methods found therein. We have spent little time on what factors may influence ANA behavior, though a few studies do examine the reasons ANA behavior occurs (e.g., Alemu et al. 2013), with much of the early literature focusing on experimental design features (number of attributes, number of levels, etc.) motivating ANA behavior (e.g., Hensher 2006a). Exploring whether other CE survey design features influence ANA behavior is also important. For instance, what role does consequential survey designs (Vossler, Doyon, and Rondeau 2012) have on ANA behavior? Koetse (2017) argues that the lack of consequentiality is a likely driver of cost attribute non-attendance, but little work has been done to formally test this hypothesis. Similarly, more attention can been given to understand the relationship between attribute salience and ANA (Bordalo, Gennaioli, and Shleifer 2013).

The extent to which CE experimental design construction may influence ANA behavior is another understudied topic. In a comparison of ANA behavior using three different experimental designs, Yao et al. (2015) find that a Bayesian D-efficient experimental design (one that accounts for uncertainty around non-zero utility parameters) has a higher rate of full attribute attendance compared to an orthogonal design and a D-efficient experimental design based on a naïve (zero utility parameters) underlying choice model. Given the prevalence of non-Bayesian D-efficient experimental designs in the CE literature, more research is needed to better understand how influential experimental designs are on ANA behavior and whether a broader move towards Bayesian efficient designs is needed to reduce effects on choice behavior. Furthermore, the sensitivity of ANA behavior to the range of attribute levels chosen in the design is an issue that has been raised by Hensher, Rose, and Greene (2012) and Thiene, Scarpa, and Louviere (2015), but has received little formal attention to date.

Additional research is also needed to disentangle ANA from true zero preferences and to improve our understanding of the motivations for ANA behavior. In other words, more studies like Alemu et al. (2013) are needed. This line of research could take several forms, including formal analyses like those used in Alemu et al. (2013), as well as more qualitative investigations using cognitive interviewing (Willis 2005) that allow for a deeper understanding of how different types of preferences and information processing strategies manifest in stated preference data.

Another question for future research is the extent to which stated preference data are asked to perform relative to revealed preference data, given biases exist in both types of data. For example, Hoehn and Randall (1996) demonstrated that market demand data exhibits embedding effects, much like contingent valuation data. More recently, Ketcham, Kuminoff, and Powers (2016) illustrate inconsistencies in revealed preference data over health-related decisions. Future research should compare ANA in revealed and stated preference data models.

ANA is clearly an important attribute processing strategy, but it may not be the only one that individuals employ in CE studies. There are few efforts to account for both ANA and other attribute processing strategies concurrently (e.g., Hensher and Greene 2010; Hensher 2010). However, within that literature, there is evidence that other attribute-based processing behaviors occur besides ANA. For example, there is evidence of reference point effects (i.e., differential preferences for losses versus gains in attributes) (Hess, Rose, and Hensher 2008) and sensitivity to attribute levels (Truong, Adamowicz, and Boxall 2015; Erdem et al. 2015). Moreover, what if a rationally-adaptive choice model is not what is driving choice behavior for some respondents? Several studies employ LCL modeling approaches to try to identify other decision heuristics besides a weighted adding strategy, like lexicographic preferences, elimination-by-aspect, and random regret minimization (e.g., Hess, Stathopoulos, and Daly 2012; Leong and Hensher 2013).

Another area that has yet to be explored is the extent to which choice heuristics, including ANA behaviors, are static for the same individual across choice tasks. Do attribute processing strategies evolve as the individual progresses through the CE questions? While the bounded rationally-adaptive choice model has been interpreted as driving the information processing stage of the process-outcome choice formulation, the processing stage need not be a static one that only occurs prior to all choices being made, leaving the door open to potentially a more dynamic process of choice heuristic formation (e.g., Cameron and DeShazo 2010).

Finally, we recommend researchers use cognitive interviews during qualitative testing and development of CE surveys when funding is available, given they provide opportunities to investigate issues related to ANA behavior and potentially other choice heuristics and decision strategies in CE behavior. These interviews provide a setting in which individuals can be probed about the thought processes they employ while answering CE questions. To the extent ET equipment is available, these settings would also be ideal to collect eye movement data to measure visual ANA and to debrief with individuals about why they paid attention to the elements of the survey they were observed to visually pay attention to. This could aid in the improvement of CE surveys and provide important insights into the links between cognitive processes, visual attention, and attribute attendance.

References

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

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

Balcombe, K., I. Fraser, and E. McSorley. 2015. "Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments." *Journal of Applied Econometrics* 30:447-467.

Balcombe, K., I. Fraser, L. Williams, and E. McSorley. 2017. "Examining the Relationship Between Visual Attention and Stated Preferences: A Discrete Choice Experiment Using Eye-Tracking." *Journal of Economic Behavior and Organization* 144:238-257.

Bello, M., and A. Abdulai. 2016. "Impact of ex-ante hypothetical bias mitigation methods on attribute non-attendance in choice experiments." *American Journal of Agricultural Economics* 98(5):1486-1506.

Ben-Akiva, M., and S.R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand.* Cambridge, MA: The MIT Press.

Bettman, J.R., M.F. Luce, and J.W. Payne. 1998. "Constructive Consumer Choice Processes." *Journal of Consumer Research* 25(3):187-217.

Bordalo, P., N. Gennaioli, and A. Shleifer. 2013. "Salience and Consumer Choice." *Journal of Political Economy* 121(5):803-43.

Boxall, P.C., and W.L. Adamowicz. 2002. "Understanding Heterogeneous Preferences in Random Utility Models: A Latent Class Approach" *Environmental and Resource Economics* 23:421-446.

Cameron, T.A. and J.R. DeShazo. 2010. "Differential attention to attributes in utility- theoretic choice models" *Journal of Choice Modelling* 3(3):73–115.

Campbell, D., D.A. Hensher, and R. Scarpa. 2011. "Non- attendance to attributes in environmental choice analysis: a Latent class specification" *Journal of Environmental Planning and Management* 54(8):1061–76.

-----. 2012. "Cost thresholds, cut-offs, and sensitivities in stated choice analysis: Identification and implications." *Resource and Energy Economics* 34:396-411.

Campbell, D., W.G. Hutchinson, and R. Scarpa. 2008. "Incorporating discontinuous preferences into the analysis of discrete choice experiments." *Environmental and Resource Economics* 41:401-417.

Caputo, V., E.J. Van Loo, R. Scarpa, R.M. Nayga Jr., and W. Verbeke. 2018. "Comparing serial, and choice task stated and inferred attribute non-attendance methods in food choice experiments." *Journal of Agricultural Economics* 69(1):35-57.

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

Chalak, A., M. Abiad, and K. Balcombe. 2016. "Joint use of attribute importance rankings and non-attendance data in choice experiments." *European Review of Agricultural Economics* 43(5):737-760.

Chavez, D., M. Palma, and A. Collart. 2018. "Using Eye-Tracking to Model Attribute Non-attendance in Choice Experiments." *Applied Economics Letters* 25(19):1355-1359.

Collins, A.T., J.M. Rose, and D.A. Hensher. 2013. “Specification issues in a generalized random parameters attribute nonattendance model.” *Transportation Research Part B* 56:234-253.

Colombo, S., M. Christie, and N. Hanley. 2013. “What are the consequences of ignoring attributes in choice experiments? Implications for ecosystem service valuation.” *Ecological Economics* 96:25-35.

Conlisk, J. 1996. “Why bounded rationality?” *Journal of Economic Literature* 34(2):669-700.

Dawes, R.M., D. Faust, and P.E. Meehl. 1989. “Clinical versus actuarial judgment.” *Science*, 243:1668-1674.

de Bekker-Grob, E., M. Ryan, and K. Gerard. 2012. “Discrete Choice Experiments in Health Economics: A Review of the Literature.” *Health Economics*, 21:145-172.

Dellaert, B.G.C., J.D. Brazell, and J.L. Louviere. 1999. “The Effect of Attribute Variation on Consumer Choice Consistency.” *Marketing Letters* 10(2):139-147.

DellaVigna, S. 2009. “Psychology and economics: Evidence from the field." *Journal of Economic literature* 47(2):315-72.

dePalma, A., G.M. Meyers, and Y.Y. Papageorgiou. 1994. “Rational Choice Under an Imperfect Ability To Choose.” *The American Economic Review* 84(3):419-440.

DeShazo, J.R. 2002. “Designing transactions without framing effects in iterative question formats.” *Journal of Environmental Economics and Management* 43(3):360–85.

DeShazo, J.R. and G. Fermo. 2002. “Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency.” *Journal of Environmental Economics and Management* 44:123-143.

-----. 2004. “Implications of rationally- adaptive pre- choice behaviour for the design and estimation of choice models.” Working paper, School of Public Policy and Social Research, University of California at Los Angeles, August.

Erdem, S., D. Campbell, and A.R. Hole. 2015. “Accounting for Attribute-Level Non-attendance in a Health Choice Experiment: Does it Matter?” *Health Economics* 24:773-789.

Fishburn, P.C. 1974. “Lexicographic Orders, Utilities and Decision Rules: A Survey.” *Management Science* 20(11):1442-1471.

Foster, V., and S. Mourato. 2002. “Testing for Consistency in Contingent Ranking Experiments.” *Journal of Environmental Economics and Management* 44:309-328.

Gabaix, X., and D. Laibson. 2000. “A Boundedly Rational Decision Algorithm.” *The American Economic Review: Papers and Proceedings* 90(2):433-438.

Gabaix, X. 2019. “Behavioral Inattention.” *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2. North-Holland. 261-343.

George, E.I., and R.E. McCulloch. 1993. “Variable Selection via Gibbs Sampling.” *Journal of the American Statistical Association* 88(423):881-889.

Gilbride, T.J., G.M. Allenby, and J.D. Brazell. 2006. “Models for Heterogeneous Variable Selection.” *Journal of Marketing Research* 63(August):420-430.

Gilovich, T., D. Griffin and D. Kahneman (eds). 2002. *Heuristics and Biases – the Psychology of Intuitive Judgment*, Cambridge: Cambridge University Press.

Glaholt, M.G., and E.M. Reingold. 2011. "Eye Movement Monitoring as a Process Tracing Methodology in Decision Making Research." *Journal of Neuroscience, Psychology, and Economics* 4(2):125-146.

Glenk, K., J. Martin-Ortega, M. Pulido-Velazquez, and J. Potts. 2015. "Inferring attribute non-attendance from discrete choice experiments: Implications for benefit transfer." *Environmental and Resource Economics* 60:497-520.

Greene, W.H. and D.A. Hensher. 2010. "Ordered choice, heterogeneity, and attribute processing" *Journal of Transport Economics and Policy* 44(3):331–64.

-----. 2013. "Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model." *Applied Economic Letters* 45(14):1897–1902.

Hanley, N., R.E. Wright, and W.L. Adamowicz. 1998. "Using choice experiments to value the environment: Design issues, current experience, and future prospects." *Environmental and Resource Economics* 11 (3–4):413–428.

Heiner, R.A. 1965. "The origin of predictable behavior." *American Economic Review* 73(4):560-595.

Hensher, D.A. 2004. "Information Processing Strategies in Stated Choice Studies: The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes" Institute of Transport Studies, University of Sydney and Monash University, Working paper ITS-WP-04-01.

-----. 2006a. "How do respondents process stated choice experiments? Attribute consideration under varying information load" *Journal of Applied Econometrics* 21(6):861–78.

-----. 2006b. "Attribute processing in choice experiments" in B. Kanninen (ed.) *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice*, Dordrecht: Springer, 135–58.

-----. 2014. "Attribute Processing as a Behavioural Strategy in Choice Making." Ch. 12 in *Handbook of Choice Modelling*, S. Hess and A. Daly (eds.). Padstow, U.K.: Edward Elgar Publishing, 268-289.

Hensher, D.A. and A. Collins. 2011. "Interrogation of responses to stated choice experiments: Is there sense in what respondents tell us? A closer look at what respondents choose in stated choice experiments" *Journal of Choice Modelling* 4(1):62–89.

Hensher, D.A. and W.H. Greene. 2010. "Non- attendance and dual processing of common- metric attributes in choice analysis: A latent class specification" *Empirical Economics* 39(2):413–26.

Hensher, D.A. and D. Layton. 2010. "Parameter transfer of common- metric attributes in choice analysis and cognitive rationalisation: Implications for willingness to pay" *Transportation* 37(3):473–90.

Hensher, D.A. and J.M. Rose. 2009. "Simplifying choice through attribute preservation or non- attendance: Implications for willingness to pay" *Transportation Research Part E* 45(4):583–90.

-----. 2012. "The influence of alternative acceptability, attribute thresholds and choice response certainty on automobile purchase preferences" *Journal of Transport Economics and Policy* 46(3):451–68.

Hensher, D.A., A.T. Collins, and W.H. Greene. 2013. "Accounting for attribute non- attendance and common metric aggregation in a latent class mixed multinomial logit model: A warning on potential confoundment" *Transportation* 40(5):1003–20.

Hensher, D.A., J. Rose and W. Greene. 2005. "The implications on willingness to pay of respondents ignoring specific attributes" *Transportation* 32(3):203–22.

-----. 2012. "Inferring attribute non- attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design" *Transportation* 39(2):235–54.

Hess, S., and D.A. Hensher. 2010. "Using Conditioning on Observed Choices to Retrieve Individual-specific Attribute Processing Strategies." *Transportation Research Part B* 44:781-790.

-----. 2013. "Making use of respondent reported processing information to understand attribute importance: A latent variable scaling approach." *Transportation* 40:397-412.

Hess, S., and J.M. Rose. 2007. "A Latent Class Approach to Recognising Respondents' Information Processing Strategies in SP Studies." Presentation at the Oslo workshop on valuation methods in transport planning, OWVM, Oslo.

Hess, S., A. Stathopoulos, D. Campbell, V. O'Neill, and S. Caussade. 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* 40:583-607.

Hess, S., A. Stathopoulos, and A. Daly. 2012. "Allowing for heterogeneous decision rules in discrete choice models: An approach and four case studies." *Transportation* 39:565-591.

Hole, A.R. 2011. "A discrete choice model with endogenous attribute attendance." *Economics Letters* 110:203-205.

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

Hole, A.R., R. Norman, and R. Viney. 2016. "Response Patterns in Health State Valuation Using Endogenous Attribute Attendance and Latent Class Analysis." *Health Economics* 25:212-224.

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

Jonker, M.F., B. Donkers, E.W. de Bekker-Grob, and E.A. Stolk. 2018. "Effect of Level Overlap and Color Coding on Attribute Non-Attendance in Discrete Choice Experiments." *Value in Health* 21:767-771.

Johnston, R. J., K. J. Boyle, W. Adamowicz, J. Bennett, R. Brouwer, T. A. Cameron, W. M. Hanemann, N. Hanley, M. Ryan, R. Scarpa, R. Tourangeau, and C. A. Vossler. 2017. "Contemporary Guidance for Stated Preference Studies." *Journal of the Association of Resource and Environmental Economists* 4(2):319-405.

Jourdain, D., and S. Vivithkeyoonvong. 2017. "Valuation of ecosystem services provided by irrigated rice agriculture in Thailand: A choice experiment considering attribute nonattendance." *Agricultural Economics* 48:655-667.

Kehlbacher, A., K. Balcombe, and R. Bennett. 2013. "Stated Attribute Non-Attendance in Successive Choice Experiments." *Journal of Agricultural Economics* 64(3):693-706.

Ketcham, J.D., N.V. Kuminoff, and C.A. Powers. 2016. "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program: Comment." *American Economic Review* 106(12):3932-61.

Koetse, M.J. 2017. "Effects of payment vehicle non-attendance in choice experiments on value estimates and the WTA-WTP disparity." *Journal of Environmental Economics and Policy* 6(3):225-245.

Kragt, M.E. 2013. "Stated and Inferred Attribute Attendance Models: A Comparison with Environmental Choice Experiments." *Journal of Agricultural Economics* 64(3):719-736.

Kravchenko, A. 2014. "Influence of rudimentary attribute non-attendance (ANA) on choice experiment parameter estimates and design efficiency: A monte carlo simulation analysis." *Journal of Choice Modelling* 11:57-68.

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

Lagarde, M. 2013 "Investigating Attribute Non-Attendance and its Consequences in Choice Experiments with Latent Class Models." *Health Economics* 22:554-567.

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

Layton, D., and D.A. Hensher. 2010. "Aggregation of common- metric attributes in preference revelation in choice experiments and implications for willingness to pay" *Transportation Research Part D* 15 (7):394–404.

Leong, W., and D.A. Hensher. 2012. "Embedding decision heuristics in discrete choice models: A review" *Transport Reviews* 32 (3):313–31.

Lew, D.K. 2018. "Discounting Future Payments in Stated Preference Choice Experiments." *Resource and Energy Economics* 54:150-164.

-----. 2019 "Place of Residence and Cost Attribute Non-Attendance in a Stated Preference Choice Experiment Involving a Marine Endangered Species" *Marine Resource Economics* 34(3):225-245.

Lew, D.K., and J.C. Whitehead. This issue. "Attribute Non-Attendance in Choice Experiments of Marine Ecosystem Goods and Services." *Marine Resource Economics*.

Lewis, D.J., S.J. Dundas, D.M. Kling, D.K. Lew, and S.D. Hacker. 2019. "The Non-Market Benefits of Early and Partial Gains in Managing Threatened Salmon." *PLOS One* 14(8):e0220260.

Louviere J, and G. Woodworth. 1983. "Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data." *Journal of Marketing Research* 20:350–367.

Louviere, J.J. 1988. "Conjoint analysis modelling of stated preferences: a review of theory, methods, recent developments and external validity." *Journal of Transportation.Economics and Policy* 1:93–119.

Louviere, J.J., E. Lancsar. 2009. "Choice experiment in health: the good, the bad, the ugly and toward a brighter future." *Health Economics, Policy, and Law* 4:527–546.

Louviere, J.J., D.A. Hensher, and J. Swait. 2000. *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Valuation*. Cambridge, U.K.: Cambridge University Press.

Luce, R.D. 1959. *Individual Choice Behavior: A Theoretical Analysis*. New York: Wiley.

Malone, T., and J.L. Lusk. 2018 "Consequences of participant inattention with an application to carbon taxes for meat products." *Ecological Economics* 145:218-230.

Manski, C.F. 1977 "The Structure of Random Utility Models." *Theory and Decision* 8:229-254.

Mariel, P. D. Hoyos, and J. Meyerhoff. 2013. "Stated or Inferred Attribute Non-Attendance? A Simulation Approach." *Economia Agraria y Recursos Naturales* 13(1):51-67.

Mazzotta, M.J., and J.J. Opaluch. 1995. "Decision making when choices are complex: A test of Heiner's Hypothesis." *Land Economics* 71(4):500-515.

McFadden, D. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3:303-328.

Meehl, P.E. 1954. "Clinical versus statistical prediction: A theoretical analysis and a review of the evidence." Minneapolis, MN, US: University of Minnesota Press. 149 pp., http://dx.doi.org/10.1037/11281-000

-----. 1989. “Causes and effects of my disturbing little book.” *Journal of Personality Assessment*, 50(3):370-375.

Mohanty, S.K., K. Balcombe, R. Bennett, G. Nocella, and I. Fraser. 2019. “Attribute Specific Impacts of Stated Non-Attendance in Choice Experiments.” *Journal of Agricultural Economics* 70(3):686-704.

Orquin, J.L., and K. Holmqvist. 2018. “Threats to the Validity of Eye Movement Research in Psychology.” *Behavior Research Methods* 50:1645-1656.

Orquin, J.L., and S.M. Loose. 2013. “Attention and Choice: A Review of Eye Movements in Decision Making.” *Acta Psychologica* 144:190-206.

Petrolia, D., and J. Hwang. This issue. “Accounting for Attribute Non-Attendance in Three Previously-Published Choice Studies of Coastal and Marine Resources” *Marine Resource Economics*.

Puckett, S.M., and D.A. Hensher. 2008. “The Role of Attribute Processing Strategies in Estimating the Preferences of Road Freight Stakeholders.” *Transportation Research Part E* 44:379-395.

Randall, A., and J.P. Hoehn. 1996. "Embedding in Market Demand Systems." *Journal of Environmental Economics and Management* 30(3):369-80.

Rose, J.M., D.A. Hensher, W.H. Greene, and S.P. Washington. 2012. “Attribute Exclusion Strategies in Airline Choice: Accounting for Exogenous Information on Decision Maker Processing Strategies in Models of Discrete Choice.” *Transportmetrica* 8(5):344-360.

Rose J.M., D.A. Hensher, and W.H. Greene. 2005. “Recovering Costs Through Price and Service Differentiation: Accounting for Exogenous Information on Attribute Processing Strategies in Airline Choice.” *Journal of Air Transport Management* 11:400-407.

Russo, J.E., and B.A. Dosher. 1983. “Strategies for Multiattribute Binary Choice.” *Journal of Experimental Psychology* 9(4):676-696.

Ryan, M., N. Krucien, and F. Hermens. 2018. “The Eyes Have It: Using Eye Tracking to Inform Information Processing Strategies in Multi-Attributes Choices.” *Health Economics* 27(4):709-721.

Saelensminde, K. 2002. “The impact of choice inconsistencies in stated choice studies.” *Environmental and Resource Economics* 23:403–20.

Sandorf, E.D., D. Campbell, and N. Hanley. 2017. “Disentangling the influence of knowledge on attribute non-attendance.” *Journal of Choice Modelling* 24:36-50.

Scarpa, R., T. Gilbride, D. Campbell, and D.A. Hensher. 2009. “Modelling attribute non- attendance in choice experiments for rural landscape valuation” *European Review of Agricultural Economics* 36(2):151–74.

Scarpa, R., M. Thiene, and D.A. Hensher. 2010. “Monitoring choice task attribute attendance in non- market valuation of multiple park management services: does it matter?” *Land Economics* 86(4):817–39.

Scarpa, R., R. Zanoli, V. Bruschi and S. Naspetti. 2013. “Inferred and stated attribute non- attendance in food choice experiments” *American Journal of Agricultural Economics* 95(1):165–80.

Simon, H.A. 1955. “A behavioral model of rational choice.” *Quarterly Journal of Economics* 69(1):99-118.

-----. 1956. “Rational choice and the structure of the environment.” *Psychological Review* 63(2):129-138.

-----. 1959. “Theories of decision-making in economics and behavioral science.” *American Economic Review* 49(3):253-283.

-----. 1978 “Rationality as a process and as product of thought.” *American Economic Review (Papers and Proceedings)* 68(2):1-16.

Swait, J., and W. Adamowicz. 2001. “The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching.” *Journal of Consumer Research* 28:135-148.

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

Thiene, M., R. Scarpa, and J.J. Louviere. 2015. "Addressing Preference Heterogeneity, Multiple Scales and Attribute Attendance with a Correlated Finite Mixing Model of Tap Water Choice." *Environmental and Resource Economics* 62(3): 637-656.

Titz, J., A. Schotz, and P. Sedlmeier. 2018. “Comparing Eye Trackers by Correlating Their Eye-Metric Data.” *Behavior Research Methods* 50:1853-1863.

Train, K.E. 2009. *Discrete Choice Methods with Simulation*, 2nd edition. New York: Cambridge University Press.

Train, K.E., and G. Sonnier. 2005. “Mixed Logit with Bounded Distributions of Correlated Partworths.” Chapter 7 in Applications of Simulation Methods in Environmental and Resource Economics (R. Scarpa and A. Alberini, eds.), Dordrecht, The Netherlands: Springer, 117-134.

Truong, T.D., W.L. Adamowicz, and P.C. Boxall. 2015. “Modeling Non-Compensatory Preferences in Environmental Valuation.” *Resource and Energy Economics* 39:89-107.

Tversky, A. 1972. “Choice by Elimination.” *Journal of Mathematical Psychology* 9:341-367.

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

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

van Loo, E.J., R.M. Nayga, D. Campbell, H.-S. Seo, and W. Verbeke. 2018. “Using Eye Tracking to Account for Attribute Non-Attendance in Choice Experiments.” *European Review of Agricultural Economics* 45(3):333-365.

Vossler, C., M. Doyon, and D. Rondeau. 2012. “Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments.” *American Economic Journal: Microeconomics* 4(4):145-71.

Wedel, M., and R. Pieters. 2008. “A Review of Eye-Tracking Research in Marketing Research.” Ch. 5 in *Review of Marketing Research* (N.K. Malhotra, ed.), New York: M.E. Sharpe. Pages 123-147.

Weller, P., M. Oehlmann, P. Mariel, and J. Meyerhoff. 2014. “Stated and inferred attribute non-attendance in a design of designs approach.” *Journal of Choice Modelling* 11:43-56.

Whitehead, J., T. Haab, J.-C. Huang. 2011. *Preference Data for Environmental Valuation: Combining Revealed and Stated Approaches*. Routledge: New York.

Willis, G.B. 2005. *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.

Yao, R.T., R. Scarpa, J.M. Rose, and J.A. Turner. 2015. “Experimental Design Criteria and Their Behavioural Efficiency: An Evaluation in the Field.” *Environmental and Resource Economics* 62:433-455.

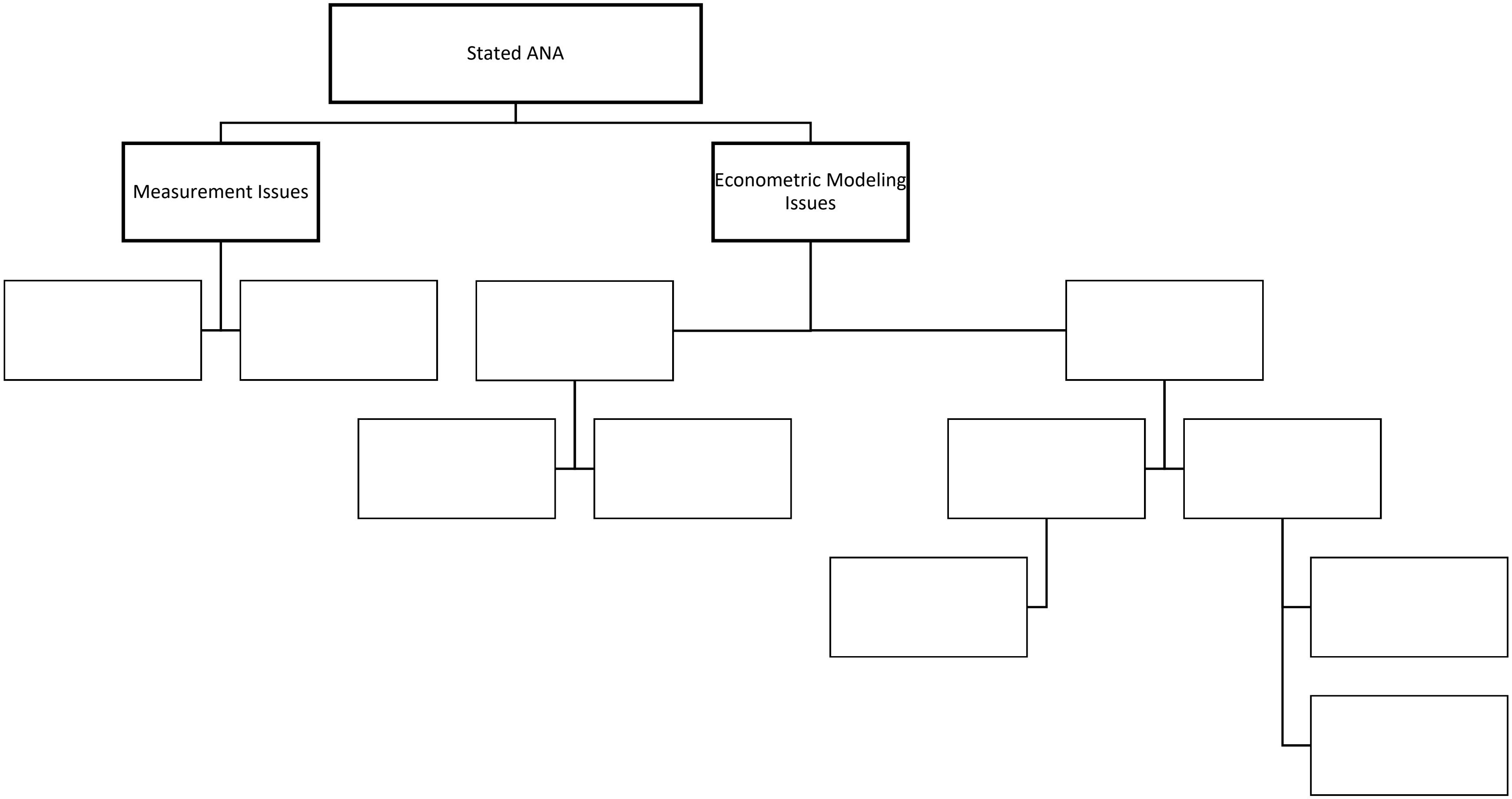

Figure 1. A Typology of Stated Attribute Non-Attendance Models and Measurement Issues

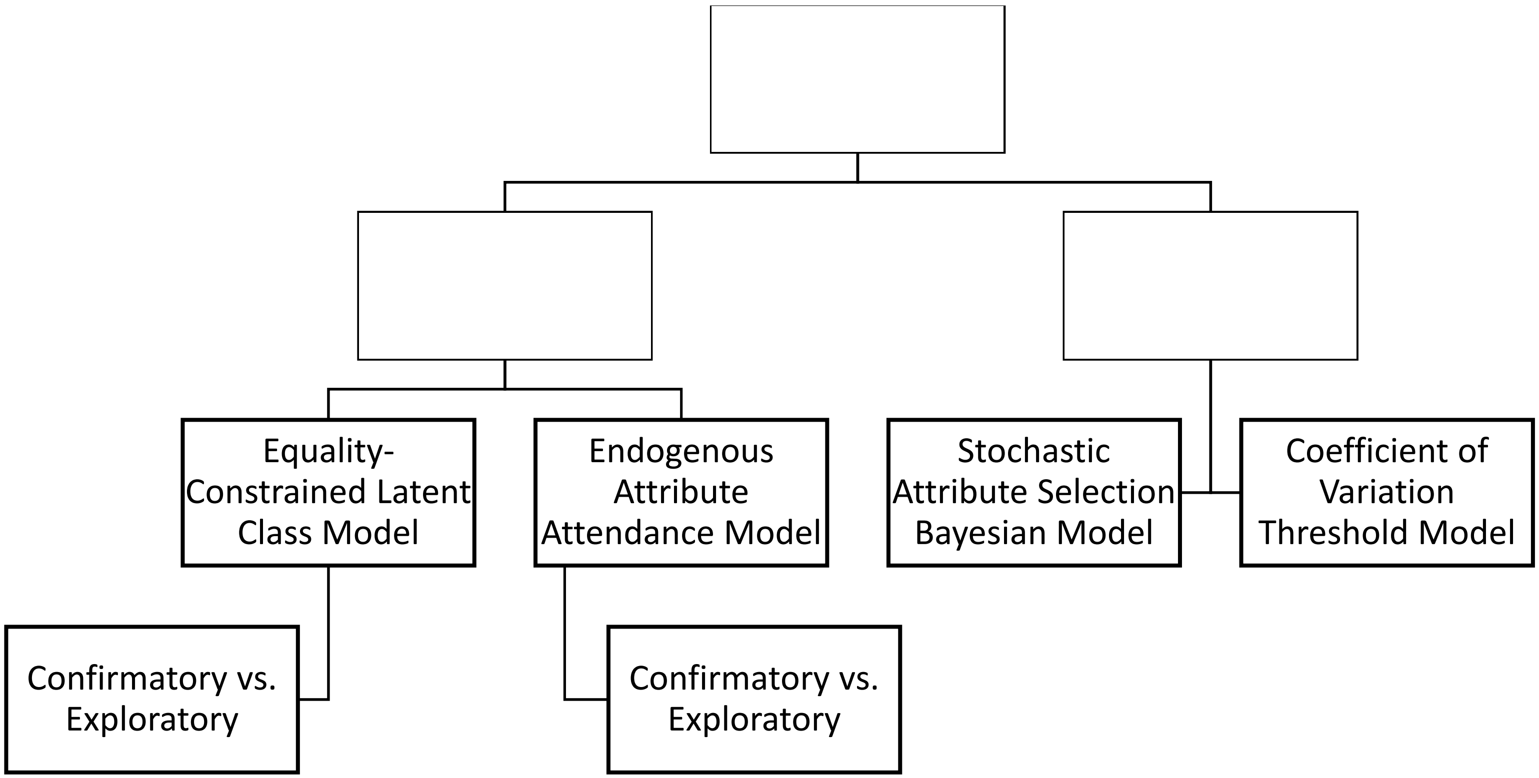

Figure 2. A Typology of Inferred Attribute Non-Attendance Models