Attribute Non-Attendance and Satisficing Behavior in Online Choice Experiments

Michael S. Jones[1], Lisa A. House[2], and Zhifeng Gao[3]

Abstract*:*

While a successful survey requires engaged and attentive respondents, careless survey completion remains a great concern in online market research. In this article, we test metrics of engagement in an online willingness-to-pay (WTP) study for fresh blueberry attributes using a major U.S. panel company and evaluate the impact that poorly behaving respondents have on subsequent data quality. In doing so, we investigate in detail the complex joint relationship between attribute non-attendance (ANA) and measures of respondent engagement in web surveying. Using fixed latent classes, an approach known as the Equality Constrained Latent Class procedure, we export individual probabilistic class assignment of all levels of attribute attendance to cross reference with respondents who fail measures of engagement and fraudulence, and analyze their composition and impact on latent classes, indicating non-attendance of individual and combinations of attributes. We also analyze engagement impacts on the tau variance parameter in the scaled mixed logit model and find strong links to unnecessarily increased heterogeneity when not properly filtering poorly behaving respondents. While WTP estimates between respondents passing and failing engagement metrics are similar with the ECLC model, filtering failing respondents in the scaled mixed logit model reduces overall WTP estimates. Results have implications for both WTP researchers and general online market researchers.

[1] Michael Jones is a Graduate Fellow at North Carolina State University (formerly a research assistant at the University of Florida); [2]Lisa House is a Professor in Food and Resource Economics and [3]Zhifeng Gao is an Associate Professor in Food and Resource Economics at the University of Florida. This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2009-55818-05077 from the USDA National Institute of Food and Agriculture.

Online surveying has firmly established itself in the last decade as a popular, low-cost, broad-reaching method to conduct market research (Dillman, Smyth, and Christian 2009). Online access panels, probability-based or otherwise, are a particularly popular recruiting tool in online market research due to their convenience and cost (Baker et al. 2010). Agricultural market and policy research is no exception and frequently relies on online panel respondents (Bartels and van den Berg 2011; Ellison and Lusk 2011; Kinsey et al. 2009; Tonsor and Wolf 2010), especially in willingness-to-pay (WTP) literature.[1]

However, compensated online research through web-panels has many potential drawbacks that must be considered to ensure the validity of data and resulting estimates. Incentive offers for participation may induce individuals to attempt to falsely "qualify" in the survey screening period and/or exert little or no effort within the survey, with respondents simply desiring to finish and collect payment. Sobering questions that researchers must ask themselves are as follows: (1) Are all of the panel respondents in my survey exerting sufficient and genuine effort in their answers? and (2) If not, what potentially negative effects are these poorly behaving respondents introducing into the resulting dataset? If poorly qualified (poor) respondents are in fact in the dataset and significantly affect data quality, then online researchers must be vigilant in controlling this problem.

Especially in WTP research with choice experiments (CEs), consistency in intra-respondent choice patterns is crucial to determining rational preferences. Recent work demonstrates that irrationality within the CE, as measured by class revealed preference axioms, is strongly predicted by poor respondent behavior such as fraudulently "qualifying" for the survey, missing simple embedded directives known as trap questions (TQs), and speeding through critical portions like the CE (Jones, House, and Gao, forthcoming). Frequent failure of the weak axiom of revealed preference (WARP) and the strong axiom of revealed preference (SARP) are long-acknowledged indicators of poor CE data quality (Lancsar and Louviere 2006) and thus provide a time-tested lens for analysis. Irrational preference patterns lead to poor estimation power of statistical models and introduce significant variance into the calculation of an individual's attribute coefficients. Incomplete preferences are further related to processing patterns described as attribute non-attendance (ANA), or ignoring one or more attributes in a choice set. ANA is receiving attention from a growing body of literature, as failing to account for this through modeling strategies, especially non-attendance of the price attribute, is linked to upwardly biased WTP estimates (Hess and Hensher 2010; Scarpa et al. 2009;). We hypothesize that a major potential biasing component in online research and driver of ANA in CEs is that poor quality panel respondents may not be providing thoughtful answers to all questions. While potential problems with ANA and poor respondent data quality have been separately acknowledged and investigated, the link between the two and the resulting effects on WTP estimates is currently an empirical question.

In this study, we survey 1,880 respondents through a major U.S. panel company about their blueberry purchasing preferences. We use a CE and measure respondent engagement through several classic identifiers of poor respondent behavior. These include low-probability screening questions to identify respondents attempting to fraudulently enter the survey, trap questions (TQs), invisible speed metrics for individual survey segments, and an age validation strategy. Jointly, these quarantining filters identify respondents who have a high probability of fraudulence and/or inattention. Each metric utilized can be easily incorporated into any online survey to filter these respondents in real time. In this study, instead of filtering failing respondents, we retain them to more critically examine their preference patterns.

We also utilize innovative scaled mixed logit models, simulating datasets both with and without implementation of poor respondent filters. This allows us to examine the impact of not screening poor respondents on results. Additionally, we will investigate the rate of ANA by respondents who do fail the screening criteria through the equality constrained latent class model.

Background

Researchers have long discussed the potential for poor quality data from online panels, probability-based or otherwise, when fraudulent or inattentive respondents are present in the sample (Miller and Baker-Prewitt 2009; Smith and Hofma-Brown 2005, 2006; Willems, van Ossenbrugge, and Vonk 2006). In one study across fourteen popular U.S. panels, Miller (2006) finds that fraudulent and inattentive respondents composed 8–25% of the total sample, with inattentive respondents appearing to cause the largest bias in data. To identify fraudulent respondents attempting to falsely qualify for surveys, researchers have used low-probability screening questions in the opening sections (Smith and Hofma-Brown 2005; Downes-Le Guin 2005). Inattentive respondents, or *satisficers,* (Krosnick 1991) have been identified with trap questions (TQs) (Downes-Le Guin et al. 2012; Gao, House, and Xie 2013), thresholds for unusually fast completion times or "speeding" (Miller and Baker-Prewitt 2009), and inconsistencies in behavioral or demographics response patterns (Author citation 2013; Miller and Baker-Prewitt 2009). *Satisficing* is a common marketing literature term originally coined by Simon (1957) to describe actors failing to exert sufficient effort to maximize profit in a decision-making process; instead, actors seek an acceptable profit threshold. In the case of web surveys, this translates to only completing the survey adequately enough to receive an incentive offer.

Removing fraudulent and inattentive respondents who fail these quarantining metrics has been shown to decrease incidence of irrational response patterns in the CE, as measured by SARP and WARP (Jones, House, and Gao, forthcoming). Fundamental to the choice-based approach to consumer theory, SARP and WARP have been argued as being superior to alternative metrics in the preference-based approach,[2] as it allows rationality to be established in a directly observable manner that obliges fewer restrictions on demand and "provides a behavioral foundation to the theory of individual decision making" (Lancsar and Louviere 2006).

Originally described by Samuelson (1938), WARP is satisfied when, given vectors of prices and attribute bundles (p_t, x_t) for *t* = 1,…,T, then bundle x_1 is *directly revealed preferred* to a bundle x_2 if $p_t x_1 \geq p_t x_2$. Bundle x_2 then cannot be subsequently *directly revealed preferred* to x_1. SARP is a more rigorous, transitive expansion that is satisfied when, given the same vectors, bundle x_1 is indirectly *revealed preferred* to x_3 if a sequence exists such that $p_t x_1 \geq p_t x_2$ and $p_t x_2 \geq p_t x_3$. While WARP is necessary for rational preferences, it is SARP, or the confirmation of a symmetric and negative semi-definite Slutsky substitution matrix, which is "necessary and sufficient" to assure existence of a rational preference set (Houthakker 1950). Additionally, the weakened generalized axiom of revealed preference (GARP) developed by Varian (1982) to allow for multi-valued indifference curves is also a necessary and sufficient condition. In a forced-choice scenario which CEs almost invariably assume (no intra-choice indication of "indifference"), GARP and SARP violations become identical.

Attribute Non-Attendance

Respondents have also long been noted to not "attend" to, or not pay attention to, all attributes in a CE (Hensher 2006; Hensher, Rose, and Greene 2005). Non-attendance, particularly of the price attribute, is linked to significantly inflated WTP estimates and is thus an important consideration for any CE researcher (Hensher, Rose, and Greene 2012; Scarpa et al. 2009). Some identification of ANA has taken the form of *stated non-attendance* (SA), or explicit supplementary questions inquiring about the importance of attributes to a specific respondent, although the empirical validity of this approach has been challenged (Hensher and Rose 2009; Hess and Hensher 2010). Instead, *inferred non-attendance* (IA), or use of model inference to identify and potentially treat ANA, has been proposed through fixed latent class methods (Hensher and Greene 2010; Hole 2010; Scarpa et al. 2009;) and stochastic attribute selection through Bayesian estimation (Scarpa et al. 2009). This article focuses on ANA specification through fixed latent classes or the equality constrained latent class (ECLC) model, designing all possible linear utility functions that could be followed by respondents, given the limited product attributes provided in the CE. This strategy creates, for a certain number of attributes K, a 2^K rule for possible linear utility functions in which respondents attend all, some, or none of the product attributes. Some utility functions may be theoretically sound, such as ignoring a qualitative component of a product, while behaviors such as price ANA challenge basic fundamentals of economic theory and greatly inhibit proper WTP estimation (Lancsar and Louviere 2006).

Several factors may drive respondents' processing strategies that probabilistically sort them into latent classes indicating ANA of single, multiple, or all CE attributes. In general, ANA has been credited to several drivers, including response to the hypothetical nature of CEs (particularly for the price attribute) (Scarpa et al. 2009), cognitively simplifying choice sets (Alemu et al. 2013), and researchers' poor specification of attribute ranges which fail to induce compensatory behavior (Hensher, Rose, and Greene 2012). As many CE experiments are administered online and provide compensation for completion, another factor may also be inducing perverse behavior among survey participants—inattention (i.e., simply not caring about the research or survey design and only attempting to collect the incentive offer). It is this aspect which we investigate in depth in this work.

Experimental Design

Our survey was designed to include low-probability screening questions, TQs, invisible speed metrics for individual survey segments, and an age validation strategy along with a choice experiment. Four versions of this survey were tested to focus on the effect of placement and frequency of TQ inclusion. The placement of interest was at the beginning of the survey and just before the choice experiment. Other than which answer to select, both the beginning and end TQ were identical. The intent was to be as courteous to the respondent as possible while hinting at an explanation for the question's inclusion. The TQ read: "Please select 'disagree' for this line. Thank you for reading carefully."

Therefore, 1,880 respondents were randomly allocated to the following four versions, with versions weighted based on past TQ research to ensure sufficient failures for analysis:

1. Control (n=429): No TQ
2. Beginning TQ Only (n=474): Placement of one TQ in the first line of a question grid in the beginning of the survey (after median 0.87 minutes)
3. End TQ Only (n=451): Placement of one TQ in the last line of a question grid just before the choice experiment (after median 6.38 minutes)
4. Beginning and End TQs (n=526): Placement of two TQs, with one in the first line of a question grid in the beginning of the survey (after median 0.84 minutes) and the other in the last line of a question grid just before the choice experiment (after median 6.39 minutes)

For the purposes of analysis, these versions are pooled to achieve adequate sample size to separately analyze participants who fail any of the poor behavior indicators. All versions have identical complimenting low-probability screening questions, age verifications, and (invisible) speeding metrics. The low-probability screener regarded fresh berry purchases in the past year. In addition to common options such as fresh blueberries, fresh strawberries and fresh raspberries, improbable options were fresh goji berries, fresh red currants, and fresh muscadine grapes. Potentially fraudulent applicants were flagged only if they chose at least two of three of these options. Due to extremely sparse and geographically diverse growing locations, it is highly unlikely that an individual actually did consume multiple (fresh) low-probability options. In another verification method, respondents were initially asked to enter their age in the screening section and then asked to select their birth year in the ending demographics section. Inattentive respondents were flagged if the birth year deviated by at least two years from reported age. Finally, if respondents answered the CE at an average rate of ≤ 4 seconds per question then inattention was also flagged.

Each survey version contained a choice experiment to elicit preferences for fresh blueberries. The blueberry products contained a simple design of three attributes: production method (organic vs. conventional), production locale ("local" vs. "U.S."), and four levels for price, summarized in table 1. The price spread chosen mirrors blueberry price variance in the United States for the Aug/Sept 2013 survey period (USDA 2013).

Table 1. Choice Experiment Attributes

Choice Experiment Attributes

Attributes	Levels
Production Method	Organic, Conventional

Production Location	"Local" (within state of residence), "U.S." (outside of state of residence)
Price (USD/pint)	\$1.49, \$2.49, \$3.49, \$4.49

A full factorial design results in $(2x2x4)^2 = 256$ choice combinations. The OPTEX procedure in SAS provided a D-optimal fractional factorial design of 13 randomly-ordered choices for respondents to make the choice task more reasonable. Figure 1 illustrates the choice presentation in the online survey. To avoid possible confusion that higher priced goods represent higher quality goods, we include in the CE instructions to "Please assume the only characteristics that differ are those which are explicitly mentioned". We also include a "cheap talk" script to attempt to diminish hypothetical bias and encourage participants to select as they would in a purchasing scenario (Lusk 2003). Importantly, the "I would choose neither" option is not offered in this CE design, as this discourages *satisficing* by avoiding difficult decisions (Carson el al. 1994) and, importantly, avoids potentially debilitating specification problems when evaluating preference axiom violations (Hougaard, Tjur, and Østerdal 2012). This follows other CE work specifically investigating data quality (McIntosh and Ryan 2002). Respondents screened for the survey are primary shoppers who have purchased blueberries in the last year, so removal of the "neither" option is assumed to not have a measureable negative effect on respondent attitudes.[3]

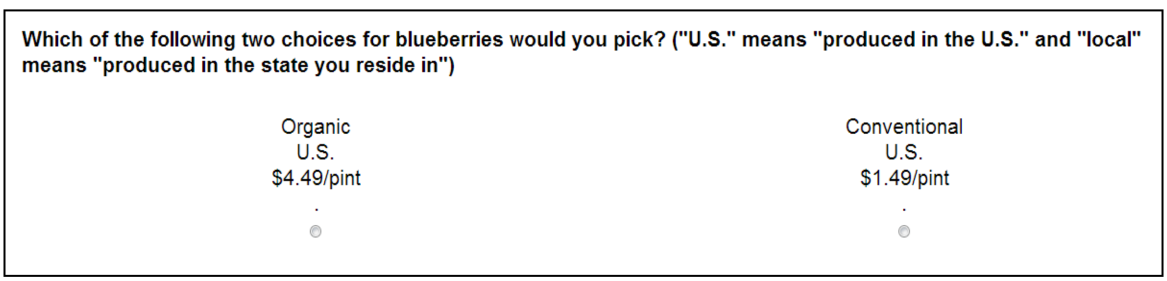

Figure 1. Survey Choice Example

Model Specification

Equality Constrained Latent Class Model

Following Hensher, Rose, and Greene (2012), fixed latent classes allow respondents to sort, based on their consideration of attributes in the choice process, into one of 2^K classes (or classes *q,* when $q = 1, \dots, Q$). This builds on the standard MNL model by probabilistically sorting respondents, following the form:

$$Prob(i,j|q) = \frac{\exp(\beta_q' x_{i,j})}{\sum_{j=1}^{J} \exp(\beta_q' x_{i,j})} \tag{1}$$

The class vector q has been described as a masking vector of form $(\delta_1, \delta_2, \delta_3, \dots)$, where each δ may take the values 0,1. Therefore the β_q vector becomes the interaction (product) of the coefficient β with the masking vector. In this context of blueberry attribute selection, we have three attributes that expand to $2^3 = 8$ classes, fixing utility functions as follows in table 2:

Table 2. Fixed Latent Classes and Corresponding Utility Functions

Latent Class	Linear Utility Functions				
1. Full Attendance	$U(Blueberries) =$	$\beta_1 price +$	$\beta_2 organic +$	$\beta_3 local +$	ε
2. Full Non-Attendance	$U(Blueberries) =$				ε
3. Attend Local Only	$U(Blueberries) =$			$\beta_3 local +$	ε
4. Attend Org. Only	$U(Blueberries) =$		$\beta_2 organic +$		ε
5. Attend Price Only	$U(Blueberries) =$	$\beta_1 price +$			ε
6. Ignore Local Only	$U(Blueberries) =$	$\beta_1 price +$	$\beta_2 organic +$		ε
7. Ignore Org. Only	$U(Blueberries) =$	$\beta_1 price +$		$\beta_3 local +$	ε
8. Ignore Price Only	$U(Blueberries) =$		$\beta_2 organic +$	$\beta_3 local +$	ε

Full Attribute Attendance (AA) describes a respondent who is most likely to sort into a utility function in which the $\beta's$ for every CE attribute are non-zero. Other classes involved single, dual, or complete ANA which may provide insight into the respondents' processing strategies. Since we make no presumption that "rational" respondents should choose any particular qualitative attribute (Lancsar and Louviere 2006), in the case of "organic" or "local" ANA we assume that respondents simply do not care about these attributes. Therefore, the fixed classes "Ignore Local Only", "Ignore Organic Only", and "Attend Price Only" are all perfectly amenable to economic theory.

The remaining latent classes, however, deviate from economic theory since respondents ignore the price attribute (at a minimum). This includes "Attend Organic Only", "Attend Local Only", "Ignore Price Only", and "Full Attribute Non-Attendance" (Full ANA). A respondent who sorts into the Full ANA category does not have any discernibly non-zero $\beta's$ for any CE attributes. This is arguably the least useful class of respondents, as no statistical information may be derived from their CE responses to determine their preferences.

Scaled Mixed Logit Model

The scaled mixed logit model is an approach arising from random parameter estimations in WTP space which account for both scale and preference heterogeneity (Greene and Hensher 2010). Following Greene and Hensher (2010), the traditional mixed multinomial logit model is represented by

$$Prob(choice_{it} = j | x_{it}, z_i, v_i) = \frac{\exp(V_{it,j})}{\sum_{j=1}^{J_{it}} \exp(V_{it,j})} \qquad (2)$$

where

$$V_{it} = \beta_i' \mathbf{x}_{it,j}$$

$$\beta_i = \beta + \Delta\mathbf{z}_i + \Gamma\mathbf{v}_i$$

$\mathbf{x}_{it,j}$ = the set *K* of attributes composing alternative *j* in choice situation *t* observed by individual *i*

$\mathbf{z}_i$ = the set of *M* characteristics of individual *i* which influence the mean of the taste parameters

$\mathbf{v}_i$ = the vector of *K* random variables with mean=0 and known variances and zero covariances

Attempts to account for scale heterogeneity as well as individual preference heterogeneity have led to expansions on the β_i's to estimate additional parameters in the generalized mixed logit model, represented by

$$\beta_i = \sigma_i[\beta + \Delta\mathbf{z}_i] + [\gamma + \sigma_i(1-\gamma)]\Gamma\mathbf{v}_i \qquad (3)$$

where

$$\sigma_i = \exp\left(\bar{\sigma} + \delta'\mathbf{h}_i + \tau w_i\right)$$

σ_i = the standard deviation of the individual-specific error term

$\mathbf{h}_i$ = the set of *L* characteristics of individual *i* which may overlap with $\mathbf{z}_i$

δ = parameters in the observed part of the heterogeneity in the scale term

w_i = unobserved heterogeneity, the distribution of which is assumed to be standard normal

$\bar{\sigma}$ = mean parameter in the variance

γ = weighting parameter indicating the level of variance in the residual preference heterogeneity which varies with scale, which must be estimated between 0 and 1

τ = the coefficient of the unobserved scale heterogeneity

We then specify the model in WTP space, following work which indicates the superiority of directly obtaining estimates in this manner (Balcombe, Chalak, and Fraser 2009; Scarpa, Thiene, and Train 2008;). Following the optimal design found by Balcombe, Chalak, and Fraser, who independently tested twenty models with fixed and random coefficients in preference and WTP space, we evaluated price in a random log-normal distribution and qualitative attributes in a normal distribution. By estimating in WTP space, the weighting parameter γ is fixed at zero and the β for price is normalized at unity. This $\gamma = 0$ specification collapses to the SMXLM (scaled mixed logit model), with

$$\beta_i = \sigma_i[\beta + \Delta\mathbf{z}_i + \Gamma\mathbf{v}_i] \qquad (4)$$

Greene and Hensher (2010) show that without explicitly accounting for observed heterogeneity $\delta'\mathbf{h}_i$, $E[\sigma_i] = \exp\left(\bar{\sigma} + \tau^2/2\right)$. The coefficient τ thus becomes the major driver of the standard deviation of the individual-specific error term and can help researchers determine the level of unobserved heterogeneity within a dataset. We operationalize this concept to analyze the intensity of the variance between passing respondents and those poor respondents who would have been quarantined by real time filters in the survey.

Results

Significantly higher frequencies of respondents flagged as "fraudulent" or "inattentive" are concentrated in certain fixed latent classes (table 3). Results indicate that 29.0% of the Full ANA class and 18.9% of the Attend Organic Only class were flagged as potentially fraudulent. These two classes alone account for over half of the fraudulent respondents in the survey. Inversely, the Full AA and Ignore Organic Only classes have significantly lower rates of fraudulence. Engagement metrics, specifically TQ failure and CE speeding, also indicate that Full ANA and Attend Organic Only classes contain significantly higher rates of inattentive respondents. The speeding metric is particularly worrisome, as little to no cognitive processing would occur during the CE with an average time of ≤ 4 seconds per question.

Table 3. Probabilistic Fixed Class Assignment vs. Quarantining Metrics

<u>**Fixed Latent Class**</u>			**Fraud**	**Inattention**			
(highest probability exported from Nlogit)	**N**	**Class %***	≥2 of 3 Low Prob. Q's (%)	Missing ≥1 TQ if shown (%)	Age Not Consistent (%)	Speed CE ≤4 sec. per Q. (%)	**Any of the four filters (%)**
1. Full AA	904	48.1	3.4	2.9	1.3	1.2	**7.9**
2. Full ANA	262	13.9	29.0	27.9	6.1	20.6	**48.5**
3. Attend Local Only	53	2.8	9.4	4.4	3.8	0.0	**13.2**
4. Attend Org. Only	53	2.8	18.9	15.9	9.4	7.5	**39.6**
5. Attend Price Only	51	2.7	11.8	9.5	2.0	3.9	**17.6**
6. Ignore Local Only	0	-	-	-	-	-	**-**
7. Ignore Price Only	240	12.8	11.3	5.9	2.1	2.5	**17.9**
8. Ignore Org. Only	317	16.7	3.8	3.1	0.0	0.9	**6.6**

Total	**1,880**	**100.0**	**8.9**	**7.3**	**2.2**	**4.3**	**15.9**

Note: Full ANA (65.0%), Attend Local Only (4.6%), Attend Organic Only (5.3%), and Ignore Price Only (8.7%) jointly encompass 83.6% of all individuals violating SARP.
Note: Full ANA (77.2%), Attend Local Only (6.2%), Attend Organic Only (6.7%), and Ignore Price Only (4.1%) jointly encompass 94.2% of all individuals violating SARP more than once.
*Class percentages will not perfectly align with output sample average probabilistic distributions since the highest class probability was designated to place each individual in a unique class.

The ANA composition shift in the ECLC fixed class assignments when quarantining metrics filter poor online panel respondents is shown in table 4. With the simplified CE design, it is possible to run this analysis separately for the total sample, "offending" (potentially quarantined) respondents, and passing respondents who were not identified as fraudulent or inattentive. Two scenarios are run, with the first being the full set of 8 possible linear utility functions. The second scenario controls only for utility functions not amenable to economic theory, namely Full ANA, non-attendance of the price attribute, and dual non-attendance of price and organic or price and local attributes (Scarpa et al. 2009).

There are significant improvements in both models' goodness-of-fit metrics after filtering (comparing the group of passing respondents who were not identified as fraudulent or inattentive with the total sample), with gains in adjusted R^2 and the scale-adjusted AIC. Offending respondents are five times more likely to fall into the Full ANA class, with over 40% of individuals probabilistically assigned in both scenarios. After filtering, full attendance increases by 8.4% and 10.6% and Full ANA decreases by 38.5% and 40.6% in the 8-class and 5-class scenarios, respectively. Interestingly, a disproportionately larger number of respondents failing quarantining metrics also seem to sort into the class which only attends to the organic attribute, and (in the 8-class scenario) a disproportionately smaller number sort into the class only ignoring the organic attribute. With an elevated level of inattentive and "speeding" respondents in the class only attending the organic attribute, this may reflect a quick processing strategy when seeing the organic attribute in the choice set. Additionally, offending respondents selecting randomly are less likely to have answer patterns resulting in insignificance of only one rather than two attribute βs, possibly explaining the lower incidence of ignoring price or organic only. Overall, the ANA sum-totals drop substantially after filtering, primarily driven by declining rates of Full ANA.

Changes in WTP before and after filtering are, by contrast, unremarkable with the ECLC model. Full sample organic and local WTP estimates are within 1% of the filtered subset. The intra-scenario proximity of WTP estimates is most likely due to the ECLC model's ability to cope with non-attendance resulting from poor respondent behavior, since censored latent classes are more likely to contain offending respondents. However, simply comparing offenders to non-offenders, the more theoretically sound 5-class model produces organic WTP estimates which are 23.6% higher for offenders. The organic WTP estimate is likely higher for offending respondents in the 8-class model after removing offenders since there are fewer offending respondents in the Ignore Price Only class. It is important to note that the significantly higher WTP estimates in the 8-class scenario are due to the fact that non-attendance of any attribute is censored, while the 5-class model permits non-attendance of qualitative attributes but not price.

Table 4. Latent Class Probability for Attribute Attendance (AA) and Attribute Non-Attendance (ANA)

Fixed Latent Classes	8 Latent Classes: Composition % (Full Linear Utility Function Possibilities)			5 Latent Classes: Composition % (Only Un-sound Classes Out)		
	Total Sample	Without Offenders	Offenders	Total Sample	Without Offenders	Offenders
1. Full AA	38.1%*	41.3%*	18.5%*	62.2%*	68.8%*	28.3%*
2. Full ANA	14.3%*	8.5%*	42.6%*	13.0%*	8.0%*	41.0%*
3. Attend Local Only	3.6%*	4.0%*	2.1%*	4.4%*	4.9%*	2.5%*
4. Attend Org. Only	4.2%*	2.9%*	13.0%*	5.8%*	3.5%*	15.6%*
5. Attend Price Only	6.1%*	6.2%*	5.5%*	-	-	-
6. Ignore Local Only[1]	0.0%	0.0%	2.2%	-	-	-
7. Ignore Org. Only	22.1%*	25.5%*	10.6%*	-	-	-
8. Ignore Price Only	11.5%*	11.6%*	5.6%*	14.6%*	14.9%*	12.6%*
L-L	-8,726.7	-6,513.2	-2,079.4	-9,199.9	-6,970.0	-2,107.9
Adjusted R^2	0.4847	0.5428	0.2282	0.4568	0.5107	0.2176
AIC/N	0.715	0.635	1.075	0.753	0.679	1.088
Observations	24,440	20,553	3,887	24,440	20,553	3,887
Sum-Total ANA Rates						
Price	33.6%	27.1%	63.3%	37.8%	31.3%	71.7%
Organic	46.1%	44.2%	60.8%	17.4%	12.9%	43.5%
Local	24.6%	17.7%	63.3%	18.8%	11.5%	56.6%
Willingness-To-Pay Estimates[2]						
WTP Organic	\$1.93/pint	\$1.96/pint	\$1.83/pint	\$1.12/pint	\$1.10/pint	\$1.36/pint

(vs. conventional) [95% CI]	[\$1.78-\$2.08]	[\$1.81-\$2.12]	[\$1.31-\$2.35]	[\$0.99-\$1.24]	[\$0.97-\$1.23]	[\$0.86-\$1.86]
WTP Local (vs. non-state U.S.) [95% CI]	\$1.29/pint [\$1.17-\$1.41]	\$1.29/pint [\$1.18-\$1.41]	\$1.32/pint [\$0.86-\$1.78]	\$1.07/pint [\$0.95-\$1.18]	\$1.07/pint [\$0.95-\$1.18]	\$1.07/pint [\$0.63-\$1.50]

Note: * denotes statistically significant classes (all at 99% confidence interval).
[1]Class 6 "Ignore Local Only" never significant and contained no individuals when exporting individual probabilistic class assignment.
[2]Confidence intervals constructed using the delta method (Hole 2007).

The SMXLM provides a different lens for analysis, with improvements in estimation exceeding that with the ECLC approach. Results are outlined in Table 5. Filtering respondents improves goodness-of-fit through both increased adjusted R^2 (0.468 to 0.531) and decreased scale-adjusted AIC (0.738 to 0.650). Importantly, the τ-variance parameter driving the individual specific standard deviation of the individual's error term (σ_i) is significantly higher in the offending population vis-à-vis passing respondents, at 1.901 and 1.102, respectively. The moderate decline in the overall τ estimate after filtering indicates a decrease in the identifiable and controllable poor respondent behavior can decrease resulting (and unnecessary) systemic heterogeneity.

Filtering poor respondents also drives a decline in the directly estimated WTP coefficients for both organic and local blueberry attributes. The disparity between offending and non-offending respondents is much more pronounced than with the previous ECLC model, as ANA censoring controls are not implemented with the scaled mixed logit model. The organic WTP estimate for offending respondents is nearly double that for passing respondents, with the filtering process decreasing organic WTP estimates by 12.3%. Local WTP is 18.4% higher for offending respondents vs. non-offending respondents, and filtering reduces overall local WTP estimates by 7.4%. With the popularity of mixed logit models and online research through web panels, this study suggests that the implications of unengaged and *satisficing* respondents has a measureable and significant impact on estimation results. The extent to which these poor respondents penetrate the study sample will naturally be proportional to the extent of this population's bias on overall estimated coefficients.

Table 5. Scaled Mixed Logit Model (SMXLM) Comparison

Coeff. in WTP Space, USD/pint	**Total Sample**	**Removing Offending Respondents**	**"Offending" Respondents**
Organic [vs. conventional] (st. err.)	2.335*** (0.062)	2.047*** (0.054)	3.959*** (0.208)
95% CI	[2.296 – 2.374]	[2.013 – 2.081]	[3.877 – 4.041]

<u>Local</u> [vs. non-state U.S.] (st. err.)	1.425*** (0.028)	1.319*** (0.022)	1.562*** (0.099)
95% CI	[1.418 – 1.432]	[1.313 – 1.325]	[1.558 – 1.566]
Standard Deviations			
Organic	3.080***	2.422***	2.569***
(st. err.)	(0.047)	(0.037)	(0.202)
Local	0.515***	0.422***	0.126
(st. err.)	(0.040)	(0.033)	(0.184)
τ-Variance Parameter (st. err.)	1.131*** (0.016)	1.102*** (0.014)	1.901*** (0.071)
$\bar{\sigma}$	0.977	0.974	0.824
(st. dev.)	(1.391)	(1.318)	(2.260)
L-L	-9,009.31	-6,676.48	-2,124.92
Adj. R^2	0.468	0.531	0.210
AIC/n	0.738	0.650	1.096
Observations (n)	24,440	20,553	3,887

Note: *** and ** indicate significance at the 99% and 95% confidence level. Standard errors are in parentheses, except for $\bar{\sigma}$ which displays the standard deviation. γ-parameter estimation is fixed in the WTP space specification. WTP coefficients are negative in program output, but converted to intuitive positive values.

Conclusions and Recommendations

Poor respondents who fraudulently enter surveys and/or have little to no engagement are present in many online panel samples and may introduce significant unnecessary variance into model estimation and bias WTP estimates. These damaging respondents are identifiable through real-time quarantining methods such as trap questions, low-probability screening questions, speeding metrics, and age verification strategies. This article illustrates that offenders are much more likely to be present in fixed latent classes, estimated through the ECLC approach, that correspond to utility functions which are contrary to economy theory and indicate no significant β estimates for CE attributes. Removing these unengaged respondents significantly decreases attribute non-attendance, especially full non-attendance, and improves goodness-of-fit metrics in all CE models investigated.

Similarly, potentially quarantined respondents exhibit considerably higher scale heterogeneity, as measured through the τ-variance parameter in the SMXLM, than those passing quarantining metrics. The expected value of the individual-specific standard deviation of the individual's error term is thus reduced after filtering quarantined respondents. Therefore, this study demonstrates that a moderate part of this systemic heterogeneity can be avoided simply through more vigilant sample management techniques. Furthermore, filtering poor respondents reduced WTP estimates by 12.3% and 7.4% for organic and local attributes, respectively, in the SMXLM. As designing the most robust models to estimate attribute WTPs is a goal of upmost importance for CE researchers, the findings from this study indicate that the outlined procedures may contribute to more conservative estimates. And while this WTP study utilizes a CE model as an example, sample data quality is likely measurably improved for most econometric approaches when inattentive and fraudulent respondents are properly quarantined.

References

Abidoye, B., H. Bulut, J. Lawrence, B. Mennecke, and A. Townsend. 2011. "U.S. Consumers' Valuation of Quality Attributes in Beef Products." *Journal of Agricultural and Applied Economics* 43(1):1-12.

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

Andreoni, J., and J. Miller. 2002." Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica* 70(2):737-753.

Baker, R., S. Blumberg, J. Brick, M. Couper, M. Courtright, J. Dennis, D. Dillman, M. Frankel, P. Garland, R. Groves, C. Kennedy, J. Krosnick, P. Lavrakas, S. Lee, M. Link, L. Piekarski, K. Rao, R.Thomas, and D. Zahs, D. 2010. "Research Synthesis: AAPOR Report on Online Panels." *Public Opinion Quarterly* 74(4):711-781.

Balcombe, K., A. Chalak, and I. Fraser. 2009. "Model Selection for the Mixed Logit with Bayesian Estimation." *Journal of Environmental Economics and Management* 57(2):226-237.

Bartels, J., and I. van den Berg. 2011. "Fresh Fruit and Vegetables and the Added Value of Antioxidants: Attitudes of Non-, Light, and Heavy Organic Food Users." *British Food Journal* 113(1):1339-1352.

Carson, R.T., J.J. Louviere, D.A. Anderson, P. Arabie, D.S. Bunch, D.A. Hensher, R.M. Johnson, W.F. Kuhfeld, D. Steinberg, J. Swait, H. Timmermans, and J.B. Wiley. 1994. "Experimental Analysis of Choice." *Marketing Letters* 5(4):351-368.

Dillman, D., J. Smyth, and L. Christian. 2009. *Internet, Mail, and Mixed-mode Surveys: The Tailored Design Method.* Hoboken, NJ: Wiley & Sons.

Downes-Le Guin, T. 2005. "Satisficing Behavior in Online Panels." Paper presented at the MRA Annual Conference & Symposium, Chicago, IL, June.

Downes-Le Guin, T., R. Baker, J. Mechling, and E. Ruylea. 2012. “Myths and Realities of Respondent Engagement in Online Surveys.” *International Journal of Market Research* 54(5):1-21

Ellison, B., and J. Lusk. 2011. “Taxpayer Preferences for USDA Expenditures.” *Choices: The Magazine of Food, Farm, and Resource Issues* 26(2).

Gao, Z., L.A. House, X. Bi. 2013. “Data Quality of Online Survey and its Impact on Willingness to Pay Estimates.” Unpublished working paper. (Available from editor upon request).

Gao, Z., L. House, and J. Xie. 2013. “Online Survey Data Quality and its Implication for Willingness-to-Pay: A Cross-Country Comparison.” *Selected Paper prepared for presentation at the Agricultural & Applied Economics Association’s 2013 AAEA & CAES Joint Annual Meeting,* Washington, DC, USA, August 4-6. http://ageconsearch.umn.edu/bitstream/150777/2/AAEA-3002-Online%20Data%20Quality-Gao.pdf (Accessed February 21, 2014).

Gao, Z., and T. Schroeder. 2009. “Effects of Label Information on Consumer Willingness-to-Pay for Food Attributes.” *American Journal of Agricultural Economics* 91(3): 795-809.

Gao, Z., T. Schroeder, and X. Yu. 2010. “Consumer Willingness-to-Pay for Cue Attribute: The Value Beyond Its Own.” *Journal of International Food and Agribusiness Marketing* 22(1):108-124.

Greene, W., and D. Hensher. 2010. “Does Scale Heterogeneity across Individuals Matter? An Empirical Assessment of Alternative Logit Models.” *Transportation* 37(3):413-428.

Hartl, J., and R. Herrmann. 2009. “Do They Always Say No? German Consumers and Second-Generation GM Foods.” *Agricultural Economics* 40:551-560.

Hensher, D.A. 2006. “How Do Respondents Process Stated Choice Experiments? Attribute Consideration under Varying Information Load.” *Journal of Applied Econometrics* 21:861-878.

Hensher, D.A., and J. Rose. 2009. “Simplifying Choice through Attribute Preservation or Non-Attendance: Implications for Willingness to Pay.” *Transportation Research Part E: Logistics and Transportation Review* 45:583-590.

Hensher, D.A., J. Rose, and W. Greene. 2005. “The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes.” *Transportation* 32(3):203-222.

Hensher, D.A., J. Rose, and W. Greene. 2012. “Inferring Attribute Non-Attendance Form Stated Choice Data: Implications for Willingness to Pay Estimates and a Warning for Stated Choice Experiment Design.” *Transportation* 39:235-245.

Hess, S., and D.A. Hensher. 2010. “Using Conditioning on Observed Choices to Retrieve Individual-Specific Attribute Processing Strategies.” *Transportation Research Part B: Methodological* 44(6):781-790.

Hole, A. 2010. “A Discrete Choice Model with Endogenous Attribute Attendance.” University of Sheffield, Sheffield, UK.

Hole, A.R. 2007. “A comparison of approaches to estimating confidence intervals for willingness to pay measures.” *Health Economics.* 16: 827-840.

Hougaard, J., T. Tjur, and L. Østerdal. 2012. "On the Meaningfulness of Testing Preferences Axioms in Stated preference discrete choice experiments." *European Journal of Health Economics.* 13:409-417.

Houthakker, H. 1950. "Revealed Preference and the Utility Function." *Economica* 17:159-174.

Jones, M., L.A. House, Z. Gao. 2013. "Respondent Screening and SARP: Testing Quarantining Methods for Better Data Quality in Web Panel Surveys." Unpublished working paper. (Available from editor upon request).

Kinsey, J., R. Harrison, D. Degeneffe, G. Ferreira, and S. Shiratori. 2009. "Index of Consumer Confidence in the Safety of the United States Food System." *American Journal of Agricultural Economics* 91(5):1470-1476.

Krosnick, J. 1991. "Response Strategies for Coping with Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213-236.

Lancsar, E., and J. Louviere. 2006. "Deleting 'Irrational' Responses from Discrete Choice Experiments: A Case of Investigating or Imposing Preferences?" *Health Economics* 15:797-811

Loose, S., A. Peschel, and C. Grebitus. 2013. "Quantifying Effects of Convenience and Product Packaging on Consumer Preferences and Market Share of Seafood Products: The Case of Oysters." *Food Quality and Preference* 28:492-504.

Loose, S., and H. Remaud. 2013. "Impact of Corporate Social Responsibility Claims on Consumer Food Choice: A Cross-Cultural Comparison." *British Food Journal* 115(1):142-166.

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

McIntosh, E., and M. Ryan. 2002. "Using Discrete Choice Experiments to Derive Welfare Estimates for the Provision of Elective Surgery: Implications of Discontinuous Preferences." *Journal of Economic Psychology* 23:367-382.

Miller, J. 2006. "Research Reveals Alarming Incidence of 'Undesirable' Online Panelists." *Research Conference Report, September-October 2006, RFL Communications. Synthesized from presentation at Research Industry Summit: Improving Respondent Cooperation, Chicago, IL, USA Sept 28-29, 2006.*

Miller, J., and J. Baker-Prewitt. 2009. "Beyond 'Trapping' the Undesirable Panelist: The Use of Red Herrings to Reduce Satisficing." Paper presented at 2009 CASRO Panel Quality Conference, New Orleans, LA, February. **[Burke Pub. No. CP51].**

Moon, W., S. Balasubramanian, and A. Rimal. 2011. "Health Claims and Consumers' Behavioral Intentions: The Case of Soy-based Food." *Food Policy* 36:480-489.

Olynk, N., G. Tonsor, and C. Wolf. 2010. "Consumer Willingness to Pay for Livestock Credence Attribute Claim Verification." *Journal of Agricultural and Resource Economics* 35(2):261-280.

Samuelson, P. 1938. "A Note on the Pure Theory of Consumer Behavior." *Economica* 5:61-71.

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

Scarpa, R., T. Gilbride, D. Campbell, and D.A. Hensher. 2009. "Modeling Attribute Non-Attendance in Choice Experiments for Rural Landscape Valuation." *European Review of Agricultural Economics* 36(2):151-174.

Schaafsma, M., R. Brouwer, and J. Rose. 2012. "Directional Heterogeneity in WTP Models for Environmental Valuation." *Ecological Economics* 79:21-31.

Simon, H. (1957). *Models of man.* New York: Wiley.

Smith, R., and H. Hofma-Brown. 2005. "Assessing the Quality of Data from Online Panels: Moving Forward with Confidence." White Paper. Harris Interactive, Rochester, NY.

Smith, R., and H. Hofma-Brown. 2006. "Comparing Metrics and Assessing Claims." *Proceedings of the ESOMAR World Research Conference, Panel Research* 317:9-21.

Tonsor, G. 2011. "Consumer Inferences of Food Safety and Quality." *European Review of Agricultural Economics* 38(2):213-235.

Tonsor, G., and C. Wolf. 2010. "Drivers of Resident Support for Animal Care Oriented Ballot Initiatives." *Journal of Agricultural and Applied Economics* 42(3):419-428.

Uzea, A., J. Hobbs, and J. Zhang. 2011. "Activists and Animal Welfare: Quality Verifications in the Canadian Pork Sector." *Agricultural Economics* 62(2):281-304.

USDA (U.S. Department of Agriculture). 2013. *Agricultural Marketing Service Database: Weekly Advertised Fruit & Vegetable Retail* Prices. http://www.marketnews.usda.gov/portal/fv?paf_dm=full&reportConfig=true&paf_gear_id=1200002&dr=1&repType=wiz&startIndex=1&type=Retail&portal=fv&class=ALL&region=NATIONAL&organic=ALL&commodity=ALL (accessed February 21, 2014)

van Doorn, J., and P. Verhoef. 2011. "Willingness to Pay for Organic Products: Differences between Virtue and Vice Foods." *International Journal of Research in Marketing* 28:167-180.

Varian, H. 1982. "The Non-Parametric Approach to Demand Analysis." *Econometrica* 50(4):945-973.

Willems, P., R. van Ossenbruggen, and T. Vonk. 2006. "The Effects of Panel Recruitment and Management on Research Results." Paper presented at the ESOMAR Panel Research Conference, Barcelona, Spain, Nov. 29 – Dec. 1.

Wolf, C., G. Tonsor, and N. Olynk. 2011. "Understanding U.S. Consumer Demand for Milk Production Attributes." *Journal of Agricultural and Resource Economics* 36(2):326-342.

FOOTNOTES / ENDNOTES

[1] For recent examples, please see Abidoye et al. 2011; Gao and Schroeder, 2008; Gao, Schroeder, and Yu 2010; Hartl and Herrmann 2009; Loose, Peschel, and Grebitus 2013; Loose and Remaud 2013; Moon, Balasubramanian, and Rimal 2011; Olynk, Tonsor, and Wolf, 2010; Schaafsma, Brouwer, and Rose 2012; Tonsor 2011; Uzea, Hobbs, and Zhang 2011; van Doorn and Verhoef 2011; Wolf, Tonsor, and Olynk 2011; .

[2] Lancsar and Louviere (2006) discuss and criticize preference-based approaches such as monotonicity, local non-satiation, convexity, and continuity.

[3] Previous research on blueberry consumption using a choice experiment showed only 6% of respondents selecting “none”, and most selected “none” when presented a choice between two “imported” production locations. The “imported” location was thus intentionally excluded from this design to simplify the experiment and encourage participation.