EVALUATING DATA QUALITY IN DISCRETE CHOICE EXPERIMENTS

by

Courtney Bir

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Agricultural Economics

West Lafayette, Indiana

December 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Nicole Olynk Widmar, Chair
Department of Agricultural Economics

Dr. Michael S. Delgado
Department of Agricultural Economics

Dr. Nathanael M. Thompson
Department of Agricultural Economics

Dr. Elizabeth Yeager
Department of Agricultural Economics

Approved by:
Dr. Nicole Olynk Widmar

ACKNOWLEDGMENTS

I would like to thank my committee members for their advice and mentorship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Although data collection through discrete choice experiments conducted using surveys are commonly used in research, aiming to improve data quality is still serviceable and necessary. Three distinct experiments were conducted with the objectives of improving data quality by better tailoring experiments to market conditions as well as decreasing complexity and fatigue. First, consumer willingness-to-pay (WTP) for yogurt attributes was estimated using a survey targeted to be nationally representative of the US. A novel approach was used to allow for self-selection into the choice experiment for commonly purchased types of yogurt. On average, respondents were willing-to-pay a positive amount for requiring pasture access and not permitting dehorning/disbudding for both traditional and Greek yogurt. Respondents had positive WTP for Greek yogurt labeled free of high fructose corn syrup, and were willing-to-pay more for low-fat yogurt when compared to nonfat for both yogurt types.

Second, a new WTP data collection method, employing component discrete choice experiments in place of traditional larger experimental designs, was proposed and compared to the traditional method to elicit yogurt consumer's WTP for attributes in yogurt. The new WTP data collection method was designed with the objective of decreasing complexity by having respondents participate in fewer choice scenarios. Incidences of attribute non-attendance (ANA), a potential simplifying heuristic that results from complexity, occurred less frequently for all attributes in the new WTP data collection method with one exception. Exhibiting ANA for any attribute was negatively correlated with the time respondents took to complete the choice experiment.

Finally, through the use of a new best-worst scaling (BWS) data collection method, consumer preferences for fluid dairy milk attributes were elicited and results as well as measures of data quality were compared to the traditional method of BWS. Nine attributes of fluid milk were included in this study: container material, rbST-free, price, container size, fat content, humane handling of cattle, brand, required pasture access for cattle, and cattle fed an organic diet. The top (price) and bottom (container material) attributes in terms of relative ranking did not change between the new BWS data collection method and the traditional BWS method. The new BWS data collection method resulted in fewer incidences of ANA for all attributes except one. There

was not a statistical difference in the number of transitivity (an axiom of consumer theory) violators, between the new and traditional BWS methods.

CHAPTER 1. BACKGROUND AND MOTIVATION

‘Data! Data! Data!’ he cried impatiently. ‘I can’t make bricks without clay.’

-Sherlock Holmes

Data is the all-important foundation of research and analytics, but how is data quality impacting the ‘bricks’ we are making? Survey instruments are a popular method of collecting data regarding consumer preferences, especially for preferences surrounding hypothetical products that may not (yet) be present in the marketplace. Such hypothetical decisions, which come with their own set of biases (Hensher, Rose, and Greene, 2005), are often elicited through discrete choice experiments. Discrete choice experiments, whether they include best-worst scaling designs, or willingness-to-pay elicitations, can vary in terms of the number of attributes, presentation, and attribute presentation combinations. All of the elements of a particular design can impact data quality. Other issues such as the length of surveys (Galesic and Bosnjak, 2009), simplification heuristics, and fatigue (Hensher, Rose, and Greene, 2005; Hess and Hensher, 2010) can also impact the quality of data used to inform industry and policy decision makers.

The careful consideration for product attributes is an important consideration when designing choice experiments (DeShazo and Fermo, 2002). Although products may be closely related, differences in preference between products can still occur. Olynk and Ortega (2013) evaluated US consumer willingness-to-pay (WTP) for USDA, retail, or industry verified pasture access, antibiotic use, and rbST in ice cream and yogurt. A statistically significant and positive WTP was found for all attributes and verifiers studied (Olynk and Ortega, 2013). Additionally, statistically significant differences were found between the WTP for the attributes in the two products, with respondent’s willing-to-pay more for animal welfare improving attributes in yogurt when compared to ice cream (Olynk and Ortega, 2013). Based on this analysis, although other dairy products have been studied, it is possible that the results of those studies cannot be extended to other dairy products. For example, Napolitano et al. (2010) employed a WTP model to evaluate Italian consumers’ preferences for pecorino cheese and found that consumers prefer organic, and similarly Krystallis and Chryssohoidis (2005) found a preference for organic feta cheese in Greek consumers. Other studies have focused on specific animal welfare related attributes. Bir et al. (2019b) studied a combination of animal welfare and product attributes in fluid dairy milk. What if very similar products have different pricing structure, such as traditional and Greek yogurt?

What differences exist between consumers and non-consumers of the products studied? The assumptions made regarding the product studied, the impact of respondent consumption as it relates to product familiarity and preferences, and the attributes included may all impact results, and the overall quality of data collected.

Beyond the products studied, and how purchasing behavior may impact data quality and results, all survey data collection efforts must consider fatigue. People make thousands of decisions daily, and the impact of that bombardment by decision making has been focused upon by the business world. The 44th US president Barack Obama only wears blue or grey suits, Facebook founder Mark Zuckerberg has a signature gray T-shirt, and both busy leaders claim removing wardrobe choices helps them combat decision fatigue (Baer, 2015). Although psychologists may debate whether decision making can really cause fatigue, or if minimizing decisions simply provides more daily structure and less uncertainty (Hobson, 2017), the idea that having made multiple decisions has some impact on human behavior has become mainstream. The term decision fatigue has even had an update, with articles and discussions focused on choice minimalism as opposed to avoiding fatigue (Hobson, 2017). This awareness has penetrated home organization through Netflix with Marie Kondo's methods tailored to minimize clutter, and decrease decision making long term (Belikova, 2019). Even the structure of the workday is now being questioned. Previously, an 8 hour work day was established to protect factory workers, but as the type of work done has changed, it has been found that people have the greatest focus for approximately 50 minutes with short 17 minute breaks (Bradberry, 2016). With new developments regarding the impact of decision making, fatigue, and optimal performance, revisiting such ideas in data collection and choice experiments is timely. With increasing attention on data quality and whether surveys are eliciting "true" responses and/or preferences, revisiting fatigue and decision making in survey based data collection is necessary.

There are several methods that could be employed to measure or compare data quality, two of which include measures of attribute non-attendance and transitivity violations. Simplifying heuristics are often associated with attribute-non-attendance (ANA). ANA can occur when a respondent simplifies the choice task by ignoring an attribute (Scarpa et al., 2012). This simplification becomes an issue because choice experiments are based on random utility theory, and ignoring an attribute may alter the marginal effect (Scarpa et al, 2012). Methods to account for ANA, including stated and inferred ANA, have been widely used in willingness-to-pay (WTP) literature (Carlsson et al., 2007; Napolitano et al., 2008; Olynk et al., 2010). WTP can either increase or decrease when accounting for ANA (Layton and Hensher, 2008). Inferred ANA requires the evaluation of individual level standard deviations in relation to the individual coefficients called the coefficient of variation, and employs a threshold to determine occurrences of ANA (Hess and Hensher, 2010). Widmar and Ortega (2014) employed inferred ANA to determine the effects using different thresholds for stated ANA while evaluating WTP for various livestock production attributes for dairy and ham products. The different thresholds investigated in their study (1, 2 and 3) resulted in only small changes in the WTP estimates (Widmar and Ortega, 2014). Stated ANA requires an additional question asking respondents directly if they ignored any of the attributes included in the choice experiment (Hole, 2011). Methods employed to determine incidences of inferred ANA in BWS have recently been adapted to BWS (Lai et al., 2019; Bir et al., 2019a). Lai et al. (2019) included accounting for ANA in both WTP and BWS in their comparison of the two discrete choice methods, WTP and BWS. Bir et al. (2019a) included incidences of ANA as one method to evaluate differences between two BWS designs. In addition to design choices that impact response efficiency, accounting for ANA may help improve data quality.

Discrete choice models, including BWS are rooted in random utility theory (Scarpa et al., 2012; Johnson et al., 2013). The axioms of consumer theory, including transitivity (Varian, 1978), can be used as one measure of data quality in choice experiments. Transitivity implies that if A is preferred to B and B is preferred to C, then A must be preferred to C (Varian, 1978). Issues related to response efficiency, such as respondent fatigue, confusion or misunderstanding, and inattention potentially resulting from hypothetical bias may result in violations of transitivity. Lagerkvist (2013) found the possibility of transitivity violations in choice experiments, but did not determine the number of transitivity violations or violators. Bir et al. (2019a) developed a Python algorithm employing directed graphs to determine the number of violations of transitivity at the individual level in four BWS designs. They found the number of violations differed between BWS experimental designs studied (namely two designs that presented either three or eight attributes per choice scenario), as well as by the order in which the BWS designs were presented to respondents. Accounting for the number of transitivity violations and the impact of those violations on results, in conjunction with other methods, is one way to measure data quality in choice experiments. The main objective of this dissertation is to evaluate and consider new strategies to improve the quality of the data that results from discrete choice experiments.

1.1 Background and motivation references

Baer, D. (2015). The scientific reason why Barack Obama and Mark Zuckerberg wear the same outfit every day. Business insider. Retrieved from: https://www.businessinsider.com/barack-obama-mark-zuckerberg-wear-the-same-outfit-2015-4

Belikova, I. (2019). Neuroscience and psychology findings support the KonMari method. Massive Science. Retrieved from: https://massivesci.com/articles/science-of-tidying-up-brain-marie-kondo/

Bir, C., Widmar, N. O., Etumnu, C., Slipchenko, N., Polcyn, A., & Wolf, C. A. (2019a). Development and implementation of transitivity violation detection algorithm in discrete choice experiments. Manuscript submitted for publication.

Bir, C., Widmar, N. O., Wolf, C., & Delgado, M. S. (2019b). Traditional attributes moo-ve over for some consumer segments: Relative ranking of fluid milk attributes. *Appetite, 134*, 162-171.

Bradberry, T. (2016). Why the 8-hour workday doesn't work. Forbes. Retrieved from: https://www.forbes.com/sites/travisbradberry/2016/06/07/why-the-8-hour-workday-doesnt-work/#7c131c3836cc

Carlsson, F., Frykblom, P., & Lagerkvist, C. J. (2007). Consumer willingness to pay for farm animal welfare: mobile abattoirs versus transportation to slaughter. *European Review of Agricultural Economics*, *34*, 321-344.

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and management*, *44*(1), 123-143.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349–360.

Greene, W. H. (2012). NLOGIT Version 6 Reference Guide. Plainview, NY: Econometric Software, Inc.

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

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

Hobson, N. (2017). The (updated) real reason why famous founders wear the same clothes every day. Forbes. Retrieved from: https://www.forbes.com/sites/nickhobson/2017/12/15/the-updated-real-reason-why-famous-founders-wear-the-same-clothes-every-day/#7d40affe17ec

Hole, A. R. (2011). Attribute non-attendance in patients' choice of general practitioner appointment. Presented at the 2nd international choice Modelling Conference, Leeds. Retrieved from: http://tinyurl.com/yayepff3

Johnson, F. R, Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., Bresnahan, B.W., Kanninen, B., & Bridges, J. F. P. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health*, *16*(1), 3-13.

Krystallis, A., & Chryssohoidis, G. (2005). Consumers' willingness-to-pay for organic food: Factors that affect it and variation per organic product type. *British Food Journal*, *107*(5), 320-343.

Lai, J., Widmar, N. O., Bir, C., Wolf, C. M. (2019). Comparing Attribute Non-Attendance Between Best-Worst Scaling and Willingness-to-pay Choice Experiments. Manuscript submitted for publication.

Lagerkvist, C. J. (2013). Consumer preferences for food labelling attributes: Comparing direct ranking and best–worst scaling for measurement of attribute importance, preference intensity and attribute dominance. *Food Quality and Preference*, *29*(2), 77-88.

Layton, D., & Hensher, D. A. (2008). Aggregation of common-metric attributes in preference revelation in choice experiments and implications for willingness-to-pay. Institute of transport and logistics studies, the Australian key centre in transport and logistics management, the University of Sydney, Sydney Australia. ISSN 1832-570X.

Napolitano, F., Braghieri, A., Piasentier, E., Favotto, S., Naspetti, S., & Zanoli, R. (2010). Cheese liking and consumer willingness-to-pay as affected by information about organic production. *Journal of Dairy Research*, *77*(3), 280-286.

Olynk, N. J., & Ortega, D. L. (2013). Consumer preferences for verified dairy cattle management practices in processed dairy products. *Food Control*, *30*(1), 298-305.

Olynk, N. J., Tonsor, G. T., & Wolf, C. A. (2010). Consumer willingness to pay for livestock credence attributes claim verification. *Journal of Agricultural and Resource Economics*, *35*(2), 261-280.

Scarpa, R., Zanoli, R., Bruschi, V., & Naspetti, S. (2012). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics*, *95*(1), 165-180.

Varian, H. R. (1978). Microeconomic analysis. New York, NY: W. W. Norton.

Widmar, N. J. O., & Ortega, D. L. (2014). Comparing consumer preferences for livestock production process attributes across products, species, and modeling methods. *Journal of Agricultural and Applied Economics*, *46*(3), 375-391.

CHAPTER 2. IT'S ALL GREEK TO ME: US CONSUMER DEMAND FOR YOGURT ATTRIBUTES

2.1 Greek and traditional yogurt willingness-to-pay introduction

Yogurt, one of the many popular products that can be made from dairy milk, is an ancient food, with health benefits appearing in writing as early as 6000 B.C. (Fisberg and Machado, 2015). By today's standards yogurt is broadly defined as a fermented milk product that provides digested lactose (Fisberg and Machado, 2015). Yogurt can contain active cultures or be heat treated, which kills the beneficial live and active yogurt cultures (NYA, 2019). Today's yogurt products come in a variety of fat contents including: low-fat (made from low-fat milk or part skim milk with between 0.5 and 2 percent milk fat) and nonfat (made from skim milk with less than 0.5 percent milkfat) (NYA, 2019). Yogurt also comes in many forms such as traditional, Greek, whipped, drinkable, and fruit on the bottom. Greek yogurt, which has more protein than traditional yogurt, has fueled the growth of the yogurt market after first being introduced to the US in 2007 by Chobani (Meyer, 2019).

Yogurt consumption in the US lags behind other places such as Canada and Europe (Meyer, 2019; Watson, 2019). However, yogurt production in the US steadily increased from 912 million pounds in 1989 to over 4.7 billion pounds in 2014 (Figure 1). Since 2014, yogurt production in the US has slowed with 2018 production reaching just below 4.4 billion pounds (USDA NASS, 2019). Although this slight recent downward trend may be evidence that yogurt sales have begun to stagnate, yogurt companies are hoping to expand sales by creating more varieties and convincing people to eat yogurt on occasions beyond breakfast (Meyer, 2019). Additionally, expansion into more child friendly flavors and packaging such as candy flavorings and yogurt tubes is believed to result in future sales growth (Watson, 2019).

Although yogurt has been studied extensively from a health perspective, such as for immunologic effects (Meydani and Ha, 2000), effects on metabolism (Wang et al., 2013), and general gut health (Fisberg and Machado, 2015), little exists in the literature regarding consumer preferences for yogurt, and especially consumer preferences for Greek yogurt. Moro et al. (2015) studied Italian consumer willingness-to-pay (WTP) for probiotics and catechin-enriched yogurt and found a positive WTP for both attributes, while Olynk and Ortega (2013) found that US consumers had a positive willingness-to-pay for USDA, retail, or industry verified pasture access, antibiotic use, and rbST free in ice cream and yogurt. Additionally, statistically significant differences were found between the WTP for the attributes in the two products, with respondents willing-to-pay more, on a price adjusted percentage basis, for credence attributes in yogurt when compared to ice cream (Olynk and Ortega, 2013).

This work aims to build on previous findings by evaluating consumer WTP for attributes associated with pasture access, dehorning/disbudding of dairy cattle, and Greek and traditional yogurt that is labeled free of high fructose corn syrup and/or nonfat. Bir et al. (2019) used a best-worst scaling model to elicit consumer preferences for container material, rbST free, price, container size, fat content, humane handling, brand, required pasture access, and cattle fed an organic diet in fluid dairy milk. Top attributes in terms of preference revealed public interest in both product attributes themselves, such as fat content, as well as animal-related attributes, such as humane handling or requiring access to pasture (Bir et al., 2019). Other studied dairy production practices include dehorning, which is practiced on dairy operations to improve safety for both people and animals, decrease aggressive behavior, and decrease carcass damage; however, the process of dehorning is painful and often done without pain medication (AVMA, 2014). Twelve percent of respondents in a nationally representative 2017 survey indicated that they had altered their dairy consumption due to animal welfare concerns, and tail docking and dehorning were perceived as the least beneficial and most negative implications for dairy cattle welfare of those studied (Widmar et al., 2017). Although both are yogurt products, Greek and traditional yogurt are priced differently, typically come in different sized containers, and may be consumed for different reasons and by individuals from different demographic groups. In addition to animal welfare attributes, the yogurt attribute high-fructose corn syrup was also included in the WTP experiment. High-fructose corn syrup is a modified corn syrup that has an increased amount of fructose made from the enzymatic conversion of glucose, and is widely used as a sweetener (Merriam Webster, 2019). Although there are no scientific links between high-fructose corn syrup, obesity and other negative health effects in humans, some health conscious consumers have recently become skeptical and are avoiding high-fructose corn syrup in products (Parker and Nwosu, 2010).

The objectives of this study are to 1) develop background regarding consumer shopping/eating behaviors and evaluate when, why, and what type of yogurt products people in the US consume; 2) elicit and compare consumer WTP for traditional and Greek yogurt attributes; and 3) determine if differences in WTP for traditional and Greek yogurt attributes can be ascribed to demographics and shopping behavior. These results contribute to the literature by evaluating an understudied, but commercially important, dairy product – yogurt.

2.2 Greek and traditional yogurt willingness-to-pay methods

An online survey designed in Qualtrics was administered from June 11-21, 2019 to gather demographic information and dairy consumption behaviors, with a special focus on yogurt purchases and consumption. To further understand preferences for yogurt, two WTP experiments featuring Greek or traditional yogurt were included. Kantar, a company which hosts an opt-in online panel was used to obtain survey respondents: a total of 894 respondents completed the survey with the associated yogurt WTP experiments. The sample was targeted to be nationally representative of the US in terms of gender, income, education and geographical region of residence as defined by the US Census Bureau Regions and Divisions using quotas in Qualtrics (US Census Bureau, 2016). Using the test of proportions, the statistical representativeness of the survey respondents was evaluated for the targeted demographics.

To understand dairy shopping and consumption behavior, respondents were asked how much they spend on food in an average week, how frequently they eat out for each meal, what snacks they purchase outside the home, and what types of information they look for on packaging. Respondents were asked how frequently (if at all) they purchased specific yogurt products for their household. If they responded that they purchased the yogurt product at least monthly, they were asked additional specific questions about that product such as what type they purchased, and for what consumption purpose. Respondents were also asked a series of yogurt purchasing questions for their households, including if they had decreased yogurt consumption for any reason including: animal welfare/handling, health, budgetary, and food safety concerns. To develop an understanding of respondent perceptions of dairy cattle management practices, they were asked to indicate their level of agreement that practices such as docking (removal) of tails of dairy cows/cattle, confining dairy cows/cattle indoors, dehorning (removal of horns), and disbudding (removal of horn buds), decreased dairy cow/cattle welfare on a scale from 1 (very strongly agree) to 7 (very strongly disagree). Means and standard deviations were calculated, and the levels of agreement across the production practices were statistically compared using a t-test.

2.2.1 Willingness-to-pay for Greek and traditional yogurt

After responding to questions regarding household dairy consumption including yogurt, and prior to participating in a WTP choice experiment, respondents were asked to indicate the type of yogurt they most commonly shopped for: Greek yogurt, traditional yogurt, or neither. Only one selection was allowed. This question differs slightly from the demographics and general purchasing section of the survey because it asked about the individual, not the household's purchasing behavior, and the most common purchase. Additional proportion testing was conducted within demographic categories for those who purchased yogurt, and in particular purchased Greek or traditional yogurt. If the respondent selected Greek yogurt, they participated in a WTP choice experiment where they were asked to choose between two Greek yogurt purchasing scenarios or the option "I do not choose to purchase either option A or B". If they selected traditional yogurt, they participated in a WTP experiment of traditional yogurt purchasing scenarios. If the respondent selected neither, they were randomly assigned to participate in either the Greek or traditional WTP choice experiment. The SAS OPTEX program was used to design a main effects experiment and determine the specific combination of attribute levels seen by respondents in the choice experiment (SAS, 2014; Lusk and Norwood 2005). The specific design was chosen by maximizing D-efficiency, which was 85.49 for both the Greek and traditional yogurt designs. The final design resulted in 12 choice scenarios (questions), and respondents participated in all 12 choice scenarios for either Greek or traditional yogurt. With the exception of the attributes that were presented in each choice scenario, respondents were informed that the two traditional or two Greek yogurt products presented in each scenario had the same characteristics in terms of color, brand, and flavor. Information explaining each of the four attributes in the choice experiment were shown to respondents prior to presenting any questions and are available in Appendix A. The cheap talk script as proposed by Lusk (2003) is intended to minimize hypothetical bias, and was employed in both WTP experiments.

Attributes included were the same for the traditional and Greek yogurt WTP experiments: price, required pasture access or pasture access not required, dehorning/disbudding not permitted or dehorning/disbudding permitted, labeled free of high fructose corn syrup or no high fructose corn syrup labeling claim, and nonfat or lowfat. For Greek yogurt the prices presented per 5.3 oz. cup were $0.72, $1.00, and $1.29. For traditional yogurt the prices presented per 6.0 oz. cup were

$0.40, $0.79 and $1.14. Prices and cup sizes were determined by observing prices and available sizes in the marketplace in April, 2019.

Choice experiments are based in random utility theory. The probability that respondent *n* chooses alternative *i*, which represents maximizing utility (*U*) with deterministic component V_{nit}, *if* $U_{nit} > U_{njt}\ \forall j \neq i$ is represented by (Train, 2009):

$$P_{nit} = Prob\big(V_{nit} + \varepsilon_{nit} > V_{njt} + \varepsilon_{njt}; \forall_j \in C, \forall_j \neq i\big) \quad (1)$$

Given the underlying distribution of the error term, Equation 1 can be condensed through algebraic manipulation to:

$$P_{nit} = \frac{\exp(V_{nit})}{\sum_j \exp(V_{njt})}. \quad (2)$$

The random utility of a selection for either traditional or Greek yogurt is defined as:

$$V_{it} = \beta_1 Price_{it} + \beta_2 ReqPasture_{it} + \beta_3 NoDehorning_{it} + \beta_4 FreeOfFructose_{it} + \beta_5 Nonfat_{\ it} + \beta_6 Optout_{it} \quad (3)$$

where *Price* is the price a respondent is willing-to-pay for traditional or Greek yogurt, and *Optout* is a constant which represents the respondent's disutility from having to walk away from purchasing either Greek or traditional yogurt, *ReqPasture* is the effects coded term for required pasture access, *NoDehorn* is the effects coded term for dehorning/disbudding not permitted, *FreeOfFructose* is the effects coded term for labeled free of high fructose corn syrup and Nonfat is the effects coded term for nonfat as opposed to *lowfat*. For the RPL model, the mean WTP is calculated as the negative of the ratio of the coefficient for the particular attribute and the coefficient of price, for example the WTP for required pasture access can be calculated as:

$$WTP = -2\frac{\beta_2}{\beta_1}. \quad (4)$$

The -2 in Equation 4 accounts for the effects coding of the various levels of the attributes. In this experiment, all attributes had two levels and were coded with -1 and 1, unlike typical 0,1 dummy variable coding (Adamowicz et al., 1994). This coding prevents the left out dummy variable from being incorporated into the intercept, which occurs under the traditional dummy variable estimation setup (Adamowicz et al., 1994). The disutility in terms of dollars of walking away from the purchase of traditional or Greek yogurt (*OptOut*) is calculated as:

$$WTP = -\frac{\beta_6}{\beta_1}. \quad (5)$$

A likelihood-ratio (LR) test was used to determine if those who purchased and those who did not purchase Greek (traditional) yogurt could be pooled for analysis in the Greek (traditional) yogurt models (Louviere et al., 2000). Using the Krinsky and Robb method of parametric bootstrapping, ninety-five percent confidence intervals were determined for each attribute in each model to account for variability in estimation (Krinsky and Robb, 1986; Olynk and Ortega, 2013). Within each model, either traditional or Greek yogurt, overlapping confidence intervals were compared to determine if there were statistically significant differences between WTP for each attribute (Schenker and Gentleman, 2001). To statistically compare WTP for each attribute between the traditional and Greek models, the complete combinatorial method as outlined by Poe and Loomis (2005) was employed.

In order to better understand the relationship between demographics, shopping behavior, and WTP for both traditional and Greek yogurt, two seemingly unrelated regressions (SURs) were employed. SURs were used because it was likely that the error terms for the WTP estimates within the individual equations were correlated (Greene, 2013; Zellner, 1962). The same model structure was used for both Greek and traditional yogurt, defined as:

$$WTPReqPast = \beta_1 Male + \beta_2 BuysYogurt + \beta_3 AnimalWelfare + \beta_4 HighIncome + \beta_5 Child + \beta_6 CattleConfine$$

$$WTPNoDehorn = \beta_1 Male + \beta_2 BuysYogurt + \beta_3 AnimalWelfare + \beta_4 HighIncome + \beta_5 Child + \beta_6 CattleDehorn + \beta_6 CattleDisbud$$

$$WTPFreeOfFructose = \beta_1 Male + \beta_2 BuysYogurt + \beta_4 HighIncome + \beta_5 Child$$

$$WTPNonFat = \beta_1 Male + \beta_2 Milk + \beta_3 BuysYogurt + \beta_4 HighIncome + \beta_5 Child$$

$$WTPOptOut = \beta_1 Male + \beta_2 BuysYogurt + \beta_4 HighIncome + \beta_5 Child \qquad (6)$$

where *Male* indicates male gender, *BuysYogurt* indicates purchasing yogurt outside the home, *AnimalWelfare* indicates reading animal welfare labeling on milk, meat, or dairy products, *HighIncome* indicates an income of $75,000 or higher, *Child* indicates having a child in the household, *CattleConfine* indicates the level of agreement that confining dairy cows/cattle indoors decreases welfare, *CattleDehorn* indicates the level of agreement that dehorning (removal) of horns decreases welfare, *CattleDisbud* indicates the level of agreement that disbudding (removal) of horn buds decreases welfare, and *Milk* indicates that the respondent purchases 2 percent, 1 percent, or fat free dairy milk. The Breusch-Pagan test was conducted to determine if the individual equations within the SUR models were correlated (Breusch and Pagan, 1980).

2.3 Greek and traditional yogurt willingness-to-pay results

The demographics of the 894 respondents who completed the survey instrument and WTP experiments statistically varied from the US census in only a few demographic categories (Table 1). The percentage of respondents who were 18-24 (9%) was statistically lower than the US census (13%). A lower percentage of respondents did not graduate from high school (6%), and a higher percentage of respondents graduated from high school (33%), when compared to the US census, 13% and 28% respectively. A higher percentage of respondents were from the South (39%), and a lower percentage of respondents were from the Midwest (21%), when compared to the US census, 21% and 38% respectively.

On average, respondents spent $176.86 (SD 479.28, n=879) per week on total food consumption including at home, in restaurants, take-outs, etc. Approximately half of respondents (56%) did not consume breakfast outside the home during a typical week where as 25% consumed 1-2 breakfasts outside the home (Table 2). Forty percent of respondents did not consume lunch out in a typical week, while 37% consumed at least 1-2 meals. A lower percentage (33%) of respondents consumed zero dinners out in a given week, while 42% consumed 1-2 meals out. Juxtaposed to breakfast consumption outside the home, only 43% of respondents never purchased coffee or tea outside the home in a typical week, and 35% purchased coffee or tea 1-2 times. Fifty-four percent of respondents never purchased ice cream outside the home, 67% never purchased juice/smoothies outside the home, and 71% of respondents never purchased yogurt cups outside the home. For all products, the next commonly selected category was 1-2 times. Most respondents indicated that they read information when reviewing meat, egg, or milk product packaging. Sixty-six percent of respondents indicated they looked for price, 60% looked for the product expiration or "sell by" date, 38% looked for nutritional information, and only 14% looked for animal welfare information. Sixty-seven percent of respondents purchased at least some fluid dairy milk (skim, lowfat 1%, reduced fat 2%, and whole milk) in a typical week. Of those, 6% purchased 3 or more gallons of fluid dairy milk a week. Of the respondents who purchased fluid dairy milk (n=717), 61% of respondents purchased 2% reduced fat milk, and 1% purchased lowfat milk or fat free skim milk. The mean level of agreement that confining dairy cows/cattle indoors decreased dairy cattle welfare was statistically lower than all other dairy cattle welfare practices, indicating respondents were most concerned about this practice (Figure 2). There were no statistically significant differences between the production practices of docking (removal) tails of dairy cows/cattle, dehorning (removal of horns), and disbudding (removal of horn buds).

Five hundred and forty seven (61% of) respondents indicated that they purchased/consumed yogurt themselves (Table 1). A statistically higher percentage of women (65%) purchased yogurt for themselves when compared to men (56%). Lower percentages of respondents 55 and older purchased yogurt when compared to all other age categories. Higher percentages of respondents with incomes of over $50,000 purchased yogurt when compared to the lower income categories. Lower percentages of respondents who did not graduate from high school and lower percentages of respondents from the Midwest purchased yogurt when compared to all other education and region categories, respectively. Traditional yogurt was purchased most commonly by 347 respondents, 39% of all respondents or 63% of yogurt shoppers. Higher percentages of respondents aged 25-34 purchased traditional yogurt when compared to all other age groups. A lower percentage of those who did not graduate from high school purchased traditional yogurt when compared to all other education categories. Greek yogurt was purchased most commonly by 200 respondents, 22% of the total sample or 37% of yogurt consumers. Lower percentages of respondents with an income of less than $50,000 purchased Greek yogurt when compared to all other income categories. Fifty-three percent of respondents indicated that someone in their household purchased yogurt at least 4 times a year, 42% indicated they did not, and 5% indicated that they did not know. Of those respondents the majority (63%) had not reduced yogurt consumption over the past three years. Twenty percent indicated that they had reduced yogurt consumption due to food safety concerns, and only 13% had reduced consumption due to animal welfare/handling concerns (Table 2).

Respondents who purchased yogurt (n=475) were asked which specific types of yogurt products their household purchased as well as the frequency (Table 3). Few respondents did not recognize or did not know the yogurt products included, ranging from 3% to 9%. Forty-five percent of yogurt purchasers purchased individual traditional yogurt cups weekly, while 41% purchased individual Greek yogurt cups monthly. For traditional yogurt in large tubs, Greek yogurt in large tubs, drinkable yogurt, and yogurt tubes, high percentages of yogurt buying respondents never purchased these products (ranging from 51% to 54%). High percentages of respondents who purchased individual traditional yogurt cups (n=332), purchased fruit-on the bottom cups (51%), low-fat (40%), and blended (42%). For individual Greek yogurt cup purchasers (n=332), 47%

purchased fruit-on the bottom cups, and 36% purchased blended cups. A high percentage of respondents (39%) who purchased yogurt in large tubs (n=196) purchased plain yogurt. Drinkable yogurt that contained active yogurt cultures or was blended was purchased by a high percentage of respondents who purchased drinkable yogurt (n=182), 41% and 36% respectively. High percentages of respondents who purchased yogurt tubes (n=179) purchased regular fat content (41%) or blended (40%). For all yogurt products, low percentages of respondents, ranging from 10% to 18%, purchased heat-treated products. The most common uses for all yogurt products studied were to consume as a snack, to consume with/as breakfast, and to consume with/as lunch (Table 4). For Greek yogurt in large tubs, an additional category, for use in a recipe was selected by 81% of Greek yogurt in large tub purchasers (n=196). Fed to a pet as a treat or meal, was selected the least for all products studied; however, it was still selected by 23% of individual traditional yogurt cup purchasers, which was the lowest product use (n=392).

2.3.1 Willingness-to-pay for Greek and traditional yogurt production and production process attributes

For both the traditional and Greek yogurt models, those who did and did not consume that type of yogurt could not be pooled (LR=143, df=20, p<0.001 for Greek, and LR=233, df=20, p<0.001 for traditional). Respondents were willing-to-pay a statistically significant amount for all attributes for the traditional yogurt, with the exception of yogurt labeled free of high fructose corn syrup (Table 5). Respondents had a higher WTP for both required pasture access ($0.46) and dehorning/disbudding not permitted ($0.58) when compared to fat content in traditional yogurt. There was not a statistically significant difference in mean WTP for pasture access and dehorning/disbudding not permitted for traditional yogurt. For the attribute nonfat, respondents were willing-to-pay -$0.14 when compared to lowfat in traditional yogurt. Walking away from a traditional yogurt purchasing opportunity (Optout) resulted in a disutility of $1.50.

For Greek yogurt, respondents were willing-to-pay a higher amount for required pasture access ($1.18) and dehorning/disbudding not permitted ($1.19) when compared to the attributes labeled free of high fructose corn syrup, and nonfat. Respondents were not willing-to-pay a positive amount for nonfat Greek yogurt, relative to lowfat yogurt. Walking away from the Greek yogurt purchasing opportunity (Optout) resulted in a disutility of $1.13. When comparing the normalized WTP between traditional and Greek yogurt, respondents were willing-to-pay more for required pasture access, and dehorning/disbudding not permitted in Greek yogurt than in traditional yogurt.

For both the SUR for traditional and Greek WTP the residuals of the individual equations were correlated as indicated by a Breusch-Pagan test for independent equations of less that <0.001 for both models. The correlations between each equation for the two models are available in Appendix B. All individual models were statistically significant for both Greek and traditional yogurt, with the exception of the model for WTP for nonfat traditional yogurt (Table 6). Being male decreased WTP for required pasture access and dehorning/disbudding not permitted for both traditional and Greek yogurt. Buying yogurt outside the home decreased WTP for required pasture access, and dehorning/disbudding not permitted for both traditional and Greek yogurt. Additionally, buying yogurt outside the home increased WTP for nonfat Greek yogurt. Looking at animal welfare labeling did not have a statistically significant impact on WTP for required pasture access or dehorning/disbudding not permitted for traditional yogurt, but increased WTP for required pasture access in Greek yogurt. Having an income above $75,000 decreased WTP for required pasture access for both traditional and Greek yogurt. Additionally, having an income above $75,000 increased WTP for labeled free of high fructose corn syrup in traditional yogurt. Having a child in the household decreased WTP for dehorning/disbudding not permitted in traditional and Greek yogurt and decreased the disutility experienced from walking away from a traditional or Greek yogurt buying opportunity. Furthermore, having a child decreased the WTP for required pasture access in Greek yogurt. Interestingly, the level of agreement that confining dairy cows/cattle indoors decreases dairy cattle welfare did not statistically significantly impact WTP for required pasture access in traditional or Greek yogurt. The level of agreement that disbudding (removal) of horn buds decreases welfare did not have a statistically significant impact on WTP for dehorning/disbudding not permitted for either traditional or Greek yogurt.

2.4 Greek and traditional yogurt willingness-to-pay discussion

Although the survey respondents had a slightly higher education, which is a common occurrence in online surveys (Szolnoki et al., 2013), in general the demographics closely matched the US census. Reported weekly food shopping spending including at home, in restaurants, and take-outs, was somewhat higher than previous findings. McKendree et al. (2013) found that on average respondents spent $132.77 on food, which is lower than the $176.86 reported in this study. The bureau of labor statistics reported average food expenditure (at home and eating out) in 2018 as $7,923, which is $152 (USDL, 2019). There was a 2.5% increase in food expenditure between 2017 and 2018 (USDL, 2019). Therefore, the higher total weekly spending found in this analysis may reflect a general increase in food prices over the past six years, among other possible factors.

In a December 2016 Gallup poll, 38% of adult Americans did not eat dinner out in an average week, and 45% ate 1-2 dinners out (Saad, 2017). This study's findings that 33% of respondents consumed zero dinners out and 42% consumed 1-2 dinners out corroborated previous statistics. Interestingly, although ice cream and yogurt are both dairy products, a higher percentage of respondents purchased ice cream outside the house at least once a week when compared to yogurt. Consumers are eating less ice cream, but the ice cream they are consuming is higher-end (Leathan, 2017). Traditionally, ice cream may be seen as a treat, while yogurt is part of a meal or healthy snack (Olynk and Ortega, 2013). As people become more conscious of sugar consumption, people have begun to consume premium yogurts as a household dessert in place of ice cream (Leathan, 2017).

Yogurt did not become mainstream in the US until the 1970s, previously a product mostly sold in health food stores (Meyer, 2019; Davis et al., 2010). Sixty-one percent of respondents in this analysis indicated that they purchased yogurt. Unsurprisingly, lower percentages of those 55 and older purchased yogurt, as perhaps it wasn't part of their diet earlier in life. A higher percentage of women consumed yogurt when compared to men. Douglas et al. (2013) found that consuming a snack of Greek yogurt (with 24 g. protein) reduced hunger, increased fullness, and delayed further eating in healthy women when compared to lower protein snacks. Women, especially postmenopausal women, require more calcium, which dairy products including yogurt provide, to combat bone loss that occurs when bone breakdown exceeds formation (ODS, 2019). Although there were not differences in consumption of Greek yogurt specifically for women and men, higher percentages of respondents with an income over $50,000 purchased yogurt. Based on their results using a cross sectional analysis and Neilson data, Davis et al., (2010) proposed that price and consumer income were the main drivers of yogurt demand.

Although fewer respondents purchased Greek yogurt most commonly in this study, Greek yogurt accounts for around 45% of dollar sales for the yogurt industry (Watson, 2019). Both Greek and traditional yogurt were purchased by more respondents as individual yogurt cups when compared to large tubs. Considering the percentage of respondents who consume yogurt as a snack (which may be outside the home), purchasing the portable version makes sense. Additionally, higher percentages of respondents used tubs of yogurt (either Greek or traditional) in recipes when compared to individual cups. Yogurt naturally contains calcium and potassium and is often fortified with vitamin D, all nutrients listed as nutrients of concern in Dietary Guidelines of America (Webb, 2014). The incorporation of yogurt in recipes may be due to its health benefit, but further research would be needed to understand respondents' reasoning behind the incorporation of yogurt in their diet. Heat treating, which prolongs the shelf-life of yogurt (Speck, 1977), kills the beneficial live and active yogurt cultures (NYA, 2019). Additionally, heat treating also inactivates lactase, which renders the previously edible yogurt inedible by those with lactose intolerance (Speck, 1977). Likely due to the benefits associated with non-heat treated yogurt, and the prevalence of non-heat treated yogurt in the marketplace, few respondents purchased heat treated yogurt of any kind.

Only 14% of respondents looked for animal welfare improving information, which was interesting considering a positive WTP was found for what some believe are animal welfare improving practices, requiring pasture access, and not permitting dehorning/disbudding. In the SUR, looking at animal welfare labeling was not statistically significant for either WTP for required pasture access or WTP for not permitting dehorning/disbudding. More research is needed to determine how consumers are getting the information needed to consciously select credence attributes of dairy cow production systems. It is possible that consumers research brands prior to shopping, or are influenced by commercials that promote credence attributes for specific brands or products.

Thirteen percent of respondents had reduced yogurt consumption due to animal welfare/handling concerns in this study, similar to the 12% of respondents who had altered their general dairy consumption due to animal welfare concerns found by Widmar et al. (2017). Despite many respondents not changing their consumption patterns, the mean level of agreement that confining dairy cow/cattle indoors decreases welfare was slightly lower than neutral 3.96 and was of greatest concern when compared to the other practices studied. Widmar et al. (2017) asked respondents to indicate on a scale from 1 (extremely negative impact) to 7 (extremely positive impact) the impact of dairy production attributes including access to pasture (mean score 5.6), tail docking (mean score of 3.8), and dehorning on the welfare of dairy cattle (mean score of 3.8). Required pasture access and dehorning/disbudding not permitted had positive estimated mean

WTP for traditional and Greek yogurt in this analysis. Despite differences in the Likert scale results, there were no statistically significant differences between the WTP for required pasture access or dehorning/disbudding not permitted for either yogurt type. For these particular products, being concerned about the production practice did not result in increased WTP in the SUR, illustrating how researchers should be careful when interpreting the implications of different preference elicitation methods. Similarly, Ochs et al. (2019) found differences between Likert scales, forced ranking, and best-worst scaling when studying US resident perceptions of laying hen welfare. Lagerkvist (2013) proposed that differences found between best-worst scaling and Likert responses may be due to a lack of topic knowledge on the part respondents.

Consumers were willing-to-pay a positive amount for both required pasture access and dehorning/disbudding not permitted for both traditional and Greek yogurt. Similarly, Olynk and Ortega (2013) found positive WTP for pasture access in both ice cream and yogurt, and Olynk et al. (2010) found positive WTP for pasture access in fluid dairy milk and pork chops. Although cows prefer to be in the barn during certain weather conditions given free choice, pasture access helps reduce mastitis and improve lameness problems (Von Keyserlingk et al., 2009). In the US Department of Agriculture's National Animal Health Monitoring System Dairy 2014 report, 94.3% of dairy operations surveyed disbudded or dehorned their heifer calves. One method to avoid the need to dehorn calves is to select for polled (naturally hornless) cattle. Thompson et al. (2017) estimated the costs of incorporating polled genetics into a breeding program to range from $0-$26/head. Determining if making such changes is economically profitable for dairy farms is difficult, as the main product is fluid dairy milk, which is a component of various dairy products for which respondents may have varying WTP for dehorning/disbudding not permitted. Being male decreased WTP for both pasture access and dehorning/disbudding not permitted. Increased female concern for farm animal welfare is well documented in the literature and was also found by Morgan et al. (2016), Vanhonacker et al. (2007), and McKendree et al. (2014).

Studies have been mixed regarding the health effects of high fructose corn syrup; however, while studying rats, Bocarsly et al. (2010) found that high fructose corn syrup resulted in a higher weight gain when compared to rats consuming the same amount of calories with less high fructose corn syrup. A simple Google search indicated that consumers concerned with health are worried about high fructose corn syrup with such headlines as "8 'health' foods that contain high fructose corn syrup", "23 surprising foods with high fructose corn syrup", and "Top 7 foods with hidden high-fructose corn syrup" (Elliott, 2016; Eat this, not that!, 2016; Donsky, 2019). All of the mentioned lists include yogurt, many of which include tips for purchasing yogurt without high fructose corn syrup. WTP for labeled free of high fructose corn syrup was statistically significant for Greek yogurt, but not for traditional yogurt. The difference between traditional and Greek yogurt is mainly protein content. Perhaps those respondents who primarily purchased Greek yogurt are more concerned about health related attributes as opposed to other attributes such as taste.

Respondents had a preference for lowfat traditional and Greek yogurt when compared to nonfat, as demonstrated by the negative WTP for nonfat yogurt. In Greek yogurt, higher fat content results in increased firmness and denseness, with full-fat yogurts characterized with the highest levels (Desai and Drake, 2013). The industry is beginning to push higher fat content yogurts into the marketplace with some success (Watson, 2019), although lowfat and nonfat options made up the bulk of what was available in stores in 2019. Increased fat content in "light" products can mitigate the negative lingering taste of aspartame, which is often used to decrease caloric count (King, 2000). Interestingly, purchasing 2%, 1%, or fat free dairy milk was not a statistically significant indicator of yogurt fat content preference for either traditional of Greek yogurt in the SUR models. Consumers have a preference for reduced fat/lowfat milk when compared to whole and nonfat milk, with high consumption of 2% reduced fat milk (WMMB, 2017; Harwood and Drake, 2018; Bir et al., 2019); however, these preferences are not emulated in their WTP for fat content in yogurt.

2.5 Greek and traditional yogurt willingness-to-pay conclusion

Yogurt, an understudied dairy product, was purchased by a high percentage of survey respondents (61%). Differences were found between the demographics of those who did and did not purchase yogurt. For example, a statistically higher percentage of women (65%) purchased yogurt for their own consumption when compared to men. Although Greek yogurt has risen in popularity, traditional yogurt was still commonly purchased by more respondents, 39% of all respondents commonly purchased traditional yogurt, while 22% of respondents commonly purchased Greek yogurt. Greek yogurt which has higher protein levels than traditional yogurt, is sold in different sized containers than traditional yogurt, and is priced higher.

It has been established in the literature that WTP can differ for the same attributes in different products, even if they are made from the same or closely related ingredients. This research differs from previous studies by customizing the WTP experiment that respondents participated in based on the type of yogurt they commonly purchase. Respondents were assigned to the choice experiment for the yogurt type, either traditional or Greek, they most commonly purchased. Those who were not purchasers of either product were randomly assigned to one of the choice experiments, but in the analysis those who were not yogurt shoppers were found unable to be pooled with shoppers, indicating that being an experienced buyer impacted buying behaviors exhibited. Consumers had a positive WTP for required pasture access and not permitting dehorning/disbudding for traditional and Greek yogurt. Respondents were willing-to-pay a positive amount for labeled free of high fructose corn syrup for Greek yogurt, but not traditional yogurt. Respondents were not willing-to-pay a positive amount for nonfat when compared to lowfat in traditional or Greek yogurt when compared to nonfat, as demonstrated by the negative WTP for nonfat yogurt. A SUR was employed to better understand the relationship between demographics, shopping behavior, preferences, and WTP for Greek and traditional yogurt. Looking at animal welfare labeling was not statistically significant for either WTP for required pasture access or WTP for dehorning/disbudding not permitted, indicating research is needed to determine how consumers are selecting credence attributes. For Greek and traditional yogurt, being concerned about dehorning/disbudding or pasture access as evaluated in a Likert scale, did not result in increased WTP in the SUR. This exemplifies why researchers should be careful when interpreting the implications of different preference elicitation methods.

2.6 Greek and traditional yogurt tables and figures

Table 1. Demographics, comparison to US census, and comparison of yogurt consumers.

	Sample analysis		Yogurt analysis percentage within demographic		
Demographic Variable	**Percentage of respondents n=894**	**US Census**	**Yogurt purchase[1], percentage of yogurt purchasers within demographic n=547**	**Traditional[1] Yogurt purchasers n=347**	**Greek[1] yogurt purchasers n=200**
Gender					
Male	46[2]	49	56[3]a[4]	37a [5]	20[6]a
Female	54	51	65b	41a	24a
Age					
18-24	9ǂǂ	13	68ac	42a	27a
25-34	16	18	78a	58b	20a
35-44	18	16	66c	40a	26a
45-54	18	17	62c	43a	18a
55-65	18	17	50b	26c	23a
65 +	21	19	51b	29c	22a
Income					
$0-$24,999	26ǂ	22	51a	36ab	15a
$25,000-$49,999	25	23	52a	32b	21ab
$50,000-$74,999	17	17	71b	45a	26b
$75,000-$99,999	10	12	72b	46a	26b
$100,000 and higher	21ǂ	26	70b	42a	28b
Education					
Did not graduate from high school	6ǂǂ	13	36a	20a	17ab
Graduated from high school, Did not attend college	33ǂǂ	28	58b	41b	17a
Attended College, No Degree earned	18ǂ	21	58b	37b	21bc
Attended College, Associates or Bachelor's Degree earned	32ǂ	27	65bc	38b	27bc
Attended College, Graduate or Professional Degree earned	12	12	75c	44b	31c
Region					
Northeast	18	18	62ab	36a	26a
South	39ǂǂ	21	61ab	40a	21a
Midwest	21ǂǂ	38	54b	37a	17a
West	22	24	68a	41a	27a

ǂPercentage of respondents is statistically different than the percentage of the US census at the 0.05 level ǂǂ<0.001 level

[1]Respondent reported purchase of yogurt for self, as opposed to household.

[2]For example, read this as 46% of respondents were male

[3]For example, read this as 56% of men purchased yogurt

[4]Matching letters indicate the percentage for that demographic within the demographic category are not statistically different. Differing letters indicate the percentage for that demographic within the demographic category are statistically different. For example, the percentage of men who purchase yogurt is statistically different than the percentage of women.

[5]For example, read this as 37% of men purchased traditional yogurt

[6]For example, read this as 20% of men purchased Greek yogurt

Table 2. Respondent consumption and shopping habits.

Number of meals eaten out (either take out, restaurant, or cafeteria setting), percentage of respondents n=894					
	Zero meals	1-2 meals	3-4 meals	5-6 meals	All 7 meals
Of the 7 breakfasts in a given week	56	25	10	3	5
Of the 7 lunches in a given week	40	37	14	5	4
Of the 7 dinners in a given week	33	42	15	5	5
Number of times in a typical week respondents purchase the following items outside the home, percentage of respondents n=894					
	Never	1-2 times	3-4 times	5-6 times	7 or more
Coffee or tea	43	35	11	5	5
Juice/smoothies	67	21	6	4	2
Ice cream outside the home	54	35	7	2	2
Yogurt cup outside the home	71	17	7	4	1
Percentage of respondents who assess the following pieces of information when reviewing meat, egg or milk product packaging percentage of respondents n=894					
Nutritional information					38
Price					66
Food safety information					26
Animal welfare information					14
Local food labelling					21
Product expiration "sell by" date					60
Other					3
None					14
Reasons respondents have decreased yogurt consumption over the past three years, percentage of respondents n=475					
Animal welfare/handling concerns					13
Health concerns					18
Budgetary concerns					17
Food safety concerns					20
I have not reduced yogurt consumption					63
Other					5

Table 3. Respondent's household purchasing behavior of specific yogurt products.

Frequency household purchase of specific yogurt products, percentage of respondents whose households purchase yogurt n=475

Yogurt Product	Weekly	Monthly	Never	I do not know this product/Have never heard of it
Individual traditional yogurt cups	45	37	15	3
Individual Greek yogurt cups	29	41	27	3
Traditional yogurt in large tubs	21	23	51	6
Greek yogurt in large tubs	16	25	51	8
Drinkable yogurt	18	21	54	8
Yogurt tubes (ex. Go-Gurt)	15	23	53	9

Type of specific yogurt product respondent's household's purchase, percentage of purchasers of the particular product

Yogurt Product	Plain	Fruit-on the bottom	Whipped	Non-fat	Low-fat	Regular fat content	Contains active yogurt cultures	Heat-treated	Blended
Individual traditional yogurt cups n=392	22	51	27	15	40	29	33	10	42
Individual Greek yogurt cups n=332	27	47	23	20	33	33	33	10	36
Traditional yogurt in large tubs n=208	34	31	23	17	30	32	30	15	32
Greek yogurt in large tubs n=196	39	26	20	22	24	31	32	16	28
Drinkable yogurt n=182	30			23	32	34	41	18	36
Yogurt tubes (ex. Go-Gurt) n=179	29			20	32	41	31	18	40

Note: Greyed out boxes indicate options that were not available to respondents for that particular product. They were not included because for that particular yogurt product the option was not available in the market place (did not exist/did not make sense).

Table 4. Reasons and frequency of respondent's household's purchase of yogurt products, percentage of respondents whose household purchases each product.

Product	Reason for consuming	Daily	Weekly	At least monthly	Uses product	Never
Individual traditional yogurt cups n=392 (44% of respondents)	To consume as a snack	40	30	23	93	7
	To consume with/as breakfast	22	31	22	75	25
	To consume with/as lunch	25	25	23	73	27
	to consume with/as dinner	13	22	14	48	52
	For use in a recipe	19	17	17	52	48
	Given to a pet as a treat or a meal	11	6	6	23	77
	Other	13	19	7	39	61
Individual Greek yogurt cups n=332 (37% of respondents)	To consume as a snack	40	30	23	92	8
	To consume with/as breakfast	23	36	20	78	22
	To consume with/as lunch	26	24	25	75	25
	to consume with/as dinner	16	24	14	55	45
	For use in a recipe	23	15	18	57	43
	Given to a pet as a treat or a meal	14	7	7	29	71
	Other	14	23	7	43	57
Traditional yogurt in large tubs n=208 (23% of respondents)	To consume as a snack	48	24	17	88	12
	To consume with/as breakfast	30	37	16	83	17
	To consume with/as lunch	38	19	24	80	20
	to consume with/as dinner	24	31	14	69	31
	For use in a recipe	30	28	19	77	23
	Given to a pet as a treat or a meal	18	11	9	38	63
	Other	23	27	13	63	37
Greek yogurt in large tubs n=196 (22% of respondents)	To consume as a snack	52	20	15	88	12
	To consume with/as breakfast	30	39	17	86	14
	To consume with/as lunch	38	22	20	80	20
	to consume with/as dinner	22	31	18	71	29
	For use in a recipe	39	22	20	81	19
	Given to a pet as a treat or a meal	18	9	12	39	61
	Other	22	32	10	64	36
Drinkable yogurt n=182 (20% of respondents)	To consume as a snack	54	29	14	97	3
	To consume with/as breakfast	27	45	18	90	10
	To consume with/as lunch	39	29	20	87	13
	to consume with/as dinner	27	36	12	75	25
	For use in a recipe	38	19	18	75	25
	Given to a pet as a treat or a meal	21	10	12	43	57
	Other	23	39	10	72	28
Yogurt tubes (ex. Go-Gurt) n=179 (20% of respondents)	To consume as a snack	54	26	17	97	3
	To consume with/as breakfast	30	40	16	85	15
	To consume with/as lunch	40	25	25	89	11
	to consume with/as dinner	25	37	15	77	23
	For use in a recipe	38	18	17	73	27
	Given to a pet as a treat or a meal	24	10	11	45	55
	Other	22	35	11	68	32

Table 5. Random parameters logit coefficients, standard errors, and willingness-to-pay for traditional and Greek yogurt.

	Traditional Yogurt n=347				**Greek Yogurt n=200**				**P-value comparing WTP between traditional and Greek yogurt[3]**
Attributes	**Coefficient (SE)**	**Standard deviation (SE)**	**Mean WTP[1]**	**WTP 95% confidence interval**	**Coefficient (SE)**	**Standard deviation (SE)**	**Mean WTP[2]**	**WTP 95% confidence interval**	
Required pasture access	0.275*** (0.042)	0.353*** (0.048)	$0.46	[$0.31, $0.61]	0.612*** (0.090)	0.566*** (0.075)	$1.18	[$0.75, $1.77]	0.0052
Dehorning/ disbudding not permitted	0.349*** (0.042)	0.357*** (0.046)	$0.58	[$0.43, $0.75]	0.618*** (0.080)	0.453*** (0.068)	$1.19	[$0.79, $1.79]	0.0319
Labeled free of high fructose corn syrup	0.042 (0.033)	0.014 (0.078)	$0.07	[-$0.04, $0.18]	0.205*** (0.057)	0.159 (0.125)	$0.39	[$0.15, $0.65]	0.0157
Nonfat	-0.087** (0.036)	0.080 (0.124)	-$0.14	[-$0.26, -$0.02]	-0.170** (0.067)	0.485*** (0.078)	-$0.32	[-$0.61, -$0.06]	0.8066
Optout	-1.793*** (0.195)	3.025*** (0.183)	-$1.50	[-$1.83, -$1.17]	-1.208*** (0.330)	3.615*** (0.300)	-$1.13	[-$1.74, -$0.57]	0.0178
Price	-1.197*** (0.092)				-1.070*** (0.170)				

* statistically significant at the 0.10 level, **0.05 level, and *** at the <0.001 level

[1]Prices presented to respondents for the traditional yogurt WTP choice experiment were $0.40, $0.79, $1.14 per 6 oz cup, mean $0.78

[2]Prices presented to respondents for the Greek yogurt WTP choice experiment were $0.72, $1.00, and $1.29 per 5.3 oz cup, mean $1.00

[3]Mean prices for the traditional and Greek WTP choice experiments were used to normalize prices prior to comparison

Table 6. Seemingly unrelated regressions of willingness-to-pay for traditional and Greek yogurt attributes and demographic/shopping characteristics.

		Traditional yogurt[1] n=347			Greek yogurt[2] n=200		
Attribute model	Demographic/shopping characteristic	Coefficient	SE	P-Value	Coefficient	SE	P-Value
WTP for required pasture access	Male	-0.157	0.063	0.013	-0.468	0.163	0.004
	Buys yogurt outside the home[3]	-0.260	0.071	<0.001	-0.675	0.176	<0.001
	Looks at animal welfare labeling	0.010	0.014	0.480	0.319	0.108	0.003
	Income above $75,000	-0.142	0.069	0.041	-0.329	0.166	0.047
	Has a child	-0.110	0.069	0.112	-0.467	0.182	0.010
	Level of agreement that confining[4] dairy cows/cattle indoors decreases welfare	0.000	0.002	0.990	0.009	0.020	0.658
	Constant	0.700	0.053	<0.001	1.761	0.157	<0.001
WTP for dehorning/ disbudding not permitted	Male	-0.160	0.072	0.026	-0.468	0.140	0.001
	Buys yogurt outside the home[3]	-0.253	0.081	0.002	-0.392	0.151	0.010
	Looks at animal welfare labeling	-0.018	0.053	0.739	0.011	0.037	0.774
	Income above $75,000	-0.122	0.079	0.122	-0.234	0.143	0.101
	Has a child	-0.197	0.079	0.012	-0.329	0.157	0.036
	Agreement that dehorning (removal) or horns decreases welfare[4]	0.010	0.022	0.658	-0.023	0.019	0.218
	Agreement that disbudding (removal) of horn buds decreases welfare[4]	-0.001	0.022	0.967	0.028	0.019	0.156
	Constant	0.830	0.071	<0.001	1.651	0.117	<0.001
WTP for Labeled free of high fructose corn syrup	Male	0.044	0.027	0.103	0.070	0.080	0.381
	Buys yogurt outside the home[3]	-0.046	0.030	0.122	-0.254	0.086	0.003
	Income above $75,000	0.065	0.029	0.026	0.087	0.081	0.286
	Has a child	-0.028	0.029	0.340	-0.119	0.089	0.183
	Constant	0.058	0.022	0.008	0.432	0.064	<0.001
WTP for nonfat	Male	0.038	0.040	0.336	-0.183	0.120	0.128
	Purchases 2%, 1% or fat free dairy milk	-0.018	0.032	0.569	-0.003	0.090	0.975
	Buys yogurt outside the home[3]	-0.000	0.044	0.996	0.316	0.130	0.015
	Income above $75,000	-0.003	0.044	0.942	0.017	0.122	0.887
	Has a child	-0.064	0.044	0.141	0.204	0.134	0.129
	Constant	-0.111	0.038	0.004	-0.450	0.112	<0.001
Opt-Out	Male	-0.365	0.231	0.115	-0.960	0.424	0.024
	Buys yogurt outside the home[3]	-1.534	0.259	<0.001	-1.809	0.458	<0.001
	Income above $75,000	-0.135	0.254	0.594	-0.300	0.432	0.488
	Has a child	-0.727	0.254	0.004	-1.303	0.474	0.006
	Constant	-0.489	0.190	0.010	0.317	0.342	0.355

[1]P-value for individual traditional yogurt models (top to bottom) <0.001, <0.001, 0.002, 0.6028, <0.001
[2]P-value for individual Greek yogurt models (top to bottom) <0.001, <0.001, 0.002, 0.009, <0.001
[3]Respondent purchased yogurt outside the home at least 1-2 times a week
[4]Level of agreement was indicated on a scale of 1 (very strongly agree) to 7 (very strongly disagree)

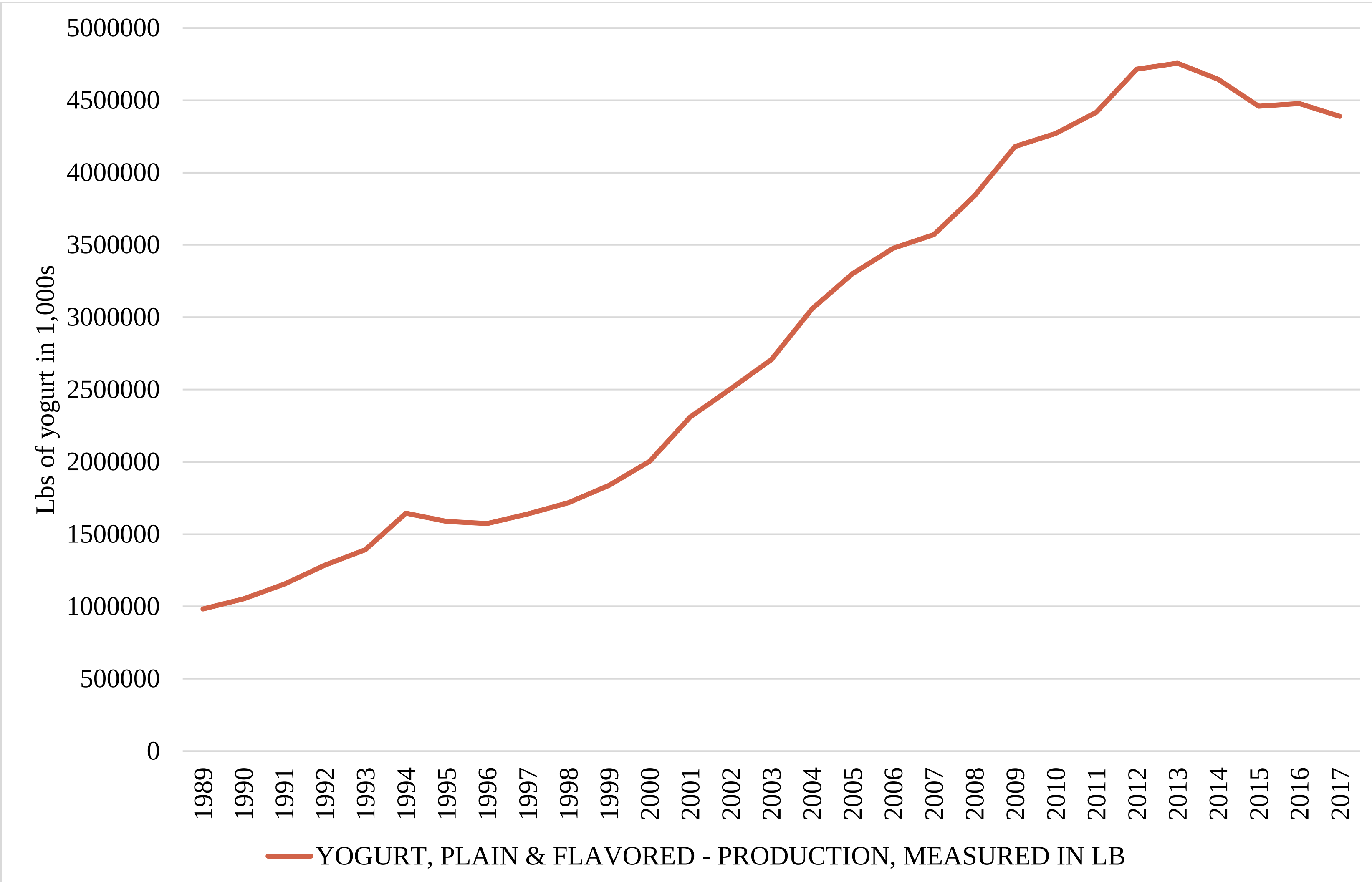

Figure 1. Plain and flavored yogurt production in the United States measured in pounds (USDA NASS 2019).

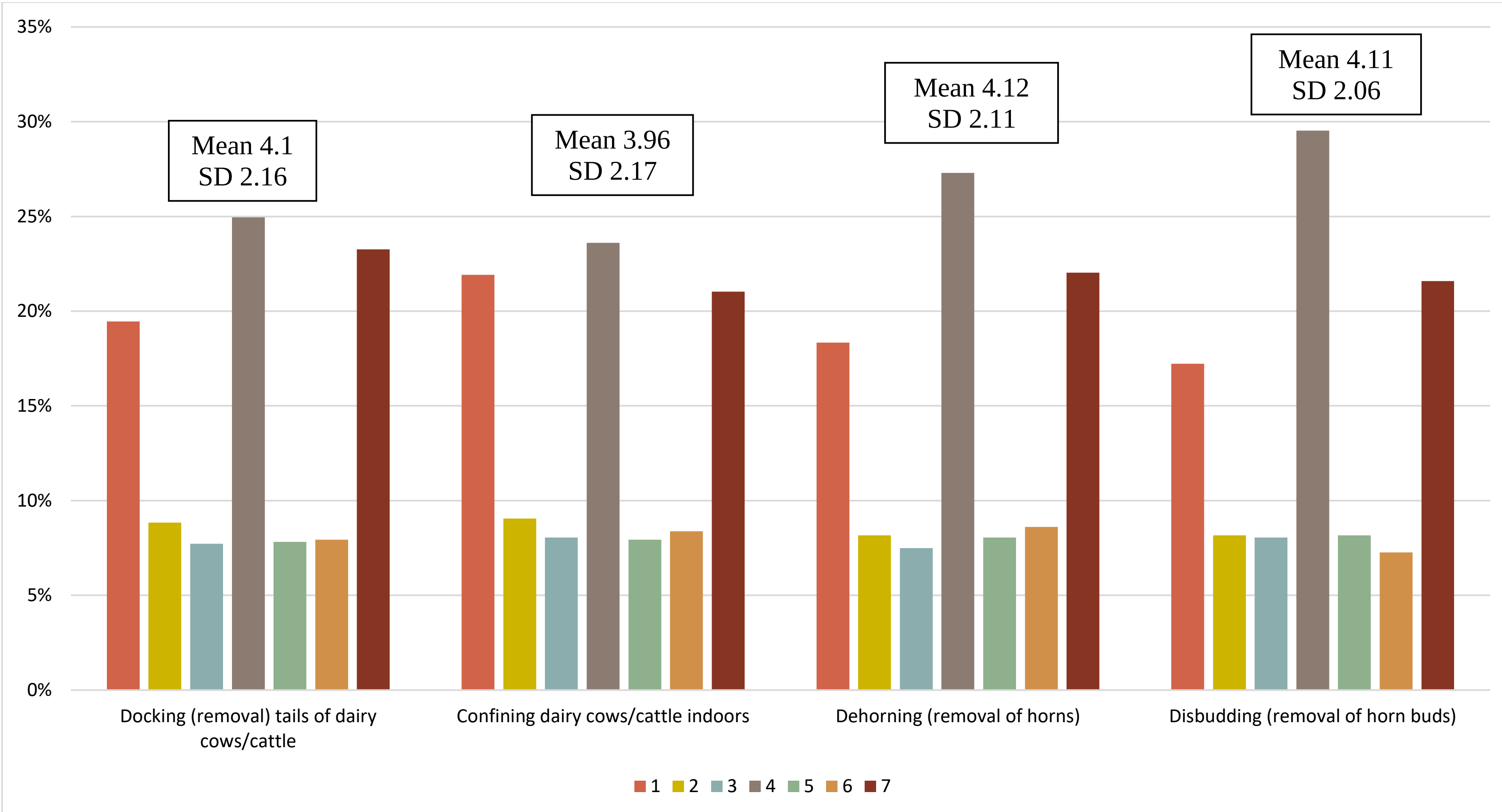

Figure 2. Respondent's level of agreement that the following practices reduces dairy cattle welfare 1 (very strongly agree) to 7 (very strongly disagree). n=894

.

2.7 Greek and traditional yogurt willingness-to-pay references

Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics Management, 26*(3), 271-292.

AVMA. (2014). Welfare Implications of Dehorning and Disbudding Cattle. Retrieved from: https://www.avma.org/KB/Resources/LiteratureReviews/Pages/Welfare-Implications-of-Dehorning-and-Disbudding-Cattle.aspx

Bir, C., Widmar, N. O., Wolf, C., & Delgado, M. S. (2019). Traditional attributes moo-ve over for some consumer segments: Relative ranking of fluid milk attributes. *Appetite, 134*, 162-171.

Bocarsly, M. E., Powell, E. S., Avena, N. M., & Hoebel, B. G. (2010). High-fructose corn syrup causes characteristics of obesity in rats: increased body weight, body fat and triglyceride levels. *Pharmacology Biochemistry and Behavior*, *97*(1), 101-106.

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies, 47*, 239–253.

Davis, C. G., Blayney, D. P., Muhammad, A., Yen, S. T., & Cooper, J. C. 2010. A cross-sectional analysis of US yogurt demand. *Journal of food distribution research, 41*, 36-45.

Desai, N. T., Shepard, L., & Drake M. A. (2013). Sensory properties and drivers of liking for Greek yogurts. *Journal of dairy science*, *96*(12), 7454-7466.

Donsky, A. (2019). Top 7 foods with hidden high-fructose corn syrup (HFCS). Naturally Savvy. Retrieved from: https://naturallysavvy.com/eat/top-7-foods-with-hidden-high-fructose-corn-syrup/

Douglas, S. M., Ortinau, L. C., Hoertel, H. A., & Leidy, H. J. (2013). Low, moderate, or high protein yogurt snacks on appetite control and subsequent eating in healthy women. *Appetite*, *60*, 117-122.

Eat this, not that! (2016). 23 surprising foods that contain high fructose corn syrup. Healthy Eating. Retrieved from: https://www.eatthis.com/foods-with-high-fructose-corn-syrup/

Elliott, B. (2016). 20 foods with high-fructose corn syrup. Healthline. Retrieved from: https://www.healthline.com/nutrition/20-foods-with-high-fructose-corn-syrup

Fisberg, M., & Machado, R. (2015). History of yogurt and current patterns of consumption. *Nutrition reviews*, *73*(1), 4-7.

Greene, W. (2013). Econometric Analysis (7th edition). Pearson Education Inc., Upper Saddle River NJ, USA.

Harwood, W. S., & Drake, M. A. (2018). Identification and characterization of fluid milk consumer groups. *Journal of Dairy Science, 101*(10), 8860-8874.

King, S. C., Lawler, P. J., & Adams, J. K. (2000). Effect of aspartame and fat on sweetness perception in yogurt. *Journal of food science*, *65*(6), 1056-1059.

Krinsky, I. & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, *68*(4), 715-719.

Lagerkvist, C. J. (2013). Consumer Preferences for Food Labelling Attributes: Comparing Direct Ranking and Best–Worst Scaling for Measurement of Attribute Importance, Preference Intensity and Attribute Dominance. *Food Quality and Preference*, *29*, 77–88.

Leathan, X. (2017). Why sales of ice cream are melting away: Amount sold across the world falls by 2.6bn litres in just two years as consumers opt for healthier lifestyles. Retrieved from: https://www.dailymail.co.uk/news/article-4673252/Global-ice-cream-sales-melting-away.html

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). Stated Choice Methods: Analysis and Applications. Cambridge, UK: Cambridge University Press.

Lusk, J. L. (2003). Effects of cheap talk on consumer willingness-to-pay for golden rice. *American journal of agricultural economics*, *85*(4), 840-856.

Lusk, J. L., & Norwood, F. B. (2005). Effect of experimental design on choice-based conjoint valuation estimates. *American Journal of Agricultural Economics, 87*(3), 771–785.

McKendree, M.G., Widmar, N.O., Ortega, D. L., & Foster, K. A. (2013). Consumer preferences for verified pork-rearing practices in the production of ham products. *Journal of Agricultural and Resource Economics, 38*(3), 397-417.

Merriam Webster. (2019). High-fructose corn syrup. Merriam Webster online dictionary. Retrieved from: https://www.merriam-webster.com/medical/high%20fructose%20corn%20syrup

Meydani, S. N., & Ha, W. (2000). Immunologic effects of yogurt. *The American journal of clinical nutrition, 71*(4), 861-872.

Meyer, Z. (2019). Yogurt makers change culture look to expand beyond breakfast to grow sales. USA Today. Retrieved from: https://www.usatoday.com/story/money/2018/12/11/grow-yogurt-makers-change-their-culture-snacking-plant-based/2130539002/

Morgan, C. J., Croney, C. C., & Widmar, N. J. (2016). Exploring relationships between ethical consumption, lifestyle choices, and social responsibility. *Advances in Applied Sociology*, *6*, 199-216.

Moro, D., Veneziani, M., Sckokai, P., & Castellari, E., (2015). Consumer Willingness-to-pay for Catechin-enriched Yogurt: Evidence from a Stated Choice Experiment. *Agribusiness, 31*(2), 243-258.

National Yogurt Association (NYA). (2019). Yogurt Varieties. Retrieved from: http://aboutyogurt.com/index.asp?bid=27

Ochs, D., Wolf, C. A., Widmar, N. and Bir, C. (2019). Is There a "Cage-Free" Lunch in US Egg Production? Public Views of Laying-Hen Housing Attributes. *Journal of Agricultural and Resource Economics, 44*, 345-361.

Office of Dairy Supplements (ODS). (2019). Calcium Fact sheet for health professionals. NIH. Retrieved from: https://ods.od.nih.gov/factsheets/Calcium-HealthProfessional/

Olynk, N. J., Tonsor, G. T., & Wolf, C. A. (2010). Consumer willingness-to-pay for livestock credence attribute claim verification. *Journal of Agricultural and Resource Economics, 35*(2), 261-280.

Olynk, N. J., & Ortega, D. L. (2013). Consumer preferences for verified dairy cattle management practices in processed dairy products. *Food Control*, *30*(1), 298-305.

Parker, K., Salas, M., & Nwosu, V. C. (2010). High fructose corn syrup: production, uses and public health concerns. *Biotechnology and Molecular Biology Reviews, 5*(5), 71-78.

Poe, G. L., Giraud, K. L., & Loomis, J. B. (2005). Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics, 87*(2), 353-365.

Saad, L. (2017). Americans' Dining-out Frequency Little Changed from 2008. Gallup. Retrieved from: https://news.gallup.com/poll/201710/americans-dining-frequency-little-changed-2008.aspx

SAS Institute Inc. 2014. SAS/QC® 13.2 User's Guide. Cary, NC: SAS Institute Inc.

Schenker, N., & Gentleman, J. F. (2001). On judging Significance of Difference by Examining the Overlap Between Confidence Intervals. *American Statistician, 53*, 182-186.

Speck, M. (1977). Heated Yogurt-Is it still yogurt?. *Journal of food protection, 40*(12), 863-865.

Szolnoki, G., & Hoffmann, D. (2013). Online, face-to-face and telephone surveys—Comparing different sampling methods in wine consumer research. *Wine Economics Policy 2*, 57-66.

Thompson, N. M., Widmar, N. O., Schutz, M. M., Cole, J. B., & Wolf, C. A. (2017). Economic considerations of breeding for polled dairy cows versus dehorning in the United States. *Journal of dairy science, 100*(6), 4941-4952.

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge, UK: Cambridge university press.

USDA NASS. (2019). Quickstats. Retrieved from: https://quickstats.nass.usda.gov/results/75CC0059-855F-32A0-AB68-7E420FE315C4

US Census Bureau. (2016). Annual Estimates of the Resident Population for Selected Age Groups by Sex for the United States, States, Counties and Puerto Rico Commonwealth and Municipios: April 1, 2010 to July 1, 2015.

USDL. (2019). Consumer expenditures—2018. Bureau of Labor statistics. Retrieved from: https://www.bls.gov/news.release/cesan.nr0.htm

Vanhonacker, F., Verbeke, W., Van Poucke, E., & Tuyttens, F. (2007). Segmentation based on consumers' perceived importance and attitude toward farm animal welfare. *International Journal of Sociology of Agriculture and Food*, *15*(3), 91-107.

Von Keyserlingk, M. A. G., Rushen, J., de Passillé, A. M., & Weary, D. M. (2009). Invited review: The welfare of dairy cattle—Key concepts and the role of science. *Journal of Dairy Science*, *92*(9), 4101-4111.

Wang, H., Livingston, K. A., Fox, C. S., Meigs, J. B., & Jacques, P. F. (2013). Yogurt consumption is associated with better diet quality and metabolic profile in American men and women. *Nutrition Research*, *33*(1), 18-26.

Watson, E. (2019). Where next for the US yogurt category? Chobani, Danone North America, Gerneral Mills, weigh in. Food. Retrieved from: file:///C:/Users/birc/Dropbox/WTP%20new%20method%202019/Yogurt%20literature/where%20next%20for%20US%20yogurt%20category.pdf (accessed September 2019)

Webb, D., Donovan. S. M., & Meydani, S. N. (2014). The role of yogurt in improving the quality of the American diet and meeting dietary guidelines. *Nutrition reviews*, *72*(3), 180-189.

Widmar, N., Morgan, C., Wolf, C. A., Yeager, E., Dominick, S., & Croney, C. (2017). US resident perceptions of dairy cattle management practices. *Agricultural Sciences*, *8*, 645-656.

Wisconsin Milk Marketing Board (WMMB). (2017). Dairy statistics. Retrieved from: http://www.americasdairyland.com/economicimpact/statistics/dairystatistics

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association 57*, 348–368.

2.8 Greek and traditional yogurt willingness-to-pay appendices

Appendix A. Verbiage seen by consumers participating in WTP choice experiment for Greek yogurt

The next portion of this survey presents you with hypothetical Greek yogurt purchasing scenarios that you could face in a retail store where you typically shop. The two products that will be presented in each scenario possess the same characteristics (e.g., similar color, brand, flavor, etc.) except for varying levels of the attributes presented below. Prices vary for each product.

For each scenario, please select the 5.3 oz cup of Greek yogurt that you would purchase, or neither, if you would not purchase either Greek cup of yogurt. For your information in interpreting alternative cups of yogurt:

Typical single-serve Greek yogurt cups are 5.3 oz in size. Thus, price refers to the cost per 5.3 oz cup of Greek yogurt: $0.72 per cup, $1.00 per cup, and $1.29 per cup.

<u>Animal welfare attributes</u>
Pasture access refers to the ability of dairy cattle to access grass pasture and not be confined solely to indoor production facilities

> -**Required pasture access** means the animal was raised on an operation providing animals with access to grass pasture
> -**Pasture access not required** indicates that no claims regarding access to grass pasture are being made

Dehorning/disbudding refers to the removal of horns/horn buds to insure dairy cows do not have horns

> -**Dehorning/disbudding not permitted** means the animal was raised on an operation that does not dehorn/disbud cattle.
> -**Dehorning/disbudding permitted** indicates that no claims regarding dehorning are being made

<u>Product attributes</u>
High-fructose corn syrup is a modified corn syrup that has an increased amount of fructose made from the enzymatic conversion of glucose and that is widely used as a sweetener

> -**Labeled free of high fructose corn syrup** means the product does not contain high fructose corn syrup
> -**No high fructose corn syrup labeling claim** means that no claims regarding the use of high fructose corn syrup are being made on this product

The amount of fat in yogurt depends on the type of milk the yogurt is made from

> - **Nonfat** indicates the yogurt is made from skim milk and contains less than 0.5% milkfat
> - **Lowfat** indicates the yogurt is made from skim milk and contains between 2% and 0.5% milkfat

The experience from previous surveys is that people often state a higher willingness-to-pay than what a person actually is willing-to-pay for the good. It is important that you make your selections like you would if you were actually facing these choices in your retail purchase decisions, noting that allocation of funds to these products means you will have less money available for other purchases.

Verbiage seen by consumers participating in WTP choice experiment for traditional yogurt

The next portion of this survey presents you with hypothetical traditional yogurt purchasing scenarios that you could face in a retail store where you typically shop. The two products that will be presented in each scenario possess the same characteristics (e.g., similar color, brand, flavor, etc.) except for varying levels of the attributes presented below. Prices vary for each product.

For each scenario, please select the 6 oz cup of traditional yogurt that you would purchase, or neither, if you would not purchase either traditional cup of yogurt. For your information in interpreting alternative cups of yogurt:

Typical single-serve traditional yogurt cups are 6 oz in size. Thus, price refers to the cost per 6 oz cup of traditional yogurt: $0.40 per cup, $0.79 per cup, and $1.14 per cup.

Animal welfare attributes
Pasture access refers to the ability of dairy cattle to access grass pasture and not be confined solely to indoor production facilities
-**Required pasture access** means the animal was raised on an operation providing animals with access to grass pasture
-**Pasture access not required** indicates that no claims regarding access to grass pasture are being made

Dehorning/disbudding refers to the removal of horns/horn buds to insure dairy cows do not have horns
-**Dehorning/disbudding not permitted** means the animal was raised on an operation that does not dehorn/disbud cattle.
-**Dehorning/disbudding permitted** indicates that no claims regarding dehorning are being made

Product attributes
High-fructose corn syrup is a modified corn syrup that has an increased amount of fructose made from the enzymatic conversion of glucose and that is widely used as a sweetener
-**Labeled free of high fructose corn syrup** means the product does not contain high fructose corn syrup
-**No high fructose corn syrup labeling claim** means that no claims regarding the use of high fructose corn syrup are being made on this product

The amount of fat in yogurt depends on the type of milk the yogurt is made from
- **Nonfat** indicates the yogurt is made from skim milk and contains less than 0.5% milkfat
- **Lowfat** indicates the yogurt is made from skim milk and contains between 2% and 0.5% milkfat

The experience from previous surveys is that people often state a higher willingness-to-pay than what a person actually is willing-to-pay for the good. It is important that you make your selections like you would if you were actually facing these choices in your retail purchase decisions, noting that allocation of funds to these products means you will have less money available for other purchases.

Appendix B. Pearson correlation between individual equations in the seemingly unrelated regression models

Correlations between individual equations in the seemingly unrelated regression for Greek yogurt

	WTP for required pasture access	WTP for dehorning disbudding not permitted	WTP for labeled free of high fructose corn syrup	WTP for nonfat	Opt-out
WTP for required pasture access	1				
WTP for dehorning disbudding not permitted	0.7474	1			
WTP for labeled free of high fructose corn syrup	0.1428	-0.0977	1		
WTP for nonfat	-0.3644	-0.2818	-0.5331	1	
Opt-out	0.7301	0.7217	0.5887	-0.5347	1

Correlations between individual questions in the seemingly unrelated regression for traditional yogurt

	WTP for required pasture access	WTP for dehorning disbudding not permitted	WTP for labeled free of high fructose corn syrup	WTP for nonfat	Opt-out
WTP for required pasture access	1				
WTP for dehorning disbudding not permitted	0.7795	1			
WTP for labeled free of high fructose corn syrup	-0.5827	-0.2926	1		
WTP for nonfat	-0.0451	0.2769	-0.0733	1	
Opt-out	0.7196	0.7072	0.1217	-0.0733	1

CHAPTER 3. PRESENTING SMALLER 'COMPONENT DESIGNS' IN PLACE OF A LARGE DISCRETE CHOICE EXPERIMENT

3.1 New willingness-to-pay data collection method introduction

The complexity of a choice task, which may result in a cumulative cognitive burden resulting in processing limits, is unobserved by the researcher and can be the result of variations in cognitive difficulty, uncertainty attributed to stimulus ambiguity, underlying preferences, among other influences (Swait and Adamowicz 2001a,b). As respondent decision effort increases with complexity, less consistent choice outcomes occur (Dellaert et al., 1999). Heiner (1983) argued that the gap between an individual's cognitive ability and the cognitive demands of a decision leads to a restriction in the range of decisions a respondent is able to consider. DeShazo and Fermo (2002) studied choice complexity in contingent valuation methods and found that an increasing cognitive burden compromised choice consistency, and therefore recommended that economists minimize the complexity of choice sets in the experimental survey design by selecting the optimal number of alternatives carefully. Dellaert et al. (1999) proposed that choice consistency was an important component in interpreting the results of willingness-to-pay (WTP) experiments, and evaluated the impact of varying the attribute levels in a WTP choice experiment. As price for the options increased, it became more difficult for respondents to decide between the utility benefits of the expenditure and the option to opt out (Dellaert et al., 1999). Shugan (1980) suggested that decision making was influenced by the respondents processing capability, the complexity of the choice and time pressure.

Due to the complexity of choice, respondents may employ processing strategies to cope with complexity when completing discrete choice tasks (Hensher, Rose, and Greene, 2005). Such strategies may include ignoring subsets of alternatives, aggregating attributes, imposing thresholds on attribute levels, and conditioning one attribute on the level of other attributes (Hensher, Rose, and Greene, 2005). Ignoring an attribute, often called attribute non-attendance (ANA), may impact the marginal effects of attribute changes (Scarpa et al., 2012). Hole (2011) outlined three approaches to ANA: stated ANA, where participants stated which attributes they did not consider after completing the choice experiment; a 'think aloud', approach where respondents were asked to verbalize ANA while completing the choice experiment, and inferred ANA. A coefficient of variation is calculated from random parameter estimates when evaluating inferred ANA, with a pre-determined threshold used as an indicator of the presence of ANA (Hess and Hensher, 2010). When evaluating data quality, ANA could serve as one method of evaluation, and as a potential proxy for the level of complexity faced by respondents.

Overall complexity for the respondent should decrease with the lessening of attributes and choice sets (DeShazo and Fermo, 2002; Hensher, 2006). This manuscript introduces a new approach to WTP data collection designed to decrease the complexity faced by respondents so as to increase data quality. WTP modeling has been used to determine preferences for animal welfare attributes in food products including yogurt (Napolitano et al., 2008), beef and chicken products (Carlsson et al., 2007), as well as milk and pork chops (Olynk et al., 2010). The number of choice scenarios presented to respondents in a WTP discrete choice experiment (CE) can be decreased by selecting a model with a lower efficiency, by decreasing the number of attributes studied, or by decreasing the number of levels within attributes (Hensher, 2006). Our new discrete CE data collection method uses fewer choice scenarios than the traditional method, but with the advantage that our improved data quality does *not* come via decreased efficiency, which is achieved through the use of component models. Specifically, respondents participate in a discrete CE design that does not include attributes in the category they self-declared as unimportant (hereby referred to as component models) through the use of a question that prompts respondents to indicate which category of attributes they do not find important (herby referred to as self-filtering question). Using the aggregated results of the component models, researchers are able to estimate WTPs for all attributes in the study. Prior to the CE, before the potential impacts of choice fatigue sets in, respondents were asked one simple question with two choices that allows for a decrease in attributes. By not including the attributes respondents declared least important prior to engaging in the choice tasks, we anticipated that incidences of ANA would decrease.

The main objective of this manuscript is to compare the results of a new WTP data collection method and the traditional method of WTP through experiments designed to elicit yogurt consumer's WTP for attributes in yogurt. The new WTP data collection method was designed with the objective of decreasing complexity by having respondents participate in fewer choice scenarios, without having to decrease design efficiency, and by limiting the number of attributes included, while in aggregate studying a larger set of attributes. ANA has been associated with complexity, simplification heuristics and fatigue (Hensher, Rose, and Greene, 2005; Scarpa et al., 2012; Hole, 2011; Hess and Hensher, 2010), so it was hypothesized that the new WTP data collection method would elicit lower incidences of ANA, and evaluating incidences of ANA was used as one measure of comparison between the two methods.

3.2 New willingness-to-pay data collection method methods

An online survey administered in Qualtrics was distributed June 11-21, 2019 to inform discrete CE data collection to allow comparison of the tradition WTP and our new WTP data collection method. Kantar, a company which hosts an opt-in online panel was used to obtain survey respondents. The data used in this analysis was part of a larger data collection effort that resulted in a total of 1,440 respondents. Respondents were asked to indicate if they purchased either Greek yogurt, traditional yogurt, or neither, and were assigned to a WTP choice experiment on that particular product based on their response. From those 1,400 respondents the random subsets that were assigned to participate in traditional choice experiments for traditional yogurt (n=347) and the choice experiment for the new WTP method of data collection for traditional yogurt (n=347) were employed in this analysis, resulting in a total sample size of 694. The remaining respondents undertook experiments for Greek yogurt which were not employed in this comparison intended to study differences between data collection methods on traditional yogurt, which is the more commonly purchased and consumed product among those studied. Respondents who did not purchase either Greek or traditional yogurt were randomly assigned to either participate in the traditional method Greek or traditional WTP designs. These respondents who likely had varied preferences from consumers of yogurt did not participate in the new WTP data collection method and therefore were not included in this analysis. Survey flow for the choice experiments included in this manuscript is available in Figure 3. A test of proportions was used to determine if there were any statistical differences between the traditional method and new WTP data collection method samples.

The attributes studied, and eventually valued through WTP estimates, in the traditional method and the new data collection method were the same, and included: required pasture access or pasture access not required; dehorning/disbudding not permitted or dehorning/disbudding permitted; labeled free of high fructose corn syrup or no high fructose corn syrup labeling claim; and nonfat or lowfat. Prices for a 6 oz cup of yogurt were \$0.40, \$0.79 and \$1.14. Concern for dairy production practices including dehorning or disbudding and pasture access were previously studied in general (Widmar et al., 2017), and specifically in fluid dairy milk (Bir et al., 2019). High fructose corn syrup is widely used as a sweetener and is made from the enzymatic conversion of glucose (Merriam Webster, 2019). The use of corn syrup in products continues to be a point of contention and of confusion for consumers, with the viral Bud Light super bowl add underlining this point (Vinjamuri, 2019). Although the use of corn syrup is a moot point, none of the corn syrup remains in the finished beer, its non-use was a point of pride for the “Bud King” who traveled the realm in the commercial searching for the corn syrup’s true brewery home (Vinjamuri, 2019). Despite consumers concerns, no findings indicating negative health effects related to the consumption of fructose corn syrup has been found in humans (Parker and Nwosu, 2010). Studying consumer WTP for yogurt while employing a new discrete choice data collection methodology to inform WTP modeling allowed for the testing of the new simpler choice tasks while adding to the knowledge base regarding an economically important dairy product.

3.2.1 Traditional willingness-to-pay methodology

The SAS OPTEX program was used to design a main effects experiment and determine the specific combination of attribute levels seen by respondents in the CE (Lusk and Norwood 2005; SAS, 2014). The specific design was chosen by maximizing D-efficiency which was 85.49 for the traditional method. Respondents who participated in the traditional WTP method completed 12 choice scenarios (questions), for which an example is presented in Figure 4. Respondents were shown information explaining each of the four attributes in the CE prior to participating in the WTP CE (see Appendix A for verbiage). The cheap talk script presented in Lusk (2003) was employed in the survey and presented to respondents immediately after the information on attributes and before starting the experiment.

Random utility theory is the basis for choice models, therefore the objective of the model is to maximize utility (U). Utility is not known by the researcher so it is defined as a deterministic component V_{nit}, *if* $U_{nit} > U_{njt} \ \forall j \neq i$. The probability that respondent *n* chooses alternative *i*, is represented by (Train, 2009):

$$P_{nit} = Prob\left(V_{nit} + \varepsilon_{nit} > V_{njt} + \varepsilon_{njt}; \forall_j \in C, \forall_j \neq i\right). \tag{7}$$

Equation (7) can be condensed through algebraic manipulation due to the underlying distribution of the error term:

$$P_{nit} = \frac{\exp(V_{nit})}{\sum_j \exp(V_{njt})}. \tag{8}$$

The deterministic component of *V* of a selection for yogurt is defined as:

$$V_{it} = \beta_1 Price_{it} + \beta_2 ReqPasture_{it} + \beta_3 NoDehorning_{it} + \beta_4 FreeOfFructose_{it} + \beta_5 Nonfat_{it} + \beta_6 Optout_{it} \quad (9)$$

where *Price* is the price a respondent is willing-to-pay for yogurt; *Optout* is a constant which represents the respondent's disutility from having to walk away from purchasing yogurt; *ReqPasture* is the effects coded term for required pasture access; *NoDehorning* is the effects coded term for dehorning/disbudding not permitted; *FreeOfFructose* is the effects coded term for labeled free of high fructose corn syrup and *Nonfat* is the effects coded term for nonfat as opposed to *lowfat*. For the RPL model estimated using NLOGIT 6 (Greene, 2019) the disutility in terms of dollars of walking away from the purchase of yogurt (OptOut) is calculated as:

$$WTP = -\frac{\beta_6}{\beta_1}. \quad (10)$$

For all other attributes WTP is calculated as the negative of the ratio of the coefficient for the particular attribute and the coefficient of price, for example the WTP for dehorning/disbudding not permitted is calculated:

$$WTP = -2\frac{\beta_3}{\beta_1}. \quad (11)$$

In Equation 11, the -2 accounts for the effects coding of the two levels of the 4 attributes. All attributes in this experiment had only two levels and were coded with -1 and 1 unlike typical 0, 1 dummy variable coding (Adamowicz et al., 1994). Unlike in the case of traditional dummy variable estimation, the effects coding prevents the left out dummy variable from being incorporated into the intercept (Adamowicz et al., 1994).

Deviating from the typical use of the mean coefficients to estimate WTP, individual WTPs for each attribute and *OptOut* were determined using equations 10 and 11. The individual level WTP for each attribute and *OptOut* was then averaged across all respondents to obtain a mean WTP. This deviation allowed for the combination of the new WTP data collection component methods, and following the same methodology in the traditional method allowed for a more accurate comparison of the new and traditional data collection methods. The standard deviation for each attribute was determined and used to calculate the 95% confidence interval (*z* score of 1.96) using the standard formula:

$$confidence\ interval_j = mean \pm (1.96 * \left(\frac{Standard\ Deviation}{SQRT(Sample\ size)}\right). \quad (12)$$

3.2.2 New willingness-to-pay data collection methodology

The main objective of the new WTP data collection method was to decrease complexity by having respondents make fewer choices and/or by limiting the number of attributes included. The new WTP data collection method achieves this objective by employing two component models, which in combination result in estimates for all attributes of interest. Prior to participating in the new WTP data collection method, respondents were presented with information regarding the attributes and their categorization (Appendix A). Respondents were then asked, "Which of the following categories do you find LEAST important when making a yogurt purchasing decision?", and were presented the options animal welfare and product attributes (Figure 3). Respondents who selected animal welfare as least important then participated in a CE that did not include the animal related attributes-required pasture access and dehorning/disbudding permitted. Respondents who selected product attributes as least important participated in a CE that did not include product related attributes including high-fructose corn syrup and fat content. The test of proportions was used to determine if there were any statistical differences between the two self-sorted samples of respondents who would participate in the two different CEs that would inform the component models. Respondents were only shown attributes within the CE that were important to them, and this approach was therefore hypothesized to decrease ANA. The SAS OPTEX program (SAS, 2014) was used to design a main effects experiment for both component models and determine the specific combination of attribute levels seen by respondents. The D-efficiency for both component models was 91.138, and respondents participated in 7 choice scenarios (questions) for either component model (examples available in Figure 4-6). Individual WTPs were estimated for the yogurt product attributes within each of the two component models individually. Once individual respondent level WTPs were determined for the two component models, a WTP of zero was imposed on the attributes not included in the model the respondent participated in. For example, if the respondent indicated that animal welfare attributes were least important, the WTP for the attributes high fructose corn syrup and fat content were estimated using their responses to the choice scenarios in the component model, and the respondent was assigned a zero WTP for pasture access required and dehorning/disbudding no permitted. Once the individual level WTPs were determined for the component models, the averages were calculated for each attribute to determine the mean WTP for the full new WTP method of data collection. Using equation 12, confidence intervals were determined for the four attributes studied and *OptOut* for the new WTP method of data collection. Differences between the WTP for attributes within the traditional WTP method, within the new WTP data collection method, and between the two methods were determined by observing overlapping confidence intervals (Schenker and Gentleman, 2001).

3.2.3 Comparing respondent decision making between willingness-to-pay experimental designs

3.2.4 Willingness-to-pay click and timing comparison

As one measure of comparison of effort that may contribute to fatigue, the number of clicks respondents made while participating in either the new WTP data collection method or the traditional data collection method were recorded. The minimum number of clicks necessary to complete the new WTP data collection and traditional WTP method was 7 and 12, respectively. This minimum number of clicks was subtracted from the number of clicks each respondent made to determine the number of superfluous clicks. The difference between the number of clicks and the minimum number were compared between the two methods using a t-test. The amount of time measured in seconds respondents spent participating in the choice experiments, not including the time spent reading instructional material, was also recorded. Naturally, it could be assumed that it would take most respondents longer to respond to 12 choice scenarios when compared to 7 choice scenarios, so the amount of time it took respondents to complete either the new WTP data collection method or the traditional WTP method was divided by either 7 or 12 respectively to determine the amount of time spent per choice scenario.

3.2.5 Willingness-to-pay attribute non-attendance

Comparing incidences of ANA is one method to compare the traditional and new WTP methods. The coefficient of variation was calculated as (Hess and Hensher, 2010):

$$\textit{Coefficient of variation} = |(\textit{individual standard deviation})/(\textit{individual coefficient})|. \quad (13)$$

A threshold of 2 was used to determine incidences of ANA for both the traditional and new WTP methods following Hess and Hensher (2010). If the coefficient of variation exceeded the threshold of 2, the attribute for that respondent was re-coded -888 to signal values that were determined to be deliberately omitted from the data set by the individual respondent (Greene, 2013). The percentage of respondents who exhibited attribute non-attendance was calculated for each model and compared between the traditional method and the new WTP data collection method using the test of proportions. All of the models, the traditional WTP model and the two component models that make up the new WTP data collection method, were then re-estimated with coded incidences of ANA. The mean WTP, standard deviations, and confidence intervals were calculated using equations 11-12 following the same procedures as the non-ANA corrected models. To evaluate the potential relationship between ANA and the time respondents take to answer choice scenarios, the Pearson correlation (Pearson and Filon, 1898) between exhibiting ANA for any attribute and the amount of time it took respondents to complete the WTP CE was estimated.

3.3 New willingness-to-pay data collection method results

There were a few statistical differences between the demographics of the respondents in the traditional versus new data collection methods (Table 7). At the <0.001 level there was a statistically lower percentage of respondents ages 45-54 and with an income of $0-$24,999 in the new method (14%, 18%) when compared to the traditional method (29%, 25%). Additionally, a higher percentage of respondents with an income of $75,000-$99,999 completed the new method (18%) when compared to the traditional method (12%). When comparing the self-selected component models of the new WTP data collection method, in general, lower percentages of respondents aged 44 and younger selected animal welfare least important when compared to product labeling. There were not statistically significant differences in gender between those who selected animal welfare and those who selected product labeling.

For the traditional method of employing CE data to estimate respondent WTP, respondents were willing-to-pay a positive amount for required pasture access and dehorning/disbudding not permitted (Table 8). The mean coefficient from the model for the attribute labeled free of high fructose corn syrup was not statistically significant. Respondents had a negative WTP for nonfat, indicating a preference for lowfat yogurt. Additionally, respondents had a negative WTP for opt-out, indicating that they experienced disutility from walking away from a yogurt buying opportunity. WTP for the attribute labeled free of high fructose corn syrup was positive for the animal welfare attributes not included component model and the mean coefficient from the model for nonfat was not statistically significant. Respondents had positive WTP for both attributes, required pasture access and dehorning/debudding not permitted, in the physical appearance attributes not included component model. In both component models, respondents had a negative WTP for *opt-out*.

By comparing confidence intervals, the WTP for all attributes were statistically significantly different between the traditional method and the new WTP data collection method with the exception of labeled free of high fructose corn syrup and nonfat (Table 9). However, the mean coefficient from the model for labeled free of high fructose corn syrup was not statistically significant in the traditional method model, and the mean coefficient from the animal welfare attributes not included component model for the nonfat was not statistically significant. For both required pasture access and dehorning/disbudding not permitted, respondent's WTP was higher in the new WTP data collection method.

On average the traditional WTP method took approximately an additional 155 seconds to complete, which is slightly over twice as long as the new WTP data collection method (Table 10). The maximum amount of time a respondent took to complete the traditional method was 5344.465 seconds, which was approximately four times longer than the maximum time it took for a respondent to complete the new WTP method. On a per-CE completed basis no statistical difference in time to complete a choice task was found between the two methods. However, statistical differences were found in the mean adjusted number of clicks it took respondents to complete the choice experiments. On average, it took respondents 22 clicks over the minimum number for the traditional method, which was statistically different than the 12 additional clicks over the minimum number for the new WTP data collection method.

For all but one attribute, the percentage of respondents who exhibited ANA was statistically smaller for the new WTP data collection method when compared to the traditional. For pasture access, the percentage of respondents who exhibited incidences of ANA was not statistically different between the traditional method (29%) and the new WTP data collection method (26%). For the traditional method, 33% of respondents exhibited ANA for dehorning/disbudding not permitted, 33% for labeled free of high fructose corn syrup and 61% for nonfat. For the new WTP data collection method, 17% of respondent exhibited ANA for dehorning/disbudding not permitted, 13% for labeled free of high fructose corn syrup and 38% for nonfat. Exhibiting ANA for any attribute was negatively correlated for the traditional method (-0.1264, p=0.0185) and the new WTP data collection method (-0.1081, p=0.0441) with the time it took respondents to complete the WTP choice experiment.

For the ANA corrected models, all mean attribute coefficients from the models were statistically significant for both the traditional method and the WTP new method component models. For both methods, respondents had a positive WTP for required pasture access, dehorning/disbudding not permitted, and labeled free of high fructose corn syrup. Respondents had a negative WTP for nonfat in both methods, and experienced disutility from walking away from a yogurt buying opportunity. Comparing overlapping confidence intervals, only the WTP for the attribute nonfat was statistically different between the ANA-corrected and non-ANA-corrected traditional models (Table 9). For the new WTP data collection method, only the WTP for the attribute labeled free of high fructose corn syrup was statistically different between the ANA-corrected and non-ANA-corrected models. Between the ANA corrected traditional and new WTP data collection methods, the WTPs for all attributes were statistically significantly different with the exception of opt-out. In the ANA-corrected new WTP data collection method, the WTP for required pasture access, dehorning/disbudding not permitted, and nonfat were statistically higher than the traditional method. For the attribute labeled free of high fructose corn syrup, respondents in the traditional method had a higher WTP.

3.4 New willingness-to-pay data collection method discussion

Surprisingly few differences were found between the self-selected component models and the demographics of the respondents. A lower percentage of respondents who were under 44 years old selected animal welfare as least important, indicating greater concern for animal welfare. McKendree et al. (2014) similarly found that younger people had greater levels of concern for farm animal welfare. Although other studies have found that women have a greater concern for animal welfare (McKendree et al., 20014; Morgan et al., 2016; Vanhonacker et al., 2007) there were not any statistical differences found between the genders and those who selected animal welfare as least important and product labeling as least important. The attributes that were included under the two categories were provided to respondents. It is possible that respondents didn't find the particular attributes within the animal welfare category important, but do care about animal welfare in general.

With the exception of attributes that were insignificant in either the traditional or new WTP data collection methods, WTPs were statistically different between the two methods for all attributes in the uncorrected models. Other studies of discrete choice models have found differences related to experimental design. Studying best-worst scaling, where respondents are asked to select the most important or least important attribute from a provided set, Byrd et al. (2018) found that results differed when respondents participated in either a show-2 or a show-3 attribute per choice scenario design. Hensher (2006) varied the number of choice sets presented, the number of alternatives in each choice set, the number of attributes per alternative and the number of levels of each attribute while evaluating WTP designs for a car commuting trip. The number of attributes per alternative and the number of alternatives in a choice set resulted in differences in mean WTP, indicating that design choices indeed had an impact on results (Hensher, 2006). Malhortra (1982) evaluated the impact of the number of attributes and number of choice sets on the results of a discrete CE of house attribute preferences. Overload, which results due to a finite limit in the processing ability of respondents, was found to occur with 10 or more alternatives in a choice set, or with 15 or more attributes included in the experiment. Although the number of attributes and choice scenarios compared in this experiment was under the threshold found by Malhotra (1982), it is possible that the decreased number of attributes and choice scenarios required in the component models of the new WTP data collection method resulted in less respondent overload.

Although it is not surprising that a CE that requires fewer choice scenarios would result in fewer respondent clicks and less time overall, when considering the impact of fatigue on survey data quality, being able to garnish the same information with less respondent clicks and time is an advantage. Employing the component models and smaller associated CEs required both less time and clicks, unadjusted, and even required fewer clicks over the minimum number when compared to the traditional method. Galesic et al. (2009) found that response rate decreased when respondents were presented the length of survey in minutes, and for longer surveys response rate for individual questions decreased. Striving for shorter surveys that result in less fatigue can help improve data quality (Galesic et al., 2009). Byrd et al. (2018) compared the amount of time it took respondents per question to complete a choice scenario with 3 attributes (approximately 14.51 seconds) and a choice scenario with 2 attributes (3.41 seconds), which were statistically different. Despite the complexity of WTP choice scenarios compared to BWS, it did not take respondents much longer to complete a WTP choice scenario in the experiments evaluated in this manuscript when compared to the show-3 BWS choice scenario in Byrd et al. (2018). It is possible that simplification heuristics (Hensher, Rose, and Greene, 2005; Swait and Adamowics, 2001) were employed by at least some respondents when presented with the added question complexity of the traditional method, decreasing the time needed to complete the task. Recall respondents in the traditional method were presented with 5 attributes at once when compared to the three attributes presented in the new WTP data collection component models, so respondents taking additional time in the traditional method per choice scenario would not have been surprising.

As hypothesized, the new WTP data collection method resulted in significantly lower incidences of ANA for all attributes with the exception of pasture access. Albeit not conclusive, taking less time to complete the CE was correlated with exhibiting ANA in both the traditional method and the new WTP data collection method employed in this analysis. Those who took less time had higher incidences of ANA which may begin to explain why added complexity did not result in additional time needed per choice scenario. Although ANA has been extensively studied in the literature, the exact nature of what ANA is capturing beyond the amount of variability is not definitive. Keller et al. (1987) suggested that the relationship between complexity and decision effectiveness may have an inverted U-shape. Respondents initially exert more effort as complexity increases, becoming more effective until the point where their effectiveness diminishes (Keller et al., 1987). This idea coupled with varying individual ability (Shugan, 1980), and complex processing strategies (Hensher, Rose, and Greene, 2005; Swait and Adamowics, 2001) begins to establish a possible relationship between complexity and ANA. The main issue with non-optimal strategies, strategies (heuristics) that diverge from utility maximization, is that discrete CEs are rooted in random utility theory and assumptions are made that the respondent is a rational utility maximizer with limitless capacity (Dellaert et al., 1999; Swait and Adamowicz, 2001a, 2001b). Unfortunately, it is unlikely that all respondents behave in this manner, so any attempt that can be made to decrease complexity and account for divergent behavior should, in general, be considered. Although there were fewer respondents who exhibited ANA in the new WTP data collection method, correcting for ANA still resulted in a statistically different WTP for one attribute.

3.5 New willingness-to-pay data collection method conclusion

The impact of decision making, fatigue, and optimal performance has infiltrated daily decision making in both personal and professional settings. This shift towards considering complexity and decision making makes revisiting these effects in discrete choice methods in a research setting timely. Discrete choice experiments are often used to elicit respondent preferences for attributes including agricultural products and the resulting choices are used to inform further analyses, such as estimating WTP for specific product attributes. This manuscript introduces a new method of discrete CE data collection with the objectives of decreasing complexity by decreasing the number of choice scenarios respondents participate in without sacrificing design efficiency or the number of attributes studied in aggregate.

Our new WTP data collection method was compared to a traditional WTP while evaluating the same attributes of traditional yogurt. In the new WTP data collection method, the use of a filter question resulted in respondents participating in a tailored WTP CE that did not include the group of attributes respondents found unimportant. Respondents were willing-to-pay a statistically significant positive amount for required pasture access, dehorning/disbudding not permitted, and a negative amount for nonfat and opt-out in the traditional method. For the new WTP method respondents were willing-to-pay a statistically significantly positive amount for required pasture access, dehorning/disbudding not permitted, and labeled free of high fructose corn syrup, and a negative WTP for optout. After correcting for ANA in both models, all mean coefficients of attributes were statistically significantly different than zero. Incidences of ANA, a potential simplifying heuristic that results from complexity, occurred less frequently for all attributes in the new WTP data collection method with the exception of pasture access. Exhibiting ANA for any attribute was negatively correlated with the time respondents took to complete the choice experiment.

The new WTP data collection method decreases the number of attributes respondents must consider, the number of choice scenarios required, and decreases the amount of time required for respondents to complete the choice experiment. These benefits may result in better quality data, and decreased incidences of fatigue and overload.

3.6 New willingness-to-pay data collection method references

Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics Management, 26*(3), 271-292.

Bir, C., Widmar, N. O., Wolf, C., & Delgado, M. S. (2019). Traditional attributes moo-ve over for some consumer segments: Relative ranking of fluid milk attributes. *Appetite, 134*, 162-171.

Byrd, E. S., Widmar, N. J. O., & Gramig, B. M. (2018). Presentation matters: Number of attributes presented impacts estimated preferences. *Agribusiness 34*(2), 377-389.

Carlsson, F., Frykblom, P., & Lagerkvist, C. J. (2007). Consumer willingness-to-pay for farm animal welfare: mobile abattoirs versus transportation to slaughter. *European Review of Agricultural Economics*, *34*, 321-344.

Dellaert, B. G., Brazell, J. D., & Louviere, J. J. (1999). The effect of attribute variation on consumer choice consistency. *Marketing Letters, 10*(2), 139-147.

DeShazo, J. R., & Fermo, G. (2002). Designing choice sets for stated preference methods: the effects of complexity on choice consistency. *Journal of Environmental Economics and management*, *44*(1), 123-143.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349–360.

Greene, W.H. (2019). Nlogit. Version 6.0, 10.

Greene, W. (2013). Econometric Analysis (7th edition). Upper Saddle River NJ, USA: Pearson Education Inc.

Heiner, R. A. (1983). The Origin of Predictable Behaviour. *American Economic Review, 73*, 560–595.

Hensher, D. A. (2006). Revealing differences in willingness-to-pay due to the dimensionality of stated choice designs: an initial assessment. *Environmental and Resource Economics, 34*(1), 7-44.

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

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

Hole, A. R. (2011). Attribute non-attendance in patients' choice of general practitioner appointment. Presented at the 2nd international choice Modelling Conference, Leeds. Retrieved from: http://tinyurl.com/yayepff3

Keller, K., & Staelin, R. (1987). Effects of Quality and Quantity of Information on Decision and Effectiveness. *Journal of Consumer Research, 14*, 200–213.

Lusk, J. L., & Norwood, F. B. (2005). Effect of experimental design on choice-based conjoint valuation estimates. *American Journal of Agricultural Economics, 87*(3), 771–785.

Lusk, J. L. (2003). Effects of cheap talk on consumer willingness-to-pay for golden rice. *American journal of agricultural economics*, *85*(4), 840-856

Malhotra, N. K. (1982). Information load and consumer decision making. *Journal of consumer research, 8*(4), 419-430.

McKendree, M. G., Croney, C. C., & Widmar, N. O. (2014). Effects of demographic factors and information sources on United States consumer perceptions of animal welfare. *Journal of animal science, 92*(7), 3161-3173.

Merriam Webster. (2019). High-fructose corn syrup. Merriam Webster online dictionary. Retrieved from: https://www.merriam-webster.com/medical/high%20fructose%20corn%20syrup

Morgan, C. J., Croney, C. C., & Widmar, N. J. (2016). Exploring relationships between ethical consumption, lifestyle choices, and social responsibility. *Advances in Applied Sociology*, *6*, 199-216.

Napolitano, F., Pacelli, C., Girolami, A., & Braghieri, A. (2008). Effect of information about animal welfare on consumer willingness-to-pay for yogurt. *Journal of dairy science, 91*(3), 910-917.

Olynk, N. J., Tonsor, G.T., & Wolf, C. A. (2010). Consumer willingness-to-pay for livestock credence attributes claim verification. *Journal of Agricultural and Resource Economics*, *35*(2), 261-280.

Parker, K., Salas, M., & Nwosu, V. C. (2010). High fructose corn syrup: production, uses and public health concerns. *Biotechnology and Molecular Biology Reviews*, *5*(5), 71-78.

Pearson, K., & Filon, L. N. G. (1898). Mathematical contributions to the theory of evolution. IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Philosophical Transactions of the Royal Society of London, Series A*, *191*, 229–311.

SAS Institute Inc. (2014). SAS/QC® 13.2 User’s Guide. Cary, NC: SAS Institute Inc.

Scarpa, R., Zanoli, R., Bruschi, V., & Naspetti, S. (2012). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics*, *95*(1), 165-180.

Schenker, N., & Gentleman, J. F. (2001). On judging Significance of Difference by Examining the Overlap Between Confidence Intervals. *American Statistician*, *53*, 182-186.

Shugan, S. M. (1980). The Cost of Thinking. *Journal of Consumer Research*, *7*, 99-111.

Swait, J., & Adamowicz, W. (2001a). 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., & Adamowicz, W. (2001b). Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice. *Organizational Behavior and Human Decision Processes*, *49*, 1–27.

Tonsor, G. T., Olynk, N., & Wolf, C. (2009). Consumer preferences for animal welfare attributes: The case of gestation crates. *Journal of Agricultural and Applied Economics, 41*(3), 713-730.

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge U.K.: Cambridge university press.

Vanhonacker, F., Verbeke, W., Van Poucke, E., & Tuyttens, F. (2007). Segmentation based on consumers’ perceived importance and attitude toward farm animal welfare. *International Journal of Sociology of Agriculture and Food*, *15*(3), 91-107.

Vinjamuri, D. (2019). Bud Light’s Super Bowl Ad: Corn Syrup, Ethics and Mistaken Identity. Forbes. Retrieved from: https://www.forbes.com/sites/davidvinjamuri/2019/02/06/bud-lights-super-bowl-ad-corn-syrup-ethics-and-mistaken-identity/#6d29b8fd3a39.

Widmar, N., Morgan, C., Wolf, C. A., Yeager, E., Dominick, S., & Croney, C. (2017). US resident perceptions of dairy cattle management practices. *Agricultural Sciences*, *8*, 645-656.

3.7 New willingness-to-pay data collection method tables and figures

Table 7. Demographics of yogurt consumers and comparison of self-selected new method WTP data collection component model

Demographic Variable	New Method n=347	Traditional Method n=347	New Method component model animal welfare selected least important[1] n=161	New Method component model Product labeling selected least important[2] n=186
Gender				
Male	45	43	44	47
Female	55	57	56	53
Age				
18-24	8	10	9	7
25-34	26	24	21 $^{\Psi\Psi}$	32 $^{\Psi\Psi}$
35-44	19	19	16 $^{\Psi\Psi}$	24 $^{\Psi\Psi}$
45-54	14 $^{\Omega\Omega}$	20 $^{\Omega\Omega}$	16 $^{\Psi}$	12 $^{\Psi}$
55-65	15	12	16	14
65 +	17	15	23 $^{\Psi\Psi}$	11 $^{\Psi\Psi}$
Income				
$0-$24,999	18 $^{\Omega\Omega}$	25 $^{\Omega\Omega}$	18	19
$25,000-$49,999	26 $^{\Omega}$	20 $^{\Omega}$	26	25
$50,000-$74,999	17	20	17	18
$75,000-$99,999	18$^{\Omega\Omega}$	12 $^{\Omega\Omega}$	20 $^{\Psi}$	15 $^{\Psi}$
$100,000 and higher	21	23	19	22
Education				
Did not graduate from high school	5 $^{\Omega}$	3 $^{\Omega}$	5 $^{\Psi}$	4 $^{\Psi}$
Graduated from high school, Did not attend college	34	35	33	34
Attended College, No Degree earned	18	17	21	16
Attended College, Associates or Bachelor's Degree earned	28	31	26	30
Attended College, Graduate or Professional Degree earned	15	14	14	16
Region				
Northeast	16	17	17	14
South	42	40	42	41
Midwest	20	20	21	18
West	23	24	20 $^{\Psi}$	27 $^{\Psi}$

$^{\Omega}$ Percentage of respondents is statistically different between the new method and traditional method samples at the 0.05 level $^{\Omega\Omega}$<0.001 level

$^{\Psi}$ Percentage of respondents is statistically different between animal welfare and product labeling selected least important 0.05 level $^{\Psi\Psi}$<0.001 level

[1]This component model did not include animal welfare attributes

[2]This component model did not include physical appearance attributes

Table 8. Traditional and new method willingness-to-pay mean model coefficients and standard errors, and average of individual respondent willingness-to-pay

	Traditional Method n=347				New Method n=347: Animal welfare selected least important n=161		New Method n=347: Product attributes selected least important n=186			
Attributes	**Coefficient (SE)**	**Standard deviation (SE)**	**Mean WTP[1]**	**Number of incidences of ANA**	**Coefficient (SE)**	**Standard deviation (SE)**	**Coefficient (SE)**	**Standard deviation (SE)**	**Mean WTP[1]**	**Number of incidences of ANA**
Required pasture access	0.275*** (0.042)	0.353*** (0.048)	$0.44	29%	————	————	0.731*** (0.107)	0.635*** (0.143)	$1.07	26%
Dehorning/ disbudding not permitted	0.349*** (0.042)	0.357*** (0.046)	$0.57	33%$^{\Omega}$	————	————	0.705*** (0.112)	0.488*** (0.106)	$1.06	17%$^{\Omega}$
Labeled free of high fructose corn syrup	0.042 (0.033)	0.014 (0.078)	$0.07	33%$^{\Omega}$	0.117** (0.051)	0.120 (0.087)	————	————	$0.09	13%$^{\Omega}$
Nonfat	-0.087** (0.036)	0.080 (0.124)	-$0.13	61%$^{\Omega}$	-0.073 (0.065)	0.167 (0.192)	————	————	-$0.06	38%$^{\Omega}$
Optout	-1.793*** (0.195)	3.025*** (0.183)	-$1.57		-2.971*** (0.332)	3.151*** (0.351)	-1.930*** (0.443)	3.811*** (0.379)	-$2.20	
Price	-1.197*** (0.092)				-1.402*** (0.142)		-0.701*** (0.167)			

* statistically significant at the 0.10 level, **0.05 level, and *** at the <0.001 level
[1]The mean WTP was calculated as the average of the individual WTP estimates
$^{\Omega}$The percentage of IANA incidences is statistically different between the traditional and new method

Table 9. Confidence intervals of preference shares for traditional method and new willingness-to-pay data collection method

	Traditional Method n=347				New WTP Data Collection Method n=347				Models have statistically different WTP
	Lower bound	Mean	Upper bound		Lower bound	Mean	Upper bound		
Required pasture access	0.376	0.439	0.502		0.782	1.073	1.363		Yes
Dehorning/ disbudding not permitted	0.497	0.567	0.638		0.723	1.058	1.394		Yes
Labeled free of high fructose corn syrup	0.048	0.073	0.097		0.070	0.087	0.103		No
Nonfat	-0.171	-0.135	-0.099		-0.121	-0.065	-0.009		No
Optout	-1.810	-1.569	-1.327		-2.588	-2.200	-1.812		Yes
	Traditional Method IANA corrected			Models w/ and w/o ANA correction are statistically different	New WTP Data Collection Method IANA corrected			Models w/ and w/o ANA correction are statistically different	Models have statistically different WTP
	Lower bound	Mean	Upper bound		Lower bound	Mean	Upper bound		
Required pasture access	0.497	0.582	0.667	No	0.942	1.309	1.677	No	Yes
Dehorning/ disbudding not permitted	0.623	0.717	0.811	No	0.872	1.275	1.679	No	Yes
Labeled free of high fructose corn syrup	0.092	0.132	0.173	No	0.036	0.045	0.055	Yes	Yes
Nonfat	-0.337	-0.260	-0.183	Yes	-0.176	-0.088	0.000	No	Yes
Optout	-1.825	-1.570	-1.315	No	-2.587	-2.145	-1.702	No	No

[1] Traditional method and New WTP Data Collection method have statistically different preference shares based on overlapping confidence intervals for either uncorrected, ANA corrected, or transitivity corrected model

Table 10. Traditional and new willingness-to-pay method click and timing analysis

	Mean	Std. Dev.	Min	Max
Unadjusted amount of time per choice experiment				
Product attributes selected least important component model	66.111	101.881	8.742	1209.586
Animal welfare selected least important component model	81.457	109.668	11.234	1175.017
New WTP data collection method	74.337	106.251	8.742	1209.586
Traditional WTP method	154.632	319.966	12.763	5344.465
Time adjusted for number of choice scenarios				
Product attributes selected least important component model	9.444[1]	14.554	1.249	172.798
Animal welfare selected least important component model	11.637[1]	15.667	1.605	167.860
New WTP data collection method	10.620	15.179	1.249	172.798
Traditional WTP method	12.886	26.664	1.063	445.372
Unadjusted number of clicks				
Product attributes selected least important component model	21.161	15.842	7	81
Animal welfare selected least important component model	18.231	18.696	7	145
New WTP data collection method	19.591	17.467	7	145
Traditional WTP method	34.147	29.502	12	276
Difference between number of clicks and minimum number				
Product attributes selected least important component model	14.161[1]	15.842	0	74
Animal welfare selected least important component model	11.231[1]	18.696	0	138
New WTP data collection method	12.591[2]	17.467	0	138
Traditional WTP method	22.147[2]	29.502	0	264

[1]Mean adjusted time per choice scenario and number of clicks were not statistically different between the two new WTP data collection component models.

[2] Mean adjusted number of clicks were statistically different between the new WTP data collection method and the traditional WTP method.

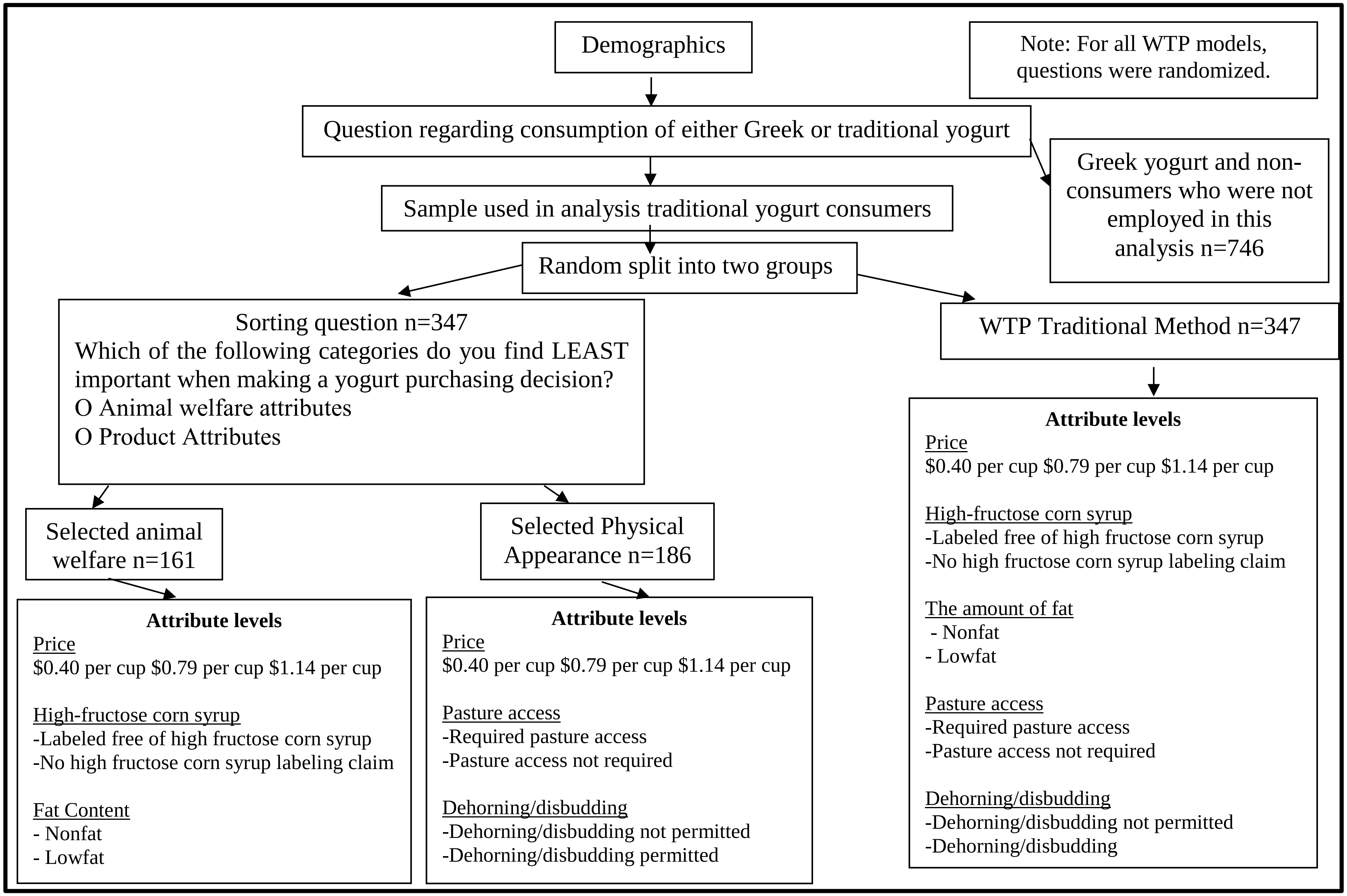

Figure 3. Survey flow including sorting questions for both willingness-to-pay experiments

I choose:

Option A	Option B	Option C
$1.14	$1.14	I do not choose to purchase either option A or B
Required pasture access	Required pasture access	
Dehorning/disbudding permitted	Dehorning/disbudding permitted	
Labeled free of high fructose corn syrup	Labeled free of high fructose corn syrup	
Nonfat	Lowfat	

Figure 4. Willingness-to-pay traditional method choice scenario example

I choose:

Option A	Option B	Option C
$1.14	$0.79	I do not choose to purchase either option A or B
Labeled free of high fructose corn syrup	Labeled free of high fructose corn syrup	
Lowfat	Nonfat	

Figure 5. Willingness-to-pay new method animal welfare selected least important example

I choose:

Option A	Option B	Option C
$1.14	$0.79	I do not choose to purchase either option A or B
Required pasture access	Required pasture access	
Dehorning/disbudding permitted	Dehorning/disbudding not permitted	

Figure 6. Willingness-to-pay new method product labeling selected least important example.

3.8 New willingness-to-pay data collection method appendices

Appendix A. Information provided to respondents
The next portion of this survey presents you with hypothetical traditional yogurt purchasing scenarios that you could face in a retail store where you typically shop. The two products that will be presented in each scenario possess the same characteristics (e.g., similar color, brand, flavor, etc.) except for varying levels of the attributes presented below. Prices vary for each product.

For each scenario, please select the 6 oz cup of traditional yogurt that you would purchase, or neither, if you would not purchase either traditional cup of yogurt. For your information in interpreting alternative cups of yogurt:

Typical single-serve traditional yogurt cups are 6 oz in size. Thus, price refers to the cost per 6 oz cup of traditional yogurt: $0.40 per cup, $0.79 per cup, and $1.14 per cup.

Animal welfare attributes
Pasture access refers to the ability of dairy cattle to access grass pasture and not be confined solely to indoor production facilities
-Required pasture access means the animal was raised on an operation providing animals with access to grass pasture
-Pasture access not required indicates that no claims regarding access to grass pasture are being made

Dehorning/disbudding refers to the removal of horns/horn buds to insure dairy cows do not have horns
-Dehorning/disbudding not permitted means the animal was raised on an operation that does not dehorn/disbud cattle.
-Dehorning/disbudding permitted indicates that no claims regarding dehorning are being made

Product attributes
High-fructose corn syrup is a modified corn syrup that has an increased amount of fructose made from the enzymatic conversion of glucose and that is widely used as a sweetener
-Labeled free of high fructose corn syrup means the product does not contain high fructose corn syrup
-No high fructose corn syrup labeling claim means that no claims regarding the use of high fructose corn syrup are being made on this product

The amount of fat in yogurt depends on the type of milk the yogurt is made from
- Nonfat indicates the yogurt is made from skim milk and contains less than 0.5% milkfat
- Lowfat indicates the yogurt is made from skim milk and contains between 2% and 0.5% milkfat

The experience from previous surveys is that people often state a higher willingness-to-pay than what a person actually is willing-to-pay for the good. It is important that you make your selections like you would if you were actually facing these choices in your retail purchase decisions, noting that allocation of funds to these products means you will have less money available for other purchases.

CHAPTER 4. DEVELOPMENT, IMPLEMENTATION, AND DATA QUALTIY ASSESMENT OF A MORE EFFICIENT METHOD OF BEST-WORST SCALING DATA COLLECTION

4.1 Best-worst scaling introduction

Discrete choice experiments are frequently used to collect data that facilitates determination and understanding of consumer preferences. Best-worst scaling (BWS), first used by Finn and Louviere to study food safety (Finn and Louviere, 1992), is one method of discrete choice experiments that results in the relative ranking of attributes. BWS scaling involves presenting respondents with a subset of a greater pool of attributes, and asking them to select the best and worst, the most important and least important, etc., from the subset provided to them. These combinations of subsets are defined as choice scenarios, and several choice scenarios are required to establish the continuum of rank. The combination of the number of attributes per choice scenario and number of choice scenarios in a given design is statistically determined, and researchers often have several design options with the same statistical power as measured by efficiency (Johnsen et al., 2013). BWS has advantages over other ranking methods, such as Likert scales, because the method forces respondents to make trade-offs (Lusk and Briggeman, 2009). Additionally, when using BWS numbers are not associated with the ranking, which avoids the issues of respondents assigning different values to numbers as well as cultural differences between numbers when conducting international studies (Auger and Louviere, 2007).

Choice experiments, including but not limited to BWS experiments, are often included as part of a longer survey instrument. Studies have shown that long surveys can cause fatigue which may result in poor data quality (Galesic et al., 2009). For example, when presented with the length of the survey in minutes response rate decreased (Galesic et al., 2019). Additionally, for longer surveys open ended questions were shorter, response rate for individual questions decreased, and there was less variability in grid type questions (Galesic et al., 2019). A full factorial design for a BWS experiment would include every possible combination of attributes, and the continuum of preference would be determined by the respondent’s choices (Louviere, Flynn and Marley, 2015). Due to length constraints, it is impractical to use the full factorial design, so researchers employ a partial factorial method often designed using readily available software programs (Flynn, 2010), such as the SAS %MktBSize macro (SAS, 2018).

Although frequently used, until recently, little discussion has surrounded the possible impact of BWS design choice on results. Byrd et al. (2018) found differences in relative rank and preference share size of the attributes between two BWS designs. The two designs included the same six attributes, but one design presented respondents with two attributes per choice scenario for a total of fifteen choice scenarios, while the other design presented respondents with three attributes per choice scenario for a total of ten choice scenarios. Bir et al. (2019a) built on this work by comparing two designs that included the same nine attributes, one that presented respondents with three attributes per choice scenario for twelve choice scenarios, and another that presented eight attributes per choice scenario for nine choice scenarios. Bir et al. (2019a) also found differences in both attribute rank and size of preference shares between both designs. Although there were differences found between two different designs in both of these experiments, Byrd et al. (2018) and Bir et al. (2019a), it is not possible to determine if the differences resulted from the number of attributes or choice scenarios presented to the respondent, or both.

Response efficiency is the measurement error that results from cognitive effects that result in inattention to choice questions or unobserved contextual influences (Johnsen et al., 2013). Statistical efficiency is that in large samples, if the distribution tends toward normality, the statistic with the least probable error is chosen (Fisher, 1922). Researchers are often forced to make trade-offs between statistical efficiency and response efficiency. Cognitive effects that result in poor-quality responses in discrete choice experiments can include simplifying heuristics (Johnsen et al., 2013; Alemu et al., 2013; Scarpa et al., 2012), respondent fatigue (Johnsen et al., 2013, Galesic et al., 2009; Day et al., 2012), confusion or misunderstanding (Johnsen et al., 2013; Day et al., 2012), and inattention resulting from hypothetical bias (Johnsen et al., 2013). Statistical efficiency can be improved by including a large number of difficult trade-off questions, the opposite of what improves response efficiency. Response efficiency improves by asking a smaller number of easier trade-off questions. Sample size also impacts statistical power, larger samples shrink the inverse of the square root of the sample size, resulting in smaller confidence intervals (Johnsen et al., 2013).

This manuscript presents a new method of BWS data collection, with the objective of improving response efficiency, and in turn, data quality. The new method strives to improve response efficiency by minimizing the number of choice scenarios presented to respondents by minimizing the number of attributes in component models while maintaining a larger number of attributes in the aggregate. Minimizing the number of choice scenarios can often be achieved by increasing the number of attributes shown to a respondents in each choice scenario within the BWS, or by decreasing the total number of attributes studied. The key to the new BWS data collection method is that it allows for a decreased number of choice scenarios while holding the number of attributes within a choice scenario constant, and without decreasing the number of attributes studied. The new method uses an initial filter question to determine the group of attributes drawn from a larger set that individual respondents do not find important. The respondent then participates in a tailored BWS design that does not include those, predetermined as unimportant attributes, resulting in a smaller experimental design overall. In aggregate, over the entire sample of respondents, the continuum of all attributes included in the study can be determined. Johnson et al. (2013) outlines that decreasing the number of choices a respondent must make, which the new method achieves, should increase response efficiency. Additionally, removing attributes that respondents would never choose as important may decrease incidences of ANA.

This analysis employs the new BWS data collection method by eliciting consumer preferences for fluid dairy milk attributes and compares results and measures of data quality to the traditional method of BWS. The objectives of this manuscript are to 1) establish a new method of BWS data collection with the objective of possibly improving data quality through response efficiency and 2) compare the new BWS data collection method to the traditional BWS method by comparing the size of preference shares, relative ranking of attributes by preference share, the number of incidences of ANA, and the number of incidences of transitivity violation and violators.

4.2 Best-worst scaling materials and methods

Consumer preferences for attributes of fluid dairy milk were used as a case study to compare results of the traditional BWS method to the new BWS data collection method. In BWS, respondents are asked to select the best or worst, most important or least important, most ethical or least ethical, etc. attribute from a subset of attributes presented (choice scenario) (Louviere, Flynn and Marley, 2015; Lusk and Briggeman, 2009). The number of attributes presented in a choice scenario can vary, and is statistically determined based on the number of attributes included in the experiment (Louviere, Flynn and Marley, 2015). In both BWS methods respondents were presented with a series of choice scenarios. Within each choice scenario respondents were asked to choose the most important and least important attribute when making a fluid dairy milk purchase out of the attributes presented to them in that particular choice scenario. Nine attributes of fluid milk were included in this study: container material, rbST-free, price, container size, fat content, humane handling of cattle, brand, required pasture access for cattle, and cattle fed an organic diet. Most consumers are familiar with fluid dairy milk, which is important when introducing a novel methodological approach for data collection.

The survey instrument, designed to collect basic demographic information as well as the traditional BWS and the new BWS data collection methods in choice experiments, was distributed in April 2016 using Qualtrics, an online survey tool. Seven hundred and fifty respondents participated in the traditional and new BWS data collection methods. Lightspeed GMI, which hosts a large opt-in panel, was used to obtain survey respondents who were required to be 18 years of age or older. The sample was targeted to be representative of the US population in terms of gender, income, education, and geographical region of residence as defined by the Census Bureau Regions and Divisions (US Census Bureau, 2015). Respondents were randomly selected to participate in either the traditional BWS method first followed by the new BWS data collection method or participate in the new BWS data collection method first followed by the traditional BWS method (Figure 7). The two groups were designed to help mitigate, in aggregate, the potential for differences in the two methods due to order effects and also allowed for each respondent to participate in both methods. Therefore, when comparing the two methods, the sub-samples were comprised of the same respondents who were presented the methods in different orders. The traditional BWS and new BWS data collection methods were compared by evaluating differences in preference share size, rank, number of incidences of ANA, and number of incidences of transitivity.

4.2 Best-worst scaling materials and methods

Consumer preferences for attributes of fluid dairy milk were used as a case study to compare results of the traditional BWS method to the new BWS data collection method. In BWS, respondents are asked to select the best or worst, most important or least important, most ethical or least ethical, etc. attribute from a subset of attributes presented (choice scenario) (Louviere, Flynn and Marley, 2015; Lusk and Briggeman, 2009). The number of attributes presented in a choice scenario can vary, and is statistically determined based on the number of attributes included in the experiment (Louviere, Flynn and Marley, 2015). In both BWS methods respondents were presented with a series of choice scenarios. Within each choice scenario respondents were asked to choose the most important and least important attribute when making a fluid dairy milk purchase out of the attributes presented to them in that particular choice scenario.

The survey instrument, designed to collect basic demographic information as well as the traditional BWS and the new BWS data collection methods in choice experiments, was distributed in April 2017 using Qualtrics, an online survey tool. Seven hundred and fifty respondents participated in the traditional and new BWS data collection methods. Lightspeed GMI, which hosts a large opt-in panel, was used to obtain survey respondents who were required to be 18 years of age or older. The sample was targeted to be representative of the US population in terms of gender, income, education, and geographical region of residence as defined by the Census Bureau Regions and Divisions (US Census Bureau, 2015). Respondents were randomly selected to participate in either the traditional BWS method first followed by the new BWS data collection method or participate in the new BWS data collection method first followed by the traditional BWS method (Figure 7). The two groups were designed to help mitigate, in aggregate, the potential for differences in the two methods due to order effects and also allowed for each respondent to participate in both methods. Therefore, when comparing the two methods, the sub-samples were comprised of the same respondents who were presented the methods in different orders. The traditional BWS and new BWS data collection methods were compared by evaluating differences in preference share size, rank, number of incidences of ANA, and number of incidences of transitivity.

4.2.1 Traditional best-worst scaling methods

The traditional BWS experiment was designed using the SAS %MktBSize macro which determines balanced incomplete block designs (SAS, 2018). With nine attributes there were a total of five balanced designs to choose from. The number of attributes presented in a choice scenario ranged from three to eight and the number of choice scenarios ranged from nine to eighteen. The selected design presented respondents with three attributes per choice scenario, for a total of twelve choice scenarios. Each attribute appeared in the design four times.

Respondent choices were employed to determine the relative share of preference, or relative level of importance of each attribute. Each attribute's location on the continuum from most important to least important was determined using the respondents' choices of the most important and least important attributes from each choice scenario. The location of attribute j on the scale of most important to least important is represented by λ_j. Thus, how important a respondent views a particular attribute, which is unobservable to researchers, for respondent *i* is:

$$I_{ij} = \lambda_j + \varepsilon_{ij} \tag{14}$$

where ε_{ij} is the random error term. The probability that the respondent *i* chooses attribute *j* as the most important attribute and attribute *k* as least important is the probability that the difference between I_{ij} and I_{ik} is greater than all differences available from the available choices. Assuming the error term is independently and identically distributed Type I extreme value, the probability of choosing a given most/least important combination takes the multinomial logit form (Lusk and Briggeman, 2009), represented by:

$$Prob(j = best \cap k = worst) = \frac{e^{\lambda_j - \lambda_k}}{\sum_{l=1}^{J}\sum_{m=1}^{J} e^{\lambda_l - \lambda_m} - J}. \tag{15}$$

The parameter λ_j, which represents how important attribute j is relative to the least important attribute, was estimated using the maximum likelihood estimation. To prevent multicollinearity one attribute must be normalized to zero (Lusk and Briggeman, 2009).

A random parameters logit (RPL) model was specified to allow for continuous heterogeneity among individuals, following Lusk and Briggeman (2009). The individual-specific parameter estimates from the RPL model were used to calculate individual-specific preference shares. The parameters are not directly intuitive to interpret, so shares of preferences are calculated to facilitate the ease of interpretation (Train, 2009). The shares of preferences are calculated as:

$$share_j = \frac{e^{\lambda_j}}{\sum_{k=1}^{J} e^{\lambda_k}} \tag{16}$$

and must necessarily sum to one across the 9 attributes. The calculated preference share for each attribute is the forecasted probability that each attribute is chosen as the most important (Wolf and Tonsor, 2013). Estimation was conducted using NLOGIT 6.0. The individual level preference shares of the RPL model for each attribute were then averaged to represent the mean preference share of the sample. Standard deviations for the preference shares of each attribute were also determined in order to calculate confidence intervals for each preference share. Confidence intervals were calculated using the following formula:

$$confidence\ interval_j = mean \pm (1.96 * \left(\frac{Standard\ Deviation}{SQRT(Sample\ size)}\right). \tag{17}$$

A 95% confidence interval was achieved by subtracting from the mean for the lower bound, adding for the upper bound, and using a *z* score of 1.96.

4.2.2 New best-worst data collection method

The objective of the new BWS data collection method was to institute a technique to establish a continuum from most important to least important attribute (in aggregate) while minimizing fatigue, through fewer choice scenarios and by minimizing the number of attributes included. The new BWS data collection method takes into account the idea that respondents each face attributes that will necessarily be at the bottom of their list of attributes ranked in importance.

According to Train (2009) the logit probability for an alternative is never exactly zero. This is clear when considering the equation for the logit choice probabilities (equation 16). When λ_i decreases, the exponential in the numerator of equation 16 approaches zero as λ_i approaches $-\infty$ and the share of preferences approaches zero. Since the probability only approaches zero, and is never exactly zero, if the attribute has no chance of being chosen by the respondent, the researcher can exclude it from the choice set (Train, 2009). The first question in the new method (Figure 8) was designed to determine the category of attributes the respondent finds least important. For the newly proposed experimental design, the 9 attributes were grouped into three attribute categories: animal welfare attributes (pasture access, humane handling), product labeling attributes (fat content, organic diet, rBST free, brand) and physical appearance attributes (container material, container size). Price remained independent of the categories and was included in all component BWS experimental designs due its consistent importance in other studies (Harwood and Drake, 2018; Bir et al., 2019b, Lusk and Briggeman, 2009). After identifying their least important category of attributes each respondent participated in a component BWS choice experiment that did not include the attributes from the category the respondent indicated was least important to them. Fundamentally the respondent was able to efficiently opt-out of seeing any attributes belonging to their self-reported least important category of attributes.

One of the benefits of the new BWS data collection method is that respondents answered fewer choice scenarios (the maximum number was ten) when participating in the new BWS data collection method when compared to the twelve choice scenarios in the traditional method, while holding the number of attributes shown in each choice scenario to constant at three. For the animal welfare and physical appearance attributes selected least important component BWS models, respondents completed seven choice scenarios. Each attribute appeared in a choice scenario three times for the animal welfare and physical appearance component BWS models. For the product labeling attributes selected least important component BWS model, respondents completed ten choice scenarios and each attribute appeared in a choice scenario six times. By intention, the three component BWS designs included different numbers of attributes. In order to combine the component BWS models, each component model (animal welfare attributes as least important model, product labeling attributes as least important model, and physical appearance attributes as least important model) was estimated separately and preference shares were calculated following equations 14-16. Once the preference shares for each individual were calculated, a preference share of zero was assigned to the attributes not included in the component model the respondent participated in, as determined by their selection of that category as least important. Using the same method as the traditional BWS, the average and standard deviation for each attribute's preference share were calculated. Confidence intervals were calculated using equation 17. The overlapping confidence interval method (Schenker and Gentleman, 2001) was used to determine if preference shares were statistically different within and between BWS models and methods.

4.2.3 Inferred attribute non-attendance

Incidences of inferred ANA were determined (Hess and Hensher, 2010). The coefficient of variation was calculated as:

$$Coefficient\ of\ variation = \left|\frac{individual\ standard\ deviation}{individual\ coefficient}\right|. \qquad (18)$$

A threshold of 1 was used to determine incidences of ANA for the traditional BWS method. Widmar and Ortega (2014) evaluated thresholds of 1, 2, and 3 when evaluating inferred ANA in WTP choice experiments. They found that the results were not very sensitive to the threshold chosen. A threshold of one was used in this analysis as a conservative choice. If the coefficient of variation exceeded the cut off of 1, the attribute for that respondent was re-coded -888. As outlined by Green (2016), the specific code -888 was used to signal values that were deliberately omitted from the data set by the individual respondent. The model was then re-estimated with incidences of ANA coded for each individual and each attribute. Individual preference shares were calculated, and the average, standard deviation, and confidence interval were calculated for each attribute using equations 16-17. For the new BWS data collection method, incidences of ANA were determined using a threshold of 1 within the component BWS models individually. Each new method component BWS model was re-estimated with the incidences of ANA coded, individual respondents were assigned preference shares of zero for attributes depending on the component

BWS model they participated in, and averages, standard deviations and confidence intervals for the aggregate of the component models were calculated using equations 16-17.

4.2.4 Best-worst scaling transitivity violations

The frequency of transitivity violations was determined for both the new BWS and traditional BWS methods. The number of transitivity violations were determined for each component model of the new method individually, and then aggregated to determine the number of violations for the new BWS data collection method. Custom Python code which relied on directed graphs was employed by Bir et al. (2019a) to determine of transitivity violations for each respondent.

Following Bir et al. (2019a) a respondents choices could be represented as:

D>A	E>D	F>A	E>G	I>D	C>F	F>I	A>C	G>C

where each letter represents an attribute and each box represents a choice scenario. Each letter (attribute) is then converted to a node. Each greater than connection between attributes is converted to an edge, which is a directed edge within the directed graph (Figures 9-11). The program stored values in an adjacency matrix during configuration (Bir et al., 2019a). Given an n-node graph G = (V, E) where V = {$v_1, v_2, \ldots, v_n$}, the adjacency matrix G is the *n x n* matrix $A_G = \{a_{ij}\}$ where

$$a_{ij} = \begin{cases} 1 & if \;\; i \to j \;\in E \\ 0 & otherwise \end{cases}. \tag{19}$$

A respondent's true preferences are never known by the researcher; therefore, it is not possible to always determine the exact number of transitivity violations that occur, at best a minimum and maximum number of possible transitivity violations can be reported. To compare between the new BWS data collection method and the traditional method, the number of violators as a percentage of respondents was calculated. For the traditional method and the aggregate new BWS data collection method, the number of respondents were the same. However, the number of respondents differed between the three component BWS models of the new BWS method, and expressing the violators as a percentage allowed for statistical comparisons. Determining a method for comparing the number of violations was not as clear-cut. The number of potential violations differed between the new BWS and traditional methods. In order to have a comparable percentage, the number of possible violations per model was calculated as the number of choice scenarios multiplied by the number of respondents who participated in that model. For example, the animal welfare selected least important model consisted of 7 choice scenarios, and 207 respondents participated in this model, so there was a total number of possible violations of 1,149. The number of possible violations were summed across the three component BWS models to determine the total number of possible violations for the new BWS method model. The percentage of violations, and violators were statistically compared using the test of proportions. In order to evaluate the impact of transitivity violators, and to compare their impact on preference shares to the uncorrected models, all models were re-estimated, as outlined in the above sections, with transitivity violators removed. Respondents who had at least one transitivity violation were considered a violator for the purposes of this analysis.

4.3 Best-worst scaling results

Seven hundred and fifty respondents participated in both the traditional BWS method and the new BWS data collection method. Demographics between the self-sorted least important attributes differed (Table 11). A higher percentage of males (49%) selected animal welfare as least important when compared to the physical appearance and product. A lower percentage of respondent aged 25-34 (12%) were in the group physical appearance selected least important. For those who animal welfare selected least important, there was a lower percentage of respondents aged 45-54 (15%). A lower percentage of respondents aged 66 and older (18%) selected physical product labeling as least important when compare to animal welfare and physical appearance.

A lower percentage of respondents with an income of \$25,000-\$49,999 selected animal welfare as least important. Conversely, for respondents with an income of \$75,000-\$99,999 a higher percentage of respondents (17%) selected animal welfare as least important when compared to the physical appearance and product labeling. For respondents with an income of \$75,000-\$99,999 a higher percentage selected animal welfare as least important when compared to physical appearance. Lower percentages of respondents with less than a high school education (7%) and an associate's degree or bachelor's degree (24%) selected product labeling as least important. Lower percentages of respondents with a graduate degree or professional degree (8%) selected physical appearance, and lower percentages of high school graduates (23%) selected animal welfare as least important.

4.3.1 Traditional and new best-worst scaling data collection method results

For both the uncorrected (for either transitivity violations or ANA) traditional and new BWS data collection methods the top attribute (price) and the bottom attribute (container material) were the same (Table 12 and 13). However, the size of the preference share for price in the uncorrected new BWS data collection method was statistically higher than the traditional method. There was not a statistical difference in the size of the preference share for the lowest ranked attribute (container material) between the two methods. There were differences, in terms of order of attributes and size of preference shares, between the uncorrected traditional and new BWS data collection methods for the middle-ranked attributes. For the traditional method, the maximum number of ties in rank, as determined by overlapping confidence intervals, is three, which occurred once. For the new BWS data collection method, the maximum number of ties in rank was two, which occurred three times. The relative order between the two methods differed for fat content, humane handling, required pasture, cattle fed an organic diet, and brand. The size of the preference shares differed for rBST free, fat content, required pasture access and cattle fed an organic diet.

4.3.2 Data quality comparisons between data collection methods

When comparing the percentage of respondents who exhibited ANA between the traditional and new BWS data collection methods, for every attribute with the exception of price, higher percentages of respondents exhibited ANA while participating in the traditional method (Table 15). The ANA-corrected traditional method was statistically different in terms of size of preference share for all attributes with the exception of: rBST free, price, and container size when compared to the uncorrected traditional method (Table 13). The ANA-corrected new BWS data collection method differed from the uncorrected new BWS data collection method in terms of size of preference share for all attributes with the exception of container size, and fat content. The size of attributes preference shares differed between the ANA-corrected traditional method and the ANA-corrected new BWS data collection method for all but one attribute, rBST free. Similar to the results of the uncorrected methods, despite difference in the size of preference shares, the highest ranked attribute (price) was the same between the two ANA-corrected methods, and the lowest ranked attribute (container material) was the same between the two ANA-corrected methods. However, for the ANA-corrected new BWS data collection method there was a tie for last between container material and brand. Rankings differed between the ANA-corrected traditional and ANA-corrected new BWS data collection method for all attributes with the exception of price, rBST free, and container material. Both ANA-corrected methods had a maximum number of three ties, and the ANA-corrected new BWS data collection method had an additional tie between two other attributes.

The percentage of transitivity violators, defined as committing at least one transitivity violation, was not statistically different between the new BWS data collection method and the traditional method (Table 14). Interestingly only 10 respondents were violators in both the new method and the traditional method, which indicates the actual people committing violations were different in the two models. The total number of violations were statistically higher in the new BWS data collection method when compared to the traditional method. When considering the maximum number of transitivity violations, again the percentage of violations were statistically higher in the new BWS data collection method. Within the three new BWS component models, differences in the number of violations and the number of violators exists. The product labeling least important component model had a higher percentage of minimum violations, maximum violations, and violators when compared to physical appearance and animal welfare selected least important component models. Animal welfare selected least important and physical appearance selected least important had the same BWS design (show 3, 7 choice scenarios). Interestingly, the models did not differ significantly in the percentage of minimum violations and the percentage of violators.

Comparing the uncorrected traditional BWS model and the transitivity-corrected traditional BWS model, the relative rank changed for three attributes: cattle fed an organic diet, container size and required pasture (Table 14). However, the size of the preference shares did not differ between the uncorrected and transitivity-corrected traditional models. For the new BWS data collection method the relative rank also changed for three attributes-required pasture, container size, and brand- between the uncorrected and transitivity-corrected models. The size of the preference share between the uncorrected and transitivity-corrected new BWS data collection models differed for humane handling, container size, and container material. Interestingly for all uncorrected and transitivity-corrected models, price was always ranked first and container material was always ranked last. There were many differences between the transitivity-corrected new BWS data collection method and the traditional method. With the exception of price and container material, the relative ranking differed for every attribute between the two transitivity-corrected models. Additionally, the size of the preference shares were statistically different between the transitivity-corrected models for all attributes, with the exception of brand.

4.4 Best-worst scaling discussion

While participating in the new BWS data collection method, it was unsurprising that higher percentages of men self-selected animal welfare as least important when compared to the other categories. Female respondents exhibited increased concern for animal welfare in studies by Morgan et al. (2016), Vanhonacker et al. (2007), and McKendree et al. (2014). In a study of Finnish consumers, Yrjola, and Kola (2004) found that respondents with lower incomes believed animal welfare was a more serious problem in Finnish agriculture. In a phone survey of US respondents Prickett (2008) found that those with a higher income were less likely to state they considered animal welfare at the grocery store. Conversely, Lagerkvist and Hess (2010) found in a meta-analysis that high income was a strong explanatory variable for consumer willingness-to-pay for farm animal welfare. In this study, there were only statistical differences in one of the lower income groups ($25,000-$49,999) and they were less likely to select that animal welfare was least important.

The results of BWS experiments are sensitive to the presentation of attributes, in terms of how many attributes are presented in a choice scenario (Bir et al., 2019a; Byrd et al., 2018). This experiment accounted for those effects by employing only models with choice scenarios that included three attributes. When adjusting the number of attributes presented in a choice scenario between two and three, Byrd et al. (2018) found that in terms of ranking, the top and bottom attributes were consistent between the two models. However, the size of the preference share for the attributes did change (Byrd et al., 2018). Note that due to the cardinal nature of preference shares, a change in preference share size is an important distinction (Wolf and Tonsor, 2013). Similar results were found by Bir et al. (2019a) when comparing models with choice scenarios of three and eight. Their results showed that the top and bottom in terms of rank remained relatively steady between the three attributes shown per choice scenario model and the eight attributes shown per choice scenario model, as well as between models presented to respondents in differing order (Bir et al., 2019a). However, the size of the preference share did vary even amongst the top and bottom ranked attributes (Bir et al., 2019a). Given previous results found in the literature, it is perhaps unsurprising that between the new BWS data collection method and the traditional BWS collection method employed in this experiment, the top and bottom attributes did not change in terms of rank. Interestingly, for the attributes brand and container material which were in the bottom, or bottom two for both the traditional method and the new BWS data collection method- the preference shares were not statistically different across methods. This consistency exists even though 38% of respondents in the new BWS data collection method did not participate in a model with container material, 35% did not participate in a model with brand and participated in fewer choice scenarios.

It is possible that the larger preference share for price resulting from the new BWS data collection method may be due to its appearance in all new BWS data component models. As previously stated, a logit probability for an alternative is never exactly zero (Train, 2009). Therefore the preference share for price was never exactly zero (the lowest for any individual respondent was 0.4%), unlike the other attributes which were set to zero if the respondent indicated they were not important. It is possible that price's inclusion in all models may have resulted in an inflation of that particular preference share; however, the rank between the two methods for price did not differ, in both cases price was ranked solidly first. In further applications of this method, depending on the attributes included, it would be possible to include all attributes in a sorting question so that all attributes have the 'opportunity' of being chosen as unimportant and therefore have a zero preference share.

Accounting for instances of ANA was one technique used to evaluate and compare the new BWS data collection method and the traditional method. It was hypothesized that because ANA may be caused by respondents simplifying choice tasks by ignoring attributes (Alemu et al., 2013; Scarpa et al., 2012) the new BWS data collection method may serve to lessen incidences of ANA. Fewer incidences of ANA did occur in the new BWS data collection method when compared to the traditional method. However, correcting for incidences of ANA still yielded results that were statistically different in terms of preference share size as well as attribute rank. Bir et al. (2019a) analyzed the effect of ANA using the same BWS experimental design used in the tradition BWS method for this work, a show three attribute per choice scenario, nine attribute, twelve choice scenario design. Bir et al. (2019a) found that for the model shown earlier to respondents the rank of the top and bottom attributes did not change between the ANA-corrected and uncorrected models; but, the rank differed for the model shown later. In this experiment, the top and bottom attributes did not change in terms of rank between the ANA-corrected and uncorrected models for the traditional and new BWS data collection methods. However, the size of the preference shares did significantly change between corrected and uncorrected models in both instances. Unlike BWS, ANA has been studied more intensively in the willingness-to-pay literature. Despite a higher level of attention, an evaluation of two of the common methods for ANA correction in willingness-to-pay, stated and inferred ANA, resulted in statistically different results (Bir et al., 2018). The causes and implications of ANA in choice experiments, both BWS and willingness-to-pay, are unclear. Further studies are needed to determine what behavior is being captured when accounting for ANA, and whether this practice results in meaningful differences in results.

Consistent with the findings of Byrd et al. (2018), Bir et al. (2019a), and the ANA analysis in this experiment, despite correcting for transitivity violations the top (price) and bottom (container material) attributes in terms of relative ranking were immovable. Studying the same nine attributes for fluid dairy milk, price sorted to the top of the relative ranking and container material sorted to the bottom of the relative ranking for both Bir et al. (2019a) and Bir et al. (2019b). The new BWS data collection had more statistically differently sized preference shares between the uncorrected and corrected for transitivity violation models when compared to the traditional BWS method. Despite this, correcting for transitivity resulted in a difference in relative ranking for three attributes in both methods, all occurring in attributes in the middle of the relative ranking. Using the same experimental design as the traditional method employed in this work, Bir et al. (2019a) reported an incidence of transitivity violators of 12% for the model shown earlier in their survey to respondents and an incidence of 9% for the model shown later to respondents. This is similar to the incidence of transitivity violators of 8.8% found in the traditional BWS method from this work. When considering the component models of the BWS new data collection method, the product labeling selected as least important model appears to be the main driver of the higher number of transitivity violators and violations. It cannot be determined if the respondents who participated in the product labeling selected least important model were fundamentally different (recall this was a self-selected group), or if it was the particular design of the BWS experiment that resulted in the significant difference. The product labeling model resulted in 10 choice scenarios, which was higher than the animal welfare selected and physical appearance selected models. However, the traditional BWS method had 12 choice scenarios, and approximately half the number of violators in terms of percentage. Interestingly, the frequency of appearance for a given attribute was greatest for product labeling selected (6) when compared to the traditional method (4), and the animal welfare and physical appearance selected models (3). Further research is necessary to evaluate the impact that nuances in BWS experimental design choice result in beyond the number of attributes presented to respondents in a choice scenario or the number of choice scenarios.

4.5 Best-worst scaling conclusion

A new BWS data collection method was introduced in this manuscript, that builds on the traditional BWS method by decreasing the number of attributes shown to individuals, and decreasing the number of choice scenarios required, while in aggregate allowing for the establishment of the continuum from least to most important for a larger number of attributes. The same set of respondents participated in both the new BWS data collection method and the traditional BWS method to determine the importance of nine different attributes when making a fluid milk purchase. The top (price) and bottom (container material) attributes in terms of relative ranking did not change between the new BWS data collection method and the traditional BWS method. Additionally, correcting for ANA and violators of transitivity did not impact the relative ranking of top and bottom attributes for either method. The relative ranking and size of preference share did differ between the new BWS data collection method, the traditional BWS method, and ANA/transitivity-corrected models. The new BWS data collection method resulted in fewer incidences of ANA for all attributes with the exception of one. However, there was not a statistical difference in the number of transitivity violators between the new and traditional BWS methods. The new BWS method provides researchers the opportunity to minimize the number of choice scenarios and attributes presented to respondents. In longer, fatigue prone survey instruments the new BWS data collection method may be useful to minimize the number of questions presented to individual respondents while maintaining in terms of transitivity violations or improving data quality in terms of ANA violations. Further research is needed to determine how the new BWS data collection method compares to the traditional method for other products, and with varying BWS composite models for the new method.

4.6 Best-worst scaling tables and figures

Table 11. Demographics of entire sample, respondents who selected animal welfare as least important, physical appearance as least important, and product labeling as least important.

Demographic Variable	US Census	Percent (%) of All Respondents n=750	Animal welfare selected least important N=207	Physical appearance selected least important N=284	Product labeling least important N=259
Gender					
Male	49%	47%	57%a	43%b	46%b
Age					
18-24	13%	12%	13%a	13%a	10%a
25-34	18%	15%	16%ab	12%a	17%b
35-44	16%	15%	16%a	16%a	14%a
45-54	17%	18%	15%a	18%b	21%b
55-65	17%	18%	17%a	16%a	20%a
66-88	19%	22%	22%ab	25%a	18%b
Annual pre-tax household income					
$0-$24,999	22%	25%	25%a	26%a	24%a
$25,000-$49,999	23%	25%	18%a	29%b	26%b
$50,000-$74,999	17%	17%	17%a	18%a	17%a
$75,000-$99,999	12%	13%	17%a	10%b	12%b
$100,000 and higher	26%	20%	23%a	17%b	21%ab
Educational Background					
Less than High School	13%	5%	3%a	4%a	7%b
High school graduate (includes equivalency)	28%	30%	23%a	33%b	32%b
Some college, no degree	21%	22%	20%a	24%a	22%a
Associate's degree or Bachelor's degree	27%	30%	37%a	31%a	24%b
Graduate or professional degree	12%	13%	17%a	8%b	15%a
Region of Residence					
Northeast	18%	19%	19%ab	21%a	17%b
South	21%	38%	40%a	38%a	38%a
Midwest	38%	21%	19%a	21%a	22%a
West	24%	22%	22%a	20%a	23%a

ꝉMatching letters indicate that demographic is not statistically different between the three self-selected categories animal welfare, product labeling, and physical appearance selected as least important. For example, the percentage of males is statistically different between animal welfare and product labeling and animal welfare and physical appearance, but is not statistically different between product labeling and physical appearance.

Table 12. Random parameters logit results, and preference shares traditional method and new best-worst data collection method.

Milk Attributes	Traditional Method N=750			New Best-Worst Data Collection Method						
	RPL			RPL animal welfare selected least important model N=207		RPL physical appearance selected least important model N=284		RPL product labeling least important model N=259		
	Coefficient	Standard Deviation	Shares of Preference	Coefficient	Standard Deviation	Coefficient	Standard Deviation	Coefficient	Standard Deviation	Preference Shares for new model[1]
Container Material	-0.592***	0.687***	5%	0.095	0.177	____	____	-0.710***	0.466***	4%
	0.042	0.052		0.083	0.248	____	____	0.055	0.066	
rBST Free	-0.104**	0.876***	9%	0.117	0.916***	0.150**	0.183	____	____	7%
	0.044	0.053		0.103	0.130	0.062	0.150	____	____	
Price	0.639	1.782***	25%	2.038***	1.904***	0.230***	0.230	1.000***	1.870***	34%
	0.0692	0.077		0.199	0.178	0.063	0.147	0.133	0.139	
Container Size	-0.077	1.052***	8%	0.851***	0.997***	____	____	-0.068	0.550***	8%
	0.050	0.063		0.113	0.122	____	____	0.056	0.066	
Fat content	0.646***	1.083***	17%	1.326***	0.973***	0.218**	1.084***	____	____	12%
	0.053	0.067		0.111	0.108	0.089	0.105	____	____	
Humane Handling	0.565***	0.576***	15%	____[2]	____	0.767***	0.661***	0.086*	0.031	15%
	0.041	0.058		____	____	0.078	0.099	0.044	0.061	
Brand	-0.451***	0.987***	6%	0.392***	0.881***	-0.895***	1.518***	____	____	5%
	0.045	0.049		0.099	0.127	0.126	0.130	____	____	
Required Pasture Access	0.061*	0.987*	8%	____	____	0.259***	0.374***	┼	┼	11%
	0.035	0.162		____	____	0.065	0.122			
Cattle fed an organic diet	┼	┼	8%	┼	┼	┼	┼	____	____	5%
								____	____	

[1]Calculated using the average of all individual respondent coefficients
[2]Crossed out boxes were not included in that BWS design
***1% significance of coefficient, **5% significance of coefficient, *1% significance of coefficient
┼ dropped to avoid multicollinearity

Table 13. Confidence intervals of preference shares for traditional method and new best-worst data collection method.

	Traditional Method with transitivity violators N=750					New Best-Worst Data Collection Method with transitivity violators N=750					Models have statistically different preference shares[1]
	Lower bound	Mean	Upper bound	Rank		Lower bound	Mean	Upper bound	Rank		
Container Material	0.043	0.045	0.047	7		0.038	0.042	0.046	6		No
rBST Free	0.083	0.088	0.094	4		0.062	0.067	0.072	4		Yes
Price	0.231	0.246	0.261	1		0.315	0.335	0.355	1		Yes
Container Size	0.079	0.083	0.087	4		0.071	0.077	0.083	4		No
Fat Content	0.166	0.174	0.182	2		0.108	0.117	0.126	3		Yes
Humane Handling	0.140	0.146	0.152	3		0.140	0.149	0.158	2		No
Brand	0.058	0.062	0.066	6		0.047	0.052	0.058	5		No
Required Pasture Access	0.078	0.080	0.083	4		0.100	0.107	0.113	3		Yes
Cattle Fed an Organic Diet	0.073	0.075	0.077	5		0.050	0.054	0.058	5		Yes
	Traditional Method ANA corrected N=750				Models w/ and w/o ANA correction are statistically different	New Best-Worst Data Collection Method ANA corrected N=750				Models w/ and w/o ANA correction are statistically different	Models have statistically different preference shares[1]
	Lower bound	Mean	Upper bound	Rank		Lower bound	Mean	Upper bound	Rank		
Container Material	0.031	0.033	0.036	8	Yes	0.069	0.072	0.075	5	Yes	Yes
rBST Free	0.084	0.094	0.103	4	No	0.080	0.083	0.086	4	Yes	No
Price	0.259	0.280	0.301	1	No	0.207	0.224	0.241	1	Yes	Yes
Container Size	0.069	0.076	0.083	5	No	0.084	0.087	0.090	4	No	Yes
Fat Content	0.198	0.213	0.227	2	Yes	0.114	0.120	0.126	3	No	Yes
Humane Handling	0.128	0.137	0.146	3	Yes	0.161	0.172	0.183	2	Yes	Yes
Brand	0.043	0.048	0.052	6/7	Yes	0.071	0.074	0.078	5	Yes	Yes
Required Pasture Access	0.060	0.064	0.067	6	Yes	0.082	0.085	0.088	4	Yes	Yes
Cattle Fed an Organic Diet	0.053	0.056	0.059	6	Yes	0.080	0.082	0.085	4	Yes	Yes
	Traditional Method without transitivity violators N=684				Models w/ and w/o violators are statistically different	New Best-Worst Data Collection Method without transitivity violators N=684				Models w/ and w/o violators are statistically different	Models have statistically different preference shares[1]
	Lower bound	Mean	Upper bound	Rank		Lower bound	Mean	Upper bound	Rank		
Container Material	0.041	0.043	0.045	7	No	0.018	0.023	0.027	7	Yes	Yes
rBST Free	0.085	0.092	0.099	4	No	0.062	0.067	0.072	5	No	Yes
Price	0.243	0.261	0.279	1	No	0.303	0.328	0.352	1	No	Yes
Container Size	0.075	0.079	0.084	5	No	0.035	0.043	0.050	6	Yes	Yes
Fat Content	0.162	0.171	0.181	2	No	0.114	0.125	0.136	3	No	Yes
Humane Handling	0.140	0.147	0.154	3	No	0.197	0.214	0.232	2	Yes	Yes
Brand	0.054	0.058	0.062	6	No	0.045	0.052	0.058	6	No	No
Required Pasture Access	0.075	0.078	0.080	5	No	0.089	0.097	0.104	4	No	Yes
Cattle Fed an Organic Diet	0.069	0.072	0.074	5	No	0.048	0.052	0.056	6	No	Yes

[1] Traditional method and New Best-Worst Data Collection method have statistically different preference shares based on overlapping confidence intervals for either uncorrected, ANA corrected, or transitivity corrected model

Table 14. Number of attribute-non-attendance and transitivity occurrences for each attribute for the traditional and new-best-worst data collection method.

	Traditional Method N=750	New Best-Worst Data Collection Method			
		Animal welfare selected Least Important n=207	Physical appearance selected least important model n=284	Product labeling least important model n=259	Total Number for New Best-Worst Data Collection Method n=750
ANA occurrences					
Container Material	229 (30.5%)[Ω]	154	___________	22	176 (23.5%)[Ω]
rBST Free	356 (47.5%)[Ω]	145	170	___________	315 (42.0%)[Ω]
Price	216 (28.8%)[Ω]	45	115	260	420 (56.0%)[Ω]
Container Size	331 (44.1%)[Ω]	69	___________	188	257 (34.3%)[Ω]
Fat content	272 (36.3%)[Ω]	30	145	___________	175 (23.3%)[Ω]
Humane Handling	207 (27.6%)[Ω]	___________	85	1	86 (11.5%)[Ω]
Brand	265 (35.3%)[Ω]	108	83	___________	191 (25.5%)[Ω]
Required Pasture Access	421 (56.1%)[Ω]	___________	184	┼	184 (24.5%)[Ω]
Cattle fed an organic diet	┼	┼	┼	___________	___________
Number of violations	**Number of respondents**				
1 violation minimum analysis	58	11	11	28	50
2 violation minimum analysis	8	1	0	8	9
3 violation minimum analysis	0	0	0	7	7
4 violation minimum analysis	0	0	0	0	0
Total number of violations (minimum)[1]	74 (0.82%[2]) [Ω]	13 (1.13%[2])a	11 (0.5%[2])a	65 (2.5%[2])b	89 (1.5%[2]) [Ω]
1 violation maximum analysis	36	9	10	18	37
2 violation maximum analysis	20	3	1	5	9
3 violation maximum analysis	5	0	0	8	8
4 violation maximum analysis	5	0	0	9	9
5 violation maximum analysis	0	0	0	2	2
6 violation maximum analysis	0	0	0	1	1
Total number of violations (maximum)[1]	111(1.2%[2]) [Ω]	15 (1.3%[2])a	12 (0.6%[2])b	104 (4.0%[2])c	131 (2.3%[2]) [Ω]
Number of respondents who committed at least 1 violation	66 (8.8%)	12 (5.7%)a	11 (3.4%)a	43 (16.6%)b	66 (8.8%)

[1]A respondents true preference is unknown by the researcher, therefore in some cases the number of possible violations is ambiguous. Therefore, the minimum and maximum number of violations is given.

[2]Percentage calculated out of total number of opportunities for violation (calculated as the number of respondents multiplied by number of choice scenarios) from left to right: 9000,1149, 1988, 2590, 5727

[3]Matching letters indicates the percentage of violations or violators is not statistically different between the new method sub-models, differing letters indicate they are statistically different across the row

┼dropped to avoid multicollinearity

[Ω]The percentage of ANA incidences or transitivity occurrences is statistically different between the traditional and new method

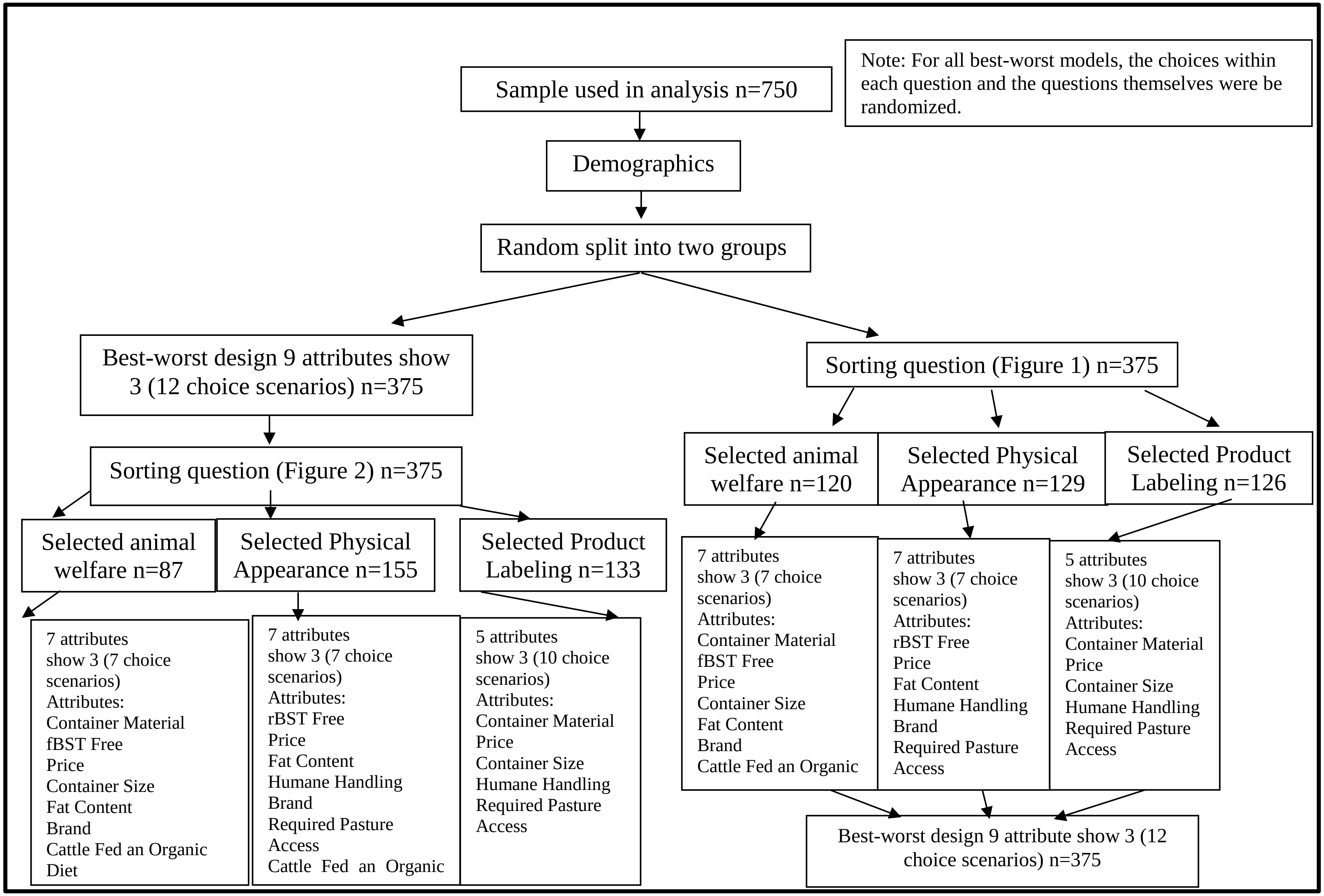

Figure 7. Best-worst scaling new data collection method survey design.

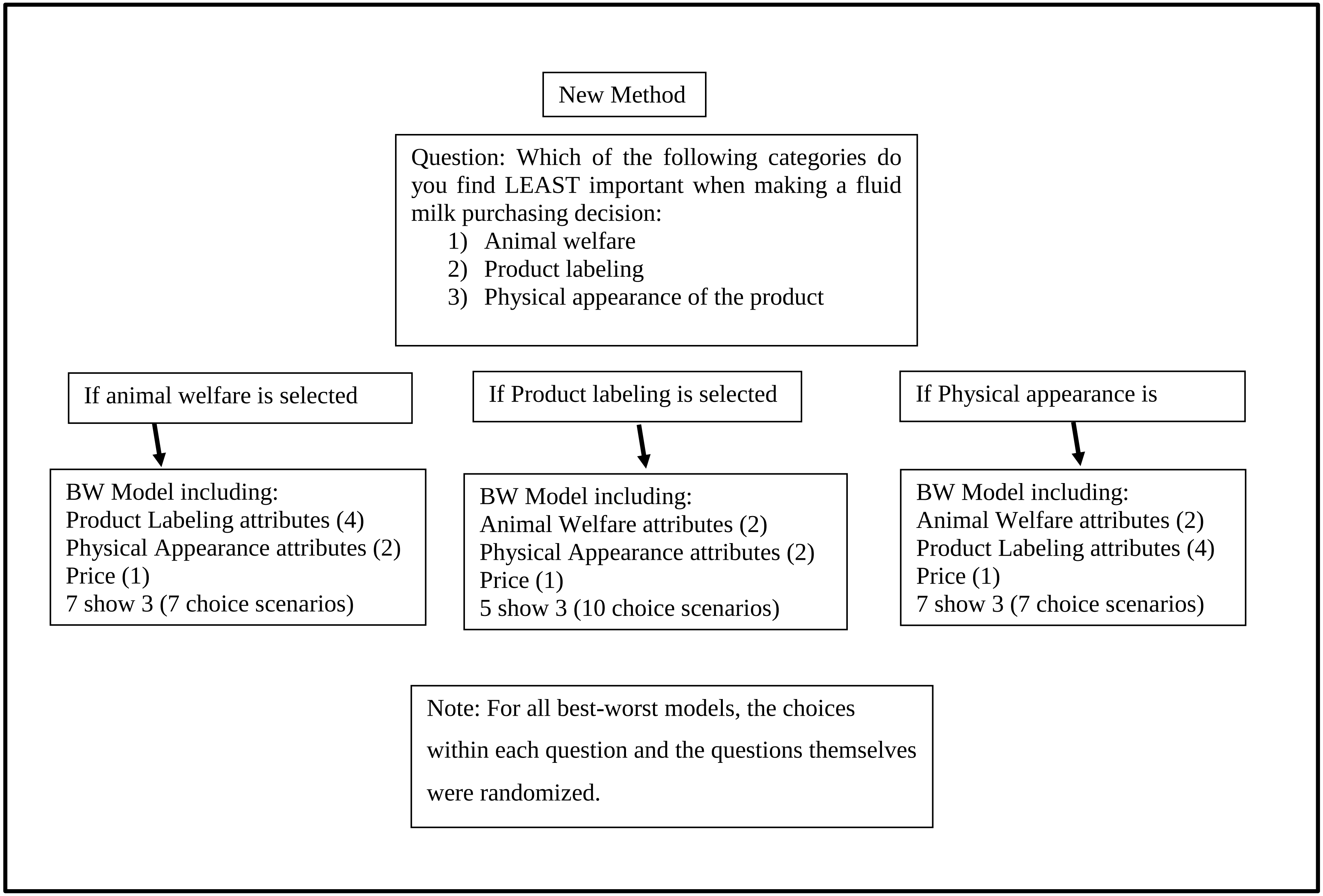

Figure 8. Flow of new best-worst data collection method.

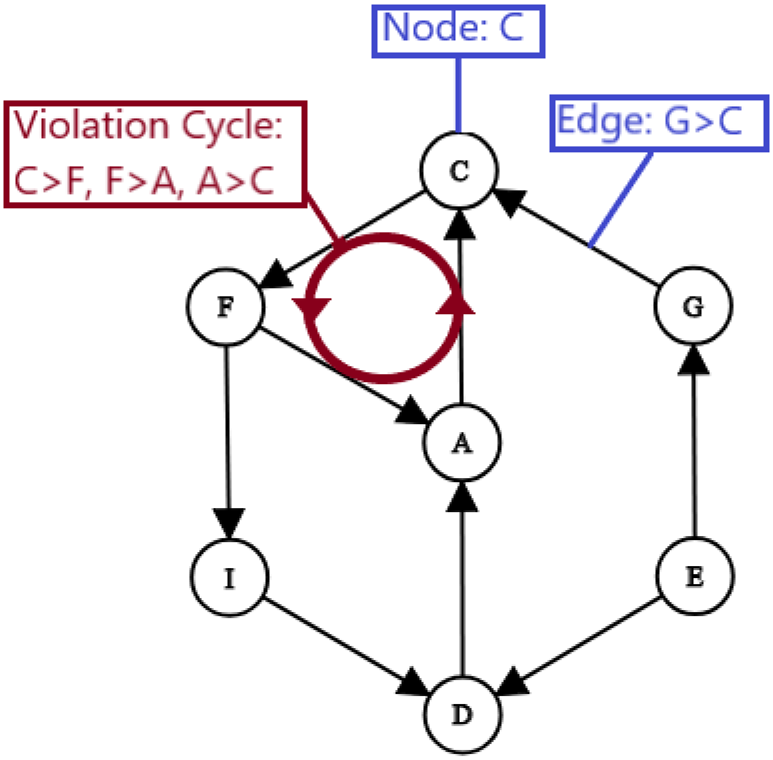

Figure 9 Anatomy of a directed graph from Bir et al. 2019a.

Note: Letters are equivalent to attributes. Circles represent nodes, and lines with arrows represent edges with direction. The red circle indicates a transitivity violation.

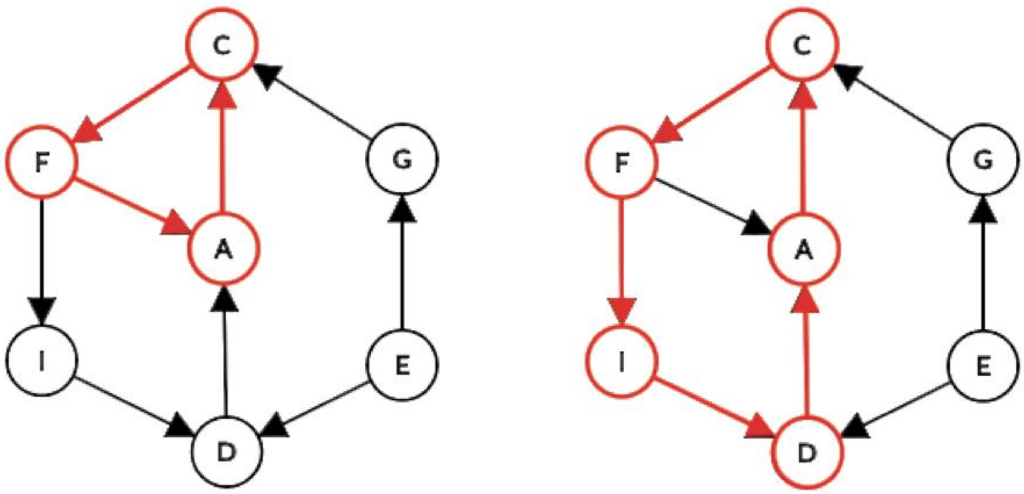

Figure 10. Example visual of the trimming process resulting in a violation of transitivity from Bir et al. 2019a.

Note: Letters are equivalent to attributes. Circles represent nodes, and lines with arrows represent edges with direction. Red lines indicate areas were the process 'circles' which indicates a violation in transitivity.

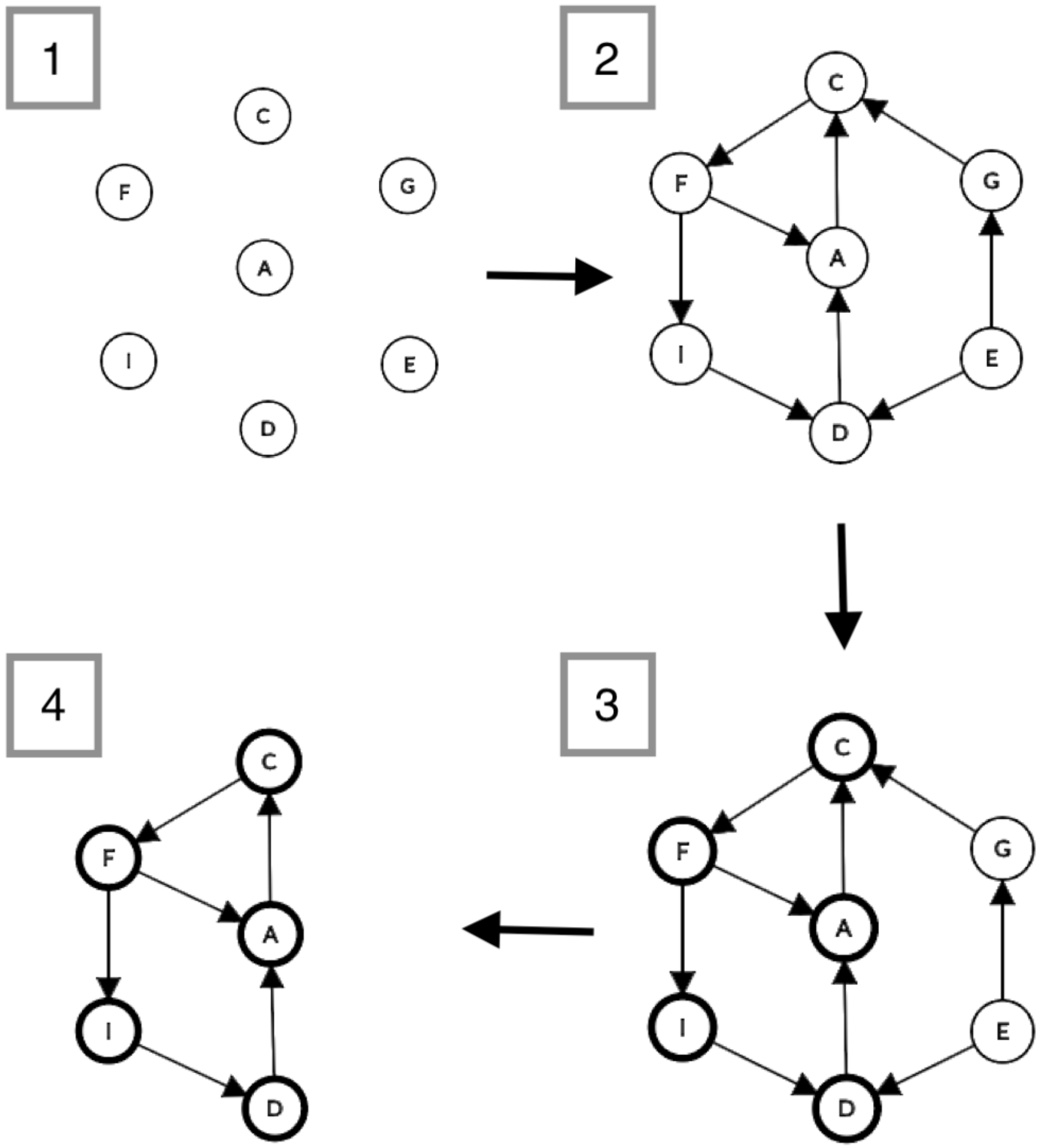

Figure 11. Steps to creating a directed graph and eliminating non-violator nodes from Bir et al. 2019a.

Note: Letters are equivalent to attributes. Circles represent nodes, and lines with arrows represent edges with direction. Step 1 is a simple statement of attributes (letters). Step 2 indicates the choices a respondent has made which establishes a rank between the attributes. Step 3 indicates the areas were the process ‘circles’ which indicates a violation in transitivity, marked by darkened nodes. Step 4 shows the trimmed nodes and edges with only transitivity violations remaining.

4.7 Best-worst scaling references

Alemu, M. H., Mørkbak, M. R., Olsen, S. B., & Jensen, C. L. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and resource economics*, 1-27.

Auger, P., Devinney, T. M., & Louviere, J. J. (2007). Using best–worst scaling methodology to investigate consumer ethical beliefs across countries. *Journal of Business Ethics, 70*(3), 299-326.

Bir, C., Widmar, N. O., Etumnu, C., Slipchenko, N., Polcyn, A., & Wolf, C. A. (2019a) Development and implementation of transitivity violation detection algorithm in discrete choice experiments. Manuscript submitted for publication.

Bir, C., Widmar, N. O., Wolf, C. A., Delgado, M. S. (2019b). Traditional attributes moo-ve over for some consumer segments: relative ranking of fluid milk attributes. *Appetite, 134*, 162-171.

Bir, C., Cummins, A. M., Widmar, N. O., & Wolf, C. A. (2018). Willingness-to-pay Estimates Informing Agribusiness Decision Making: A Cautionary Tale. *International Food and Agribusiness Management Review, 21*(7), 865-882.

Byrd, E. S., Widmar, N. J. O., & Gramig, B. M. (2018). Presentation matters: Number of attributes presented impacts estimated preferences. *Agribusiness, 34*(2), 377-389.

Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., Scarpa, R., Wang, P. (2012). Ordering effects and choice scenario awareness in repeat-response stated preference studies. *Journal of Environmental Economics and management, 63,* 73-91.

Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy & Marketing*, *11*(2), 12-25.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*(594-604), 309-368.

Flynn, T. N. (2010). Valuing citizen and patient preferences in health: recent developments in three types of best–worst scaling. *Expert review of pharmacoeconomics & outcomes research, 10*(3), 259-267.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349–360.

Harwood, W. S., & Drake, M. A. (2018). Identification and characterization of fluid milk consumer groups. *Journal of Dairy Science, 101*(10), 8860-8874.

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

Lagerkvist, C. J., & Hess, S. (2010). A meta-analysis of consumer willingness-to-pay for farm animal welfare. *European Review of Agricultural Economics, 38*(1), 55-78.

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). Best-worst scaling: Theory, methods and applications. Cambridge, U.K.: Cambridge University Press.

Lusk, J. L., & Briggeman, B. C. (2009). Food values. *American Journal of Agricultural Economics*, *91*(1), 184-196.

McKendree, M. S., Croney, C. C., Widmar, N. O. (2014). Bioethics Symposium II: Current factors influencing perceptions of animals and their welfare. *Journal of Animal Science*, *92*, 1821–1831.

Morgan, C. J., Croney, C. C., & Widmar, N. J. (2016). Exploring relationships between ethical consumption, lifestyle choices, and social responsibility. *Advances in Applied Sociology*, *6*, 199-216.

Prickett, R. W. (2008). Consumer preferences for farm animal welfare: Results from a telephone survey of US households. PhD dissertation, Oklahoma State University.

SAS. (2018). The %MktBSize Macro. Retrieved from: https://support.sas.com/rnd/app/macros/MktBSize/mktbsize.htm.

Scarpa, R., Zanoli, R., Bruschi, V., & Naspetti, S. (2012). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics*, *95*(1), 165-180.

Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *The American Statistician*, *55*(3), 182-186.

Train, K. E. (2009). Discrete choice methods with simulation. Cambridge U.K: Cambridge university press.

Yrjola, T., & Kola, J. (2004). Consumer preferences regarding multifunctional agriculture. *International Food and Agribusiness Management Review*, *7*(1030-2016-82670), 78-90.

US Census Bureau. (2015). Annual Estimates of the Resident Population for Selected Age Groups by Sex for the United States, States, Counties and Puerto Rico Commonwealth and Municipios: April 1, 2010 to July 1, 2015.

Vanhonacker, F., Verbeke, W., Van Poucke, E., & Tuyttens, F. (2007). Segmentation based on consumers' perceived importance and attitude toward farm animal welfare. *International Journal of Sociology of Agriculture and Food*, *15*(3), 91-107.

Widmar, N. J., & Ortega, D. L. (2014). Comparing Consumer Preferences for Livestock Production Process Attributes Across Products, Species, and Modeling Methods. *Journal of Agricultural and Applied Economics*, *46*(3), 1-17.

Wolf, C. A., & Tonsor, G. T. (2013). Dairy farmer policy preferences. *Journal of Agricultural and Resource Economics*, *38*(2), 220-234.