DO CONSUMERS REALLY KNOW

HOW MUCH THEY ARE WILLING TO PAY?

By

KRISTEN LYNNE KOVALSKY

Bachelor of Science in Food & Resource Economics

University of Florida

Gainesville, Florida

2010

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
May, 2012

DO CONSUMERS REALLY KNOW

HOW MUCH THEY ARE WILLING TO PAY?

Thesis Approved:

Dr. Jayson Lusk

Thesis Adviser

Dr. Bailey Norwood

Dr. Brian Adam

Dr. Sheryl A. Tucker

Dean of the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The standard economic analysis of consumer behavior assumes individuals know their preferences with certainty and behave accordingly. The assumption conveys confidence that past behaviors can be used to predict future choices, and it is the linchpin of economic welfare analysis, which assumes choices define preferences, which in turn define well-being. Unfortunately, the assumption of complete and stable preferences has been challenged by a number of empirical findings from psychology and behavioral economics.

One of the more prominent recent examples was provided by Ariely, Loewenstein, and Prelec (2003). They showed that seemingly innocuous anchors, such as a person's social security number, could significantly influence willingness-to-pay (WTP) for ordinary consumer products such as wines, chocolates, and keyboards. They interpret their results to imply that (p. 75), "The sensitivity of WTP to anchors suggests that consumers do not arrive at a choice or at a pricing task with an inventory of preexisting preferences." They go on to conclude that (p. 102), "If consumers' choices do not necessarily reflect true preferences, but are to a large extent arbitrary, then the claims of revealed preferences as a guide to public policy and the organization of economic exchange are weakened." Reviewing the behavioral economics literature, Kahneman and Thaler (2006, p. 222) reach a similar conclusion: "people do not always know what they will like; they often make systematic errors . . . and, as a result, fail to maximize their experienced utility."

Economists have responded to such findings in a variety of ways. Some have attempted to systematically model reference points and decision making and biases (e.g., Tversky and Kahneman 1991) while others have introduced differing approaches to re-conceptualize or rationalize economic welfare analysis (e.g., Bernheim and Rangel 2009; Sugden 2004). There has been comparatively little work studying the *process* by which people's preferences either are or become stable and coherent. Indeed, despite the arbitrary nature of WTP expressed in some settings, there is evidence to suggest that people are able to discover their preferences if given the opportunity to gain experience and receive feedback on their choices (Plott 1996; Plott and Zieler 2002). Plott's theory, known as the "discovered preference hypothesis," suggests that individuals might come to have coherent preferences, but these preferences may not be revealed in one-shot decisions. Rather, these preferences are discovered by information gathering, reasoning, and learning in repeated trials. Support for such a preference formation process can be found, for example, in findings that individuals with greater market experience are less likely to exhibit the endowment effect (e.g., List 2003), the preference reversal phenomena can be alleviated with market exposure (e.g., Chu and Chu 1990; Cherry, Crocker, and Shogren 2003; Cherry and Shogren 2007), and that anchoring effects can dissipate with repeated questioning (Batemen et al. 2008).

Such findings raise the possibility that instead of trying to force the theory to fit the behavior, we might adapt the analyses to a learning environment that allows individuals to discover their preferences. This is especially true of contingent valuation and other non-market valuation methods which use WTP estimates to infer welfare effects of policies. Braga and Starmer (2005) argue, for example, that behavioral anomalies do not necessarily rule out the use of WTP estimates in cost-benefit analysis, but rather highlight the need for subjects to better learn about their preferences.

In this spirit, Norwood and Lusk (2011) employed a new preference elicitation method that was meant to facilitate learning and rationality in the process of individuals stating their WTP values. In their method, the preference-discovery process is facilitated by a feedback mechanism that allows participants to review and revise their choices as the tradeoffs between choices are elucidated. Their approach forces a consistency in that statements of value and preference orderings are directly linked to utility.

In a sense, Norwood and Lusk's (2011) approach is a direct attempt to address one of the fundamental concerns Diamond and Hausman (1994) expressed with the contingent valuation method – that (p. 63) "the internal consistency problems come from an absence of preferences, not a flaw in survey methodology." The problem is presumably not that people *cannot* have preferences for the goods one often studies in contingent valuation studies, but rather (p. 62), "the lack of experience both in markets for environmental commodities and in the consequences of such decision." The method introduced by Norwood and Lusk (2011) imposes internal consistency while promoting learning and experience using an approach that provides feedback about the tradeoffs of the choices made.

Although the frameworks that allow (and even attempt to promote) preference learning have been offered in previous research such as that of Norwood and Lusk (2011) or Batemen et al. (2008), questions remain about the merits of the approaches, their ability to alleviate behavioral biases, and their predictive power. As a result, this research investigates how sensitive are WTP values obtained from a "preference learning" method similar to the one used in Norwood and Lusk (2011) to irrelevant anchors and how well the method performs in terms of predicting subsequent consumer choice as compared to other "one-shot" elicitation methods that vary the amount of preference learning that occurs during the valuation task.

The overall purpose of this research is to determine whether consumer preferences can be elicited in a systematic way that provides a rational estimation of WTP and a better prediction of consumer choice. This study uses a split-sample survey design to: (1) determine whether imposing internal consistency between preference orderings and valuations influences WTP and sensitivity to irrelevant cues and (2) determine the influence of imposing internal consistency and preference learning on predictive validity.

Background on Fluid Milk Product Attributes

Although much of the debate about contingent valuation has focused on WTP for non-market goods, practitioners also widely use preference elicitation methods for private goods for use in marketing, pricing, new-product introduction, and cost-benefit analyses (e.g., Huffman et al. 2003; Loureiro and Umberger 2003; Bernard and Bernard 2009; Olynk, Tonsor, and Wolf 2010). Preference stability and formation is equally important in such applications as well.

The empirical context for our study involves consumer preferences for a novel food attribute, an exemplar where consumers are likely unfamiliar and may well have unformed preferences that could lead to unstable choices. The agricultural industry is continuously developing new food products and methods of production to either fill niche markets or improve production efficiencies. The number of new food and beverage product introductions has followed an upward trend over the past two decades, with an average of 20,921 new products introduced each year from 2005 to 2009, compared to 9,653 new products introduced in 1992 (USDA ERS 2010). Many of these new food and beverage products feature product claims such as "gluten-free," "no trans fat," and "high-vitamin." In addition to these established product claims, over 100 *new* food and beverage product claims were identified in 2009.

Some product claims are specifically related to agricultural production practices and include things such as cage-free eggs, and grass-fed beef, and rBST-free milk, all of which are likely unfamiliar to most consumers. If consumers are indecisive and inconsistent in their preferences for these unfamiliar goods, they may violate one or more of the axioms of revealed preference, and their preference uncertainties may lead to inaccurate estimations of WTP, which could lead to biased estimates of potential market-share, optimal product prices, or benefits from a labeling police.

In this paper we use fluid milk to investigate consumer WTP for a multiattribute product. We chose fluid milk because it has a finite number of attributes and each attribute is easily defined or scaled in measureable units. For example, the attribute "Fat Content" has four discrete levels that consumers commonly refer to as Skim, Low Fat, Reduced Fat, and Whole. We also chose fluid milk because it is a staple product in the diets of many U.S. households, giving us the ability to recruit a sufficient sample of participants who consume milk on a regular basis and have underlying preferences for the various attributes of milk.

Although milk is generally considered a commodity-type product, U.S. consumers have demonstrated increasing interest in milk production methods (e.g., Bernard and Barnard 2009; Olynk, Tonsor, and Wolf 2010). Retailers have recognized the demand for differentiated milk products and have responded by offering milk products that are labeled to indicate the farm-level production practices. Examples of labels on fluid milk products include “organic,” “locally produced,” and “rBST-free.”

An “rBST-free” label on milk products indicates that the milk has been produced from cows not treated with the controversial artificial growth hormone rBST, a supplement for the naturally occurring cow hormone BST. Products with this label must qualify the claim with a mandatory supplementary statement from the FDA indicating that there is no significant difference between milk from rBST and non-rBST treated cows. However, scientific evidence from the FDA that rBST milk is safe for consumption has not inhibited consumer demand for rBST-free milk. Even mass-merchandiser Wal-Mart has responded to the consumer demand for rBST-free milk, announcing in March 2008 that it would begin sourcing its private label milk (sold under the name “Great Value”) exclusively from cows that have not been treated with rBST (Wal-Mart 2008).

Milk products that are rBST-free, organic, or locally produced are currently available in many U.S. grocery retail stores and data is available to determine the premiums that consumers are willing to pay for products with these attributes. However, some milk attributes have yet to become available in the marketplace, such as milk from cloned cows. Although the U.S. Food and Drug Administration (FDA) declared milk from cloned cows safe for consumption in January 2008 (FDA 2008), dairy companies have voluntarily refrained from selling such milk in the marketplace, reportedly due to perceived consumer concerns. If dairy companies were to choose to sell milk from cloned cows under current regulations, the FDA would not require a label to identify these products (FDA 2009), just as it does not require a label to indicate milk from cows treated with rBST. However, it is possible that dairy processors could follow in the footsteps of the rBST-free movement by choosing to label milk products as "clone-free" to indicate that cloned cows were not used in production. Existing research suggests that consumers are willing to pay premiums as large as three times those for rBST-free and organic milk in order to ensure that they are purchasing milk from non-cloned cows (Brooks and Lusk 2010). However, given the novelty of the attribute, and the general lack of familiarity with the cloning process, a key concern is whether such previous survey-based preference measures are well-formed.

CHAPTER II

METHODS AND DATA

Sample

Data was collected through an online survey administered in August 2011 by the survey-software company, Qualtrics. Participants were recruited by e-mail to complete an online survey about their preferences for milk products. In order to maintain homogeny between the different experimental treatments, participation was limited to females age 30-51 who live in the limited geographic region including Texas, Oklahoma, Kansas, or Missouri and who consume fluid milk at least once per week. In return for completing the survey, participants earned online currency that they could redeem for cash or gift cards for restaurants and online retailers.

The survey was sent to 1,395 individuals registered in the Qualtrics survey panel, 1,058 of whom met the minimum criteria and completed the survey in its entirety, implying a response rate of 76%. Participants were randomly assigned to one of three treatments. The three treatment samples were approximately equal in size with 351 participants in Treatment 1, 356 in Treatment 2, and 351 in Treatment 3. Across all treatments, 100% of the participants were female, 44% had a Bachelor's degree or higher, and 53% had children under the age of 12. In terms of race, 77% of the participants were white, 9% were black, 9% were Hispanic, and the remaining 5% were of other races. In terms of income, 30% had an annual household income less than $40,000, 42% had income between $40,000 and $79,999 and 28% had income of $80,000 or greater. Geographic representation was similar to distribution of the U.S. population in this region with 8% from Kansas, 20% from Missouri, 10% from Oklahoma, and 62% from Texas. Characteristics of the participants between treatments were relatively uniform and are outlined in table 1.

Each of the three treatments used a different elicitation method to estimate WTP for milk products with varying levels of eight attributes outlined in table 2. Participants were told to assume that all milk products in the questions were pasteurized and homogenized with Vitamins A and D added. Because some households purchase multiple types of milk for different members of the household, participants were also told to answer the questions as if they were purchasing milk for their own personal consumption.

Overview of Experiment

A brief description of our design is as follows. Treatment 1, which we call "Preference Learning + Feedback," included a series of questions designed to promote preference learning by using a version of the self-explicated conjoint method used in the marketing literature (e.g., Hoepfl and Huber 1970; Srinivasan 1988; Srinivasan and Park 1997). It has been argued that self-explicated approaches are advantageous over traditional conjoint measurement when handling a large number of attributes because the cognitive burden is minimized when participants are questioned separately on each attribute and attribute level (Green and Srinivasan 1990). We coupled the self-explicated approach with a feedback mechanism which forced internal consistency such that WTP was directly related to the prior ratings/rankings and could only be changed by reviewing and revising prior preference orderings. Treatment 2, which we call “Preference Learning,” also used the preference learning approach through the self-explicated method but dropped the feedback mechanism and simply asked participants for an outright statement of their WTP for certain milk products. Treatment 3 was the control and it only involved asking participants for outright statements of their WTP for certain milk products. By comparing WTP from Treatment 1 to Treatment 2, we can ascertain the effect of forced internal consistency. By comparing WTP from Treatment 2 to Treatment 3, we can determine the effect of preference learning. Details of each treatment are provided in the subsection that follows; a summary of the steps used to collect data in each treatment is outlined in figure 1.

Treatment 1: Preference Learning + Feedback

The initial steps in the Treatment 1 survey followed that of a conventional self-explicated approach, similar to that of Green and Helsen (1989) and Srinivasan and Park (1997). Participants began by rating the desirability of each attribute level on a scale of 1 to 10, where 1 is very undesirable and 10 is very desirable, assuming that all other characteristics are the same (e.g., see figure 2a). After evaluating the desirability of each attribute, participants were asked to indicate the relative importance of each attribute by using a constant-sum scale to allocate 100 points across the set of attributes, where attributes that are more important are allocated more points (see figure 2b). A traditional self-explicated approach would end at this point, and used the responses to infer consumer preferences for different milk products. We refer to the above rating/ranking steps as the preference learning portion of the design as they allow participants to evaluate each attribute individually, and ask themselves how important each attribute is in relation to one another using simple rating and ranking scales. Following Norwood and Lusk (2011), we went one step further and heightened the preference learning process by giving participants feedback on the consequences of their choices in terms of the WTP that was implied, and then gave them the opportunity to revise their previous ratings/rankings to achieve WTP values that were more consistent with their true desires.

Using the ratings/rankings we were able to calculate each individual's self-explicated partworth utility for a given attribute-level as:

$$(1)\ P_{ijl} = D_{ijl} W_{ij}$$

where P_{ijl} is individual i's self-explicated partworth utility for attribute j's l^{th} level, D_{ijl} is individual i's desirability rating for the l^{th} level of attribute j, and W_{ij} is individual i's importance weight for attribute j.

An individual's non-price utility for a particular milk product with a set of attributes at given levels is determined by summing the partworth utilities for all attributes at their respective levels:

$$(2)\ U_{it} = \sum_{j=1}^{J} V_{jl} P_{ijl}$$

where U_{it} is individual i's non-price utility for product t, J is the number of attributes, and V_{jl} is a dummy variable that equals 1 if product t contains the lth level of the jth attribute, and 0 otherwise.

An individual's relative WTP premium for one product over another is determined by dividing the difference in utilities by the importance weight of the price attribute:

(3) $WTP_{itk} = (U_{it} - U_{ik}) / W_{iP}$

where WTP_{itk} is individual i's WTP premium for product t over product k, U_{it} is individual i's utility for product t, U_{iv} is individual i's utility for product k, and W_{iP} is individual i's importance weight on the price attribute.

We used equation (3) to calculate each individual's WTP premium for the two milk options, Milk A over Milk B, where Milk A had attributes that are considered to be more desirable to consumers (table 3). Specifically, when compared to Milk B, Milk A had an extended expiration date, was from a branded dairy company, was produced on a farm closer to the grocery store, and was produced from conventional cows rather than cloned cows. Previous research suggests that consumers prefer milk with an extended expiration date (Tsiros and Heilman 2005), and are willing to pay premiums for branded milk (Bernard and Mathios 2005), milk that is "locally produced" (Wolf, Tonsor, and Olynk 2011), and milk that is "clone-free" (Brooks and Lusk 2010).

Before revealing the WTP values to the participant, we set the stage for our test for rationality. Following Ariely, Loewenstein, and Prelec (2003), we used an anchor test to test the sensitivity of the WTP estimations to normatively irrelevant information. Participants were asked to enter the last three digits of their phone number (figure 3a). We divided the inputted number by 100 to convert it to a dollar-and-cents[1] amount, and then asked participants if they would be willing to pay that amount as a premium for Milk

[1] The last three digits of a phone number range from 000 to 999, so the dollars-and-cents premiums used in the anchor test could range from \$0.00 to \$9.99. Although some premiums in this range may seem obtrusively high for a gallon of milk, premiums of such magnitude already exist for certain attributes (e.g., organic milk) and are possible for other attributes. For example, previous research suggests that consumers are willing to pay \$4.71 per gallon to avoid milk from cloned cows (Brooks and Lusk 2010).

A over Milk B (figure 3b). For example, if the last three digits of an individual's phone number were 376, we asked if they were willing to pay a $3.76 premium to have Milk A over Milk B. In principle, one's WTP should have nothing to do with their phone number; however, as Ariely, Loewenstein, and Prelec (2003) show, the irrelevant anchor number might be correlated with the participant's maximum stated WTP, which is obtained in the next step.

After completing the first part of the anchor test, we revealed the subject's WTP premium for Milk A over Milk B as calculated by equation (3). We asked if this was the maximum premium that they would be willing to pay (figure 3c). If they agreed that this was in fact the maximum premium that they would pay, their statements of value were consistent with their utility, so they did not revise their choices and proceeded to the next step of the survey. Contrarily, if they disagreed with this premium, their statements of value were *not* consistent with their utility, and they were therefore redirected to the screen where they entered the relative importance weight of each attribute. At this point, they adjusted their attribute importance point allocations and observed how these adjustments changed the value of the premium. For example, if an attribute is overvalued in the importance point allocations, WTP for a product with that attribute would be inflated. Once finished adjusting the attribute importance values, equation (3) was recalculated using the updated ratings and the revised WTP premium was re-stated. Respondents could go back and change again if they were still unsatisfied with the resulting WTP value.

In the final step, we included a holdout choice between four milk options so as to test the predictive validity of the previously-obtained WTP estimates. In particular, participants were shown four milk products C, D, E, and F that had varying levels of attributes and varied price levels (see table 3 or figure 5b). Participants were asked to rank the products from 1 to 4 with 1 being the product that they were most likely to purchase and 4 being the product that they were least likely to purchase.

We can determine the predicted rankings using the attribute-based partworth utility information obtained in the preference learning and feedback exercise. We then compare the predicted rankings with the actual rankings to observe the predictive validity of this preference elicitation method with a feedback mechanism.

Treatment 2: Preference Learning

Treatment 2 followed the same preference learning procedures as in Treatment 1 (figure 2), which entailed rating the desirability of each of the attribute levels and then ranking the relative importance of the attributes using a constant-sum scale. In Treatment 2, however, we did not force internal consistency by reporting to subjects their implied WTP value as determined by equation (3) nor did we give the opportunity to review and revise their choices. By comparing behavior in Treatment 2 to that in Treatment 1, we can determine the marginal effects of providing feedback on WTP through forced internal consistency.

As in Treatment 1, after rating the desirability of each attribute level and assigning importance weights for each attribute, participants in Treatment 2 continued to the anchor test for rationality. As in Treatment 1, participants were asked to enter the last three digits of their phone number and then stated whether or not they would be willing to pay that amount as a premium for Milk A over Milk B (figure 4a and 4b). Instead of responding to the calculated WTP premium by revising their importance weights as in

Treatment 1, Treatment 2 participants were simply asked to enter the maximum premium that they would be willing to pay for Milk A over Milk B (figure 4c).

Following the anchor test for rationality, participants continued to the holdout choice exercise. Participants were first shown milk products C, D, E, and F without the assigned prices. Participants were asked to enter their maximum WTP for each product (figure 5a). In the next step they were shown the identical products but with the assigned prices. As in Treatment 1, participants were asked to rank the products from 1 to 4 with 1 being the product that they are most likely to purchase and 4 being the product that they are least likely to purchase (figure 5b).

We can make two sets of holdout choice predictions for Treatment 2 participants: 1) predictions based on the provided attribute-based utility information obtained from the preference learning exercise and 2) predictions based on the consumer surplus generated from the maximum stated WTP. As in Treatment 1, we then compare the predicted rankings with the actual rankings to observe the predictive validity of this preference elicitation method *without* a feedback mechanism.

Treatment 3: Control

Treatment 3 served as the control. Treatment 3 participants did not complete the preference learning exercise nor were they provided information on their implied WTP. Instead, Treatment 3 participants began the survey with the anchor test for rationality by entering the last three digits of their phone number (figure 4a). As in Treatments 1 and 2 they were asked whether or not they would be willing to pay that amount as a premium for Milk A over Milk B (figure 4b). As in Treatment 2, they were then asked to enter the maximum premium that they would be willing to pay for Milk A over Milk B (figure 4c).

After the anchor test for rationality, participants were shown the four milk products C, D, E, and F with descriptions of all attributes except the price attribute. Participants were asked to enter their maximum WTP for each product (figure 5a). After seeing the product prices in the next step, participants were asked to rank the products from 1 to 4 as in the other treatments (figure 5b). Because we do not have any attribute-based utility information for Treatment 3 participants, we can only make predictions based on the consumer surplus generated from the maximum stated WTP. As in Treatment 2, we predict that the participants will rank the products from largest consumer surplus to smallest consumer surplus.

Interpreting the Anchor Test for Rationality

In principle, there should be no correlation between the phone number and the maximum stated WTP. A correlation between the two variables would suggest WTP is influenced by irrelevant factors. The key question we ask is which treatment yields the highest and lowest correlation between the anchor and WTP, where significant correlations are an indicator of unstable preferences. Because Treatment 1 incorporates a feedback mechanism designed to force internal consistency, we would expect via the discovered preference hypothesis that Treatment 1 will facilitate the formation of rational preferences in that participants truly understand the tradeoffs that they have made to arrive at their WTP. Contrarily, the WTP estimations in Treatment 2 and Treatment 3 are subjective estimations of what one expects their WTP would be given what they *think* they know about their underlying preferences. Without a true understanding of their preferences, it is likely that participants in Treatments 2 and 3 will make arbitrary estimations of their WTP. Therefore, we hypothesize that WTP from Treatment 1

(Preference Learning + Feedback) will exhibit smaller correlation with anchor numbers than one-shot elicitation methods (Treatment 2 and Treatment 3). We test the hypothesis

(4) $\rho_1 = \rho_2 = \rho_3$

where ρ is the correlation coefficient between maximum WTP and the random anchor number for the subscribed treatment.

Interpreting the Holdout Choice for Predictive Validity

Predictive validity is a measure of the ability of a preference elicitation method to predict subsequent choice. We can use two measures to cross validate the participants' choices in the holdout set with our predictions of their choices: 1) incidence of first-choice hits and 2) the average correlation between the predicted rank and actual rank of the four milk options. The incidence of first-choice hits most closely resembles the "real world" in that consumers typically choose only one product in any given purchase occasion. We also use a Pearson correlation to cross validate the ranked positions because it allows us to assess the overall performance of each elicitation method and evaluate how alternative methods compare with respect to predicting the specific rank position of each product in the holdout profile.

We predict the product rankings for each participant in order of consumer surplus, where the product that generates the largest consumer surplus is the first choice and the product that generates the smallest consumer surplus is the last choice. In Treatment 1, we calculate consumer surplus relative to a base product, where WTP is calculated as the sum of partworth utilities divided by the importance weight of price. In Treatment 2, we have both attribute-based utility information and WTP values estimated by the participants, so we calculate consumer surplus in two ways: 1) using the attribute-based utility information to determine consumer surplus relative to a base product, and 2) the participant's stated WTP minus price. In Treatment 3 we calculate consumer surplus as the participant's stated WTP minus price.

We theorize that participants in Treatment 2 (Preference Learning) and Treatment 3 (Control) do not really know how much they are willing to pay for each of the products in the holdout profile because they do not have a complete understanding of their underlying preferences. Therefore, their stated WTP will lead to inaccurate estimations of consumer surplus, and we will not be able to correctly predict their subsequent rankings. We expect Treatment 1 (Preference Learning + Feedback) to exhibit a higher degree of predictive validity because of the internal consistency that has been forced upon the participants' preferences.

The incidence of first-choice hits is calculated as the percent of correct hits for all participants. We compare the incidence of first-choice hits for each treatment to test the hypothesis that WTP estimates from Treatment 1 will correctly predict the first choice more frequently than one-shot elicitation methods (Treatment 2 and Treatment 3).

To calculate the Pearson correlation between the predicted ranked sets and the actual ranked sets we first calculate the correlation between the sets for each individual participant. We then calculate the mean of these correlations to obtain the overall correlation for each treatment. Methods that exhibit a higher correlation between the predicted rankings and actual rankings have a higher degree of predictive validity. We compare the correlations of each treatment to test the hypothesis that WTP estimates from a Treatment 1 will correctly predict a holdout choice more frequently than one-shot elicitation methods (Treatment 2 and Treatment 3).

CHAPTER III

RESULTS

Of the 351 people in Treatment 1 (Preference Learning + Feedback), 45% revised their importance weights after being shown their computed WTP value via equation (3). This is much lower than the 99% of participants who revised their importance weights in the method used by Norwood and Lusk (2011). One possible reason for this difference is that we used an online survey tool, so a survey administrator was not present to reinforce the instructions or answer questions. Additionally, our survey software did not give participants the opportunity to revise their choices by default. Instead, our participants had to indicate that they wanted to revise their choices, and they were then redirected to the original importance weight screen. Therefore not all participants were able to immediately observe the tradeoffs that resulted from their choices unless they chose to revise their answers. Because less than half of the participants in Treatment 1 chose to revise their importance weights we separate the results into those that chose to revise and those that did not revise in addition to examining the results as a whole.

Also in Treatment 1, 13% of the participants assigned the price attribute an importance score of zero. If an individual truly has no importance value for price, that individual would, in theory, be willing to pay an infinite amount to have the product that they most prefer. One limitation of the survey software was that we were unable to convey an infinite WTP value. Rather, the software returned $0 when calculating the WTP for participants who entered zero for the importance of the price attribute. In retrospect, it would have been preferable to return some arbitrarily large number (like $100) as WTP in the event that a participant entered zero for the importance of the price attribute.

In what follows, we report the results from Treatment 1 in four categories:

1) *All Responses* – Inclusive of all participants (351 observations)
2) *Unrevised* – Inclusive of participants who did not revise their choices (193 observations)
3) *Revised* – Inclusive of participants who revised their choices (158 observations)
4) $W_{iP} > 0$ – Inclusive of participants who assigned the price attribute an importance score greater than zero, where W_{iP} is individual *i*'s importance weight on the price attribute (307 observations).

Preference Learning

The mean importance weights assigned to each attribute are outlined in table 4. In Treatment 1 (Preference Learning + Feedback), the price attribute was allocated the most points relative to all other attributes with an average of 35 points, followed by expiration date and fat content with an average of 14 points each. The allocation of importance weights followed a similar distribution in Treatment 2 (Preference Learning), with price, expiration date, and fat content as the three most important attributes with point allocations of 28, 18, and 14, respectively. Cloning was the least important attribute to most participants in both treatments, with 46% of participants in Treatment 1 and 47% of participants in Treatment 2 indicating that cloning had zero importance when making a milk purchase decision.

It is notable that participants who revised their choices in Treatment 1 allocated an average of 16 more points to the price attribute than participants who did not revise their choices (statistically significant at the $p < 0.0001$ level). In essence, this implies that people wanted to *lower* their WTP estimates after being shown what their prior ratings/rankings implied. It also suggests that simple self explicated studies that do not provide feedback are likely to lead to upwardly biased estimates of WTP.[2]

Anchor Test for Rationality

Table 6 reports the correlation between the last three digits of the participants' phone numbers and their stated WTP for Milk A over Milk B. The correlation reported for Treatment 1 (Preference Learning + Feedback) exhibits an insignificant correlation of 0.024 while the correlations reported for Treatments 2 (Preference Learning) and 3 (Control) exhibit significant, positive correlations of 0.102 and 0.262, respectively. Although the correlations for Treatments 2 and 3 are not particularly large in magnitude we are interested in the fact that they are statistically significant and greater than the insignificant correlation reported for Treatment 1. The results from Treatment 2 and Treatment 3 are in line with those of Ariely, Loewenstein, Prelec (2003) in that the WTP estimates appear to have been influenced by arbitrary information. Although participants

[2] It has been shown that hypothetical surveys tend to lead to inflated estimations of WTP values (List and Gallet 2001; Neill et al. 1994). Although it is possible that the imposition of forced internal consistency can alleviate the effects of hypothetical bias, all three treatments were subject to a certain degree of hypothetical bias as participants did not actually have to make an economic commitment to purchase any of the products used in the survey.

in Treatment 1 were also exposed to the arbitrary information, the results do not show evidence that this information influenced the WTP estimates.[3]

In comparing across treatments, we note that Treatment 3 (Control) exhibited the largest correlation, followed by Treatment 2 (Preference Learning) and Treatment 1 (Preference Learning + Feedback). From this we are able to ascertain that preference learning alone was able to lessen the effects of anchoring, but did not induce complete rationality. With the addition of a feedback mechanism, preference learning appears to have mitigated the effects of anchoring while inducing rationality.

Holdout Choice for Predictive Validity

Table 7 reports the incidence of first-choice hits for each treatment. Treatment 1 (Preference Learning + Feedback) correctly predicted the first choice for 47.0% of participants, while Treatments 2 (Preference Learning) and 3 (Control) each correctly predicted the top choice for 58.4% of participants. Although we report a lower hit rate for Treatment 1, we note that the hit rate performs better than random chance (25%). While preference learning alone did not improve the accuracy of predicting the first choice, the addition of a feedback mechanism reduced the accuracy.

Table 7 reports the correlations between our predicted ranked positions for each product and the participants' actual ranked positions. The correlations for Treatments 1, 2, and 3 were 0.3847, 0.501, and 0.491, respectively. Each correlation represents the mean of the correlations for all participants. In comparing across treatments, a higher

[3] Participants were asked if they would be willing to pay the anchor amount as a premium for Milk A over Milk B. Their response to this question (yes or no) does not influence the analysis of the anchor test, but we would expect that people should exhibit consistency between this answer and their actual stated WTP. In other words, if a person is *not* willing to pay the anchor amount, that person should provide an actual WTP value that is *lower* than the anchor amount. Not all participants exhibited such consistency, with 77.2% of participants in Treatment 1, 90.4% of participants in Treatment 2, and 90.0% of participants in Treatments 3 exhibiting consistency between these answers.

correlation indicates greater accuracy in identified the specific rank positions of the products. As with the hit rate for the top choice, preference learning alone did not improve the accuracy of predicting ranked sets, and the addition of a feedback mechanism reduced the accuracy.

Because Treatment 1 exhibited both a lower hit rate and a lower correlation between predicted rankings and actual rankings, we are unable to conclude that methods which force internal consistency subsequently lead to a better prediction of consumer choice. Under these circumstances, the results obtained from this type of method would be less useful when determining what would happen in a market setting, such as estimating the market share for a group of products.

Although these results are not in line with our hypothesis, they concur with those of Nordgren and Dijksterhuis (2009), who found that people who deliberated on their preferences actually had *decreased* preference consistency compared to those who made non-deliberative judgments. They showed that when rating the attractiveness or quality of a series of items, people who were told to think carefully about their choices were less consistent in replicating their original ratings when evaluating the same items on a second occasion. If we assume that participants in Treatment 1 (Preference Learning + Feedback) were subject to a higher degree of deliberation because we encouraged them to reconsider their choices, they would, according to the findings of Nordgren and Dijksterhuis (2009), be less likely to rank the holdout choices in a way that was consistent with the preferences indicated by the ratings and rankings in the initial step of the survey.

Preference Rankings Across Treatments

The top choices from the holdout profiles are outlined in table 8. Across all treatments, Milk C was most frequently ranked as the top choice followed by Milk E, Milk D, and Milk F. The average rank order of products (table 9) follows a similar pattern with no statistical difference between treatments (table 10). Because the average rank order of the holdout choices is the same across treatments, we surmise that participants did not change their preferences as a result of the preference learning feedback mechanism or any other preference elicitation method. Rather, the participants held their preferences constant while possibly by changing the underlying values that they hold for the various product attributes.

Because price was the most important attribute according to the mean importance weights in Treatment 1 and Treatment 2, one might speculate that the participants were most likely to purchase the product with the lowest price and least likely to purchase the product with the highest price. This is not the case, as Milk E, the product with the lowest price, was most often ranked as the second choice, and Milk D, the product with the highest price, was most often ranked as the third choice.

Fat content was also indicated as one of the most important attributes in Treatment 1 and Treatment 2. This lends to another possible explanation to the order of choices in the holdout profile, with the speculation that each participant's top choice was the product with the fat content that they typically purchase, regardless of the other attributes. However, this theory is also dismissed, as our results do not align with the distribution of fluid milk sales by fat content. Specifically, Reduced Fat (2%) milk had the highest proportion of total fluid milk sales in 2010 with 39% of total sales, followed by Whole, Skim, and Low Fat with 29%, 17%, and 15% of fluid milk sales, respectively (USDA ERS 2011). If participants ranked the products based solely on fat content, we would expect Milk E (Reduced Fat) to be the top choice followed by Milk F (Whole), Milk C (Skim), and Milk D (Low Fat). This is not the case, as Milk C (Skim) was actually ranked as the top choice and Milk F (Whole) was ranked as the bottom choice.

Because we are unable to conclude that participants selected an "obvious" choice in the holdout profile, it suggests that the participants were considering *all* of the information and not just the two most important attributes, price and fat content. These results are in line with the notion that people become more rational by changing their values, not their underlying preferences (Cherry, Crocker, and Shogren 2003).

Willingness to Pay for Milk Product Attributes

Using the data from Treatment 1 and Treatment 2, we made estimations for marginal WTP values for selected attributes of milk products, as outlined in tables 11 and 12. Because some of the calculated WTP values were obtrusively large (as high as $459 and as low as -$64 in Treatment 1), we bound the WTP values between -$30 and $30. For example, a WTP of $459 is entered as $30, and a WTP of -$64 is entered as -$30. We bound WTP between the specific amounts of -$30 and $30 because this was the maximum amount that participants were permitted to enter in Treatments 2 and 3. As with our previous results, we also report the data inclusive and exclusive of participants who placed zero importance on the price attribute.

Using the data from Treatment 1 and excluding participants with a zero importance score for price, our results show that consumers are, on average, willing to pay a premium of $1.82 per gallon to avoid milk from cloned cows and $1.77 per gallon to avoid milk from the offspring of cloned cows, assuming that all other product attributes are the same. The difference between these values is not statistically significant, indicating that consumers do not differentiate between milk from cloned cows and milk from the offspring of cloned cows. This is in line with the findings of Brooks and Lusk (2010), however, our estimation of the premium that consumers are willing to pay to avoid milk from cloned cows is significantly lower in magnitude. We attribute our low estimation to the fact that 46% of the participants in Treatment 1 indicated that price had zero importance when making a milk purchasing decision. Because nearly half of participants said that they would not consider cloning as an important attribute, the distribution of WTP to avoid milk from cloned cows is skewed right (figure 6).

Consumers are willing to pay an increasing premium for milk as the distance from the farm to the store increases. Specifically, our results show that consumers are willing to pay a premium of $1.55 per gallon for milk produced 25 miles from the grocery store rather than milk produced over 500 miles from the grocery store. The underlying reasons why consumers prefer milk that is produced closer to the store are unknown, but we offer two possible explanations. First, it is possible that consumers associate a farm that is closer in geographical distance with a "local" or "family" farm. We chose not to use these specific words in our survey questions because of possible subjective interpretation, but previous research indicates that consumers are willing to pay premiums for "local" milk and milk from "family farms" (Wolf, Tonsor, and Olynk 2011). Second, it is possible that consumers prefer milk that is closer in geographical distance because it implies a shorter transportation distance, and thus a smaller environmental impact.

CHAPTER IV

CONCLUSION

Although surveys are common practice for measuring consumer preferences, it is argued that the values elicited from these methods lack the soundness needed to draw concrete, meaningful conclusions, and therefore lack the robustness needed to make appropriate marketing and policy recommendations. Previous research suggests that the biases that often arise in consumer preference studies may be alleviated by systematically guiding consumers to their preferences through forced internal consistency. In this approach, people review and revise their choices while observing the impact of their choices on their stated WTP. Although the theoretical framework for such preference elicitation methods has been established, a debate remains over whether the values elicited from this type of method are empirically superior in terms of rationality and predictive validity. This study sought to determine whether an elicitation method that forces internal consistency provides a more rational estimation of WTP and a better prediction of consumer choice for fluid milk products.

Our results suggest that the imposition of internal consistency yields more rational estimates of WTP, however there is no evidence to support the hypothesis that this method provides WTP estimates that lead to a higher degree of predictive validity. It is possible that by forcing internal consistency, we caused participants to deliberate over their choices, which has been shown to reduce preference consistency (Nordgren and Dijksterhuis 2009).

As it relates to WTP for milk product attributes, one interesting result is that nearly half of participants indicated that cloning had no importance when making a milk purchase decision. This is contrary to what we would expect, as previous research has indicated that consumers have a strong aversion to foods produced from cloned livestock. One possible psychological explanation for our results is that participants did not truly believe that milk from cloned cows is a realistic product, and they assumed that they would never actually encounter such a product in their grocery store.

The design of this experiment lends well to opportunities for future research. First, we had to select alternative preference elicitation methods for which to use as comparisons to the Preference Learning + Feedback method (Treatment 1). We chose to this method to preference learning *without* feedback in order to observe the marginal effects of forced consistency. Future research might compare the Preference Learning + Feedback method to an elicitation tool that does not question participants on individual attributes, such as a traditional conjoint approach where participants choose between competing product profiles. Previous studies have compared self-explicated approaches with traditional conjoint approaches but have produced mixed results (e.g. Green, Krieger, and Agarwal 1993; Leigh, MacKay and Summers 1984; Green and Helsen 1989). Our approach, which couples the self-explicated design with a feedback mechanism, has yet to be compared with other traditional stated preference methods.

Second, we had to select a measure to compare the alternative methods. One of the most notable challenges in designing this experiment was determining a way to quantitatively compare the results from different preference elicitation methods. We used an anchor test to compare the rationality of the WTP estimates across treatments and a holdout choice to determine the predictive validity of each treatment. In addition to employing variations of these techniques, there are other empirical tests, such as a test for reliability, which we could employ to establish the empirical significance of a Preference Learning + Feedback method. Future research might use alternative tests to empirically compare the results between methods.

Finally, we had to select a multi-attribute product to investigate consumer WTP. We chose fluid milk in this experiment because it has a limited number of attributes as well as the cloning attribute which was unfamiliar to participants. Future research might replicate our experiment with a different multi-attribute product to determine whether similar results are obtained.

Our ultimate goal is to identify the tools that provide decision makers with accurate information about people's preferences, but it is unlikely that we will ever be able to conclusively determine a superior preference elicitation method in absolute terms. Rather, we can work to determine the circumstances under which different methods exhibit superiority, and subsequently select the most appropriate preference elicitation method on a case-by-case basis.

REFERENCES

Ariely, D., G. Loewenstein, and D. Prelec. 2003. "'Coherent Arbitrariness': Stable Demand Curves without Stable Preferences." *The Quarterly Journal of Economics* 118(1):73-105.

Bateman, I.J., D. Burgess, W.G. Hutchinson, and D.I. Matthews. 2008. "Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness." *Journal of Environmental Economics Management* 55:127–141.

Bernard, J.C. and D.J. Bernard. 2009. "What Is It About Organic Milk? An Experimental Analysis." *American Journal of Agricultural Economics* 91(3):826.836.

Bernard, D.J. and A. Mathios. 2005. "Factors Affecting Consumer Choice and Willingness to Pay for Milk Attributes." Paper presented at AAEA annual meeting, Providence RI, 24-27 July.

Bernheim, B.D. and A. Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124:51-104.

Braga, J. and C. Starmer. 2005. "Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis." *Environmental and Resource Economics* 32:55-89.

Brooks, K. and J.L. Lusk. 2010. "Stated and Revealed Preferences for Organic and Cloned Milk: Combining Choice Experiment and Scanner Data." *American Journal of Agricultural Economics* 92(4):1229-1241.

Cherry, T.L., T.D. Crocker, and J.F. Shogren. 2003. "Rationality Spillovers." *Journal of Environmental Economics and Management* 45(1):63-84.

Cherry, T.L. and J.F. Shogren. 2007. "Rationality Crossovers." *Journal of Economic Psychology* 28(2):261-277.

Chu, Y,P. and R.L. Chu. 1990." The Subsidence of Preference Reversals in Simplified and Marketlike Experimental Settings: A Note." *American Economic Review* 80(4):902-911.

Diamond, P.A. and J.A. Hausman. 1994. "Contingent Valuation: Is Some Number better than No Number?" *Journal of Economic Perspectives* 8(4):45-64.

Green, P.E. and K. Helsen. 1989. "Cross-Validation Assessment of Alternatives to Individual-Level Conjoint Analysis: A Case Study." *Journal of Marketing Research* 26(3):346-350.

Green, P.E., A.M. Krieger, and M.K. Agarwal. 1993. "A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences." *Marketing Letters* 4(4): 369.380.

Green, P.E. and V. Srinivasan. 1990. "Conjoint Analysis in Marketing: New Developments With Implications for Research and Practice." *The Journal of Marketing* 54(4): 3-19.

Hoepfl R.T. and Huber G.P. 1970. "A study of self-explicated utility models." *Behavioral Science* 15(5):408-414.

Huffman, W.E., J.F. Shogren, M. Rousu, and A. Tegene. 2003. "Consumer Willingness to Pay for Genetically Modified Food Labels in a Market with Diverse Information: Evidence from Experimental Auctions." *Journal of Agricultural and Resource Economics* 28(3):481-502.

Kahneman, D. and R.H. Thaler. 2006. "Anomalies: Utility Maximization and Experienced Utility." *Journal of Economic Perspectives* 20(1):221–234.

Leigh, T.W., D.B. MacKay, and J.O. Summers. 1984. "Reliability and Validity of Conjoint analysis and Self-Explicated Weights: A Comparison." *Journal of Marketing Research* 21(4):456-462.

List, J.A. 2003. "Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics* 118:41-71.

List, J.A. and C.A. Gallet. 2001. "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics* 20(3): 241-254.

Loureiro, M.L. and W.J. Umberger. 2003. "Estimating Consumer Willingness to Pay for Country-of-Origin Labeling." *Journal of Agricultural and Resource Economics* 28(2):287-301.

Neill, H.R., R.G. Cummings, P.T. Ganderton, G.W. Harrison, and T. McGuckin. "Hypothetical Surveys and Real Economic Commitments." *Land Economics* 70(2): 145-154.

Nordgren, L.F. and A. Dijksterhuis. 2009. "The Devil Is in the Deliberation: Thinking Too Much Reduces Preference Consistency." *Journal of Consumer Research* 36(1):39-46.

Norwood, B.F. and J.L. Lusk. 2011. "A calibrated auction-conjoint valuation method: Valuing pork and eggs produced under differing animal welfare conditions." *Journal of Environmental Economics Management* 62:80–94.

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

Plott, C.R. 1996. "Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis." in K. Arrow, E. Colombatto, M. Perlman and C. Schmidt, Eds., *The Rational Foundations of Economic Behavior* London: Macmillan and St. Martin Press, p. 225-250.

Plott, C.R. and K. Zeiler. 2005. "The Willingness to Pay - Willingness to Accept Gap, The 'Endowment Effect,' Subject Misconceptions and Experimental Procedures for Eliciting Valuations." *American Economic Review* 95(3):530-545.

Srinivasan, V. 1988. "A Conjunctive-Compensatory Approach to the Self-Explication of Multiattributed Preferences." *Decision Sciences* 19(2):295-305.

Srinivasan, V. and C.S. Park. 1997. "Surprising Robustness of the Self-Explicated

Approach to Customer Preference Measurement." *Journal of Marketing Research* 34(2):286–291.

Sugden, R. 2004. "The Opportunity Criterion: Consumer Sovereignty without the Assumption of Coherent Preferences." *American Economic Review* 94(4):1014-1033.

Tsiros, M. and C.M. Heilman. 2005. "The Effect of Expiration Dates and Perceived Risk on Purchasing Behavior in Grocery Store Perishable Categories." *The Journal of Marketing* 69(2):114-129.

Tversky, A. and D. Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics* 106(4):1039-1061.

U.S. Food and Drug Administration. 2008. FDA Issues Documents on the Safety of Food from Animal Clones. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2008/ucm116836.htm (accessed February 10, 2012).

USDA Economic Research Service. 2010. Food Marketing System in the U.S.: New Product Introductions. http://www.ers.usda.gov/Briefing/FoodMarketingSystem/new_product.htm (accessed March 13, 2012).

USDA Economic Research Service. 2011. "Fluid milk sales by product." http://www.ers.usda.gov/Publications/ldp/LDPTables.htm (accessed February 21, 2012).

U.S. Food and Drug Administration. 2009. Animal Cloning Consumer FAQs. http://www.fda.gov/AnimalVeterinary/SafetyHealth/AnimalCloning/ucm055516.htm (accessed March 13, 2012).

Wal-Mart Stores, Inc. 2008. "Wal-Mart Offers Private Label Milk Produced without Artificial Growth Hormone," http://walmartstores.com/pressroom/news/8147.aspx (accessed February 10, 2012).

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

TABLES AND FIGURES

Table 1. Characteristics of Survey Participants

		Mean		
		Treatment 1[a]	**Treatment 2**[b]	**Treatment 3**[c]
Variable	Definition	Preference Learning + Feedback	Preference Learning	Control
Gender	1 if female; 0 if male	1.000	1.000	1.000
College	1 if obtained bachelor's degree or higher; 0 otherwise	0.439	0.427	0.440
Children	1 if has children under the age of 12; 0 otherwise	0.516	0.562	0.514
Race	1 if white ethnicity; 0 otherwise	0.780	0.756	0.758
Income1	1 if household income is < $40,000; 0 otherwise	0.323	0.274	0.303
Income2	1 if household income is $40,000 to $79,999; 0 otherwise	0.406	0.469	0.395
Income3	1 if household income is $80,000 or greater	0.271	0.257	0.303
Kansas	1 if resides in Kansas; 0 otherwise	0.083	0.053	0.114
Missouri	1 if resides in Missouri; 0 otherwise	0.214	0.202	0.171
Oklahoma	1 if resides in Oklahoma; 0 otherwise	0.114	0.101	0.094
Texas	1 if resides in Texas; 0 otherwise	0.590	0.643	0.621
Milk: Never	1 if never consumes milk; 0 otherwise	0.000	0.000	0.000
Milk: 1-2	1 if consumes milk 1-2 times per week; 0 otherwise	0.269	0.285	0.312
Milk: 3-4	1 if consumes milk 3-4 times per week; 0 otherwise	0.215	0.262	0.218
Milk: 5-6	1 if consumes milk 5-6 times per week; 0 otherwise	0.229	0.211	0.201
Milk: > 6	1 if consumes milk more than 6 times per week; 0 otherwise	0.287	0.242	0.269
Age1	1 if age is 30 to 34 years; 0 otherwise	0.219	0.278	0.274
Age2	1 if age is 35 to39 years, 0 otherwise	0.268	0.205	0.239
Age3	1 if age is 40 to 45 years, 0 otherwise	0.259	0.287	0.202
Age4	1 if age is 46 to 51 years, 0 otherwise	0.254	0.230	0.285

[a] Number of observations is 351

[b] Number of observations is 356

[c] Number of observations is 351

Table 2. Attributes and Attribute Levels Used in Surveys

Attribute	Levels
Price per gallon	\$2.99, \$3.99, \$, 4.99, \$5.99, \$6.99
Fat Content	Skim (0%), Low Fat (1%), Reduced Fat (2%), Whole (3.25%)
Brand	Great Value, Prairie Farms, Markey Pantry, Borden, Best Choice, Hiland
Organic	Not Organic, Organic
Farm Location	25 miles from store, 50 miles from store, 100 miles from store, 500 miles from store, > 500 miles from store
Expiration Date	Expires within 3 days, Expires within 5 days, Expires within 7 days, Expires within 14 days
Hormone Use	Produced without artificial growth hormones, Produced with artificial growth hormones
Cloning	Not produced from cloned cows or offspring of cloned cows, Produced from offspring of cloned cows, Produced from cloned cows

Table 3. Product Profiles Used in Surveys

Milk	Price	Fat Content	Brand	Organic	Expiration Date	Farm Location	Cloning
A			Borden		In 14 days	50 miles from store	Not produced from cloned cows or offspring of cloned cows
B			Great Value		In 5 days	> 500 miles from store	Produced from cloned cows
C	\$3.99	Skim		Not organic			Not produced from cloned cows or offspring of cloned cows
D	\$5.99	Low Fat		Organic			Not produced from cloned cows or offspring of cloned cows
E	\$2.99	Reduced Fat		Not organic			Produced from offspring of cloned cows
F	\$4.99	Whole		Organic			Produced from cloned cows

Table 4. Mean Importance Weights

	Treatment 1 Preference Learning + Feedback					Treatment 2 Preference Learning	
	All Responses	Percent Zero[a]	Unrevised	Revised	$W_{iP} > 0$	All Responses	Percent Zero
Price per gallon	35	13%	28	44	40	28	10%
Fat Content	14	23%	16	11	13	14	24%
Brand	9	36%	11	6	7	10	31%
Organic	9	38%	9	7	8	9	38%
Farm Location	6	39%	6	6	5	6	39%
Expiration Date	14	16%	16	13	13	18	17%
Hormone Use	8	37%	9	8	8	10	39%
Cloning	5	46%	5	5	5	6	47%

[a] Indicates the percent of participants who allocated zero points to the specified attribute

Table 5. Mean Importance Weights *t*-test[a] *p*-values

	TR 1 / TR 1 Unrevised / Revised	TR 1 / TR 2 All Responses / All Responses	TR 1 / TR 2 Revised / All Responses
Price per gallon	0.0000**[c]	0.0000**	0.0000**
Fat Content	0.0049**	0.8975	0.0420*[b]
Brand	0.0038**	0.3823	0.0073**
Organic	0.2209	0.7210	0.2533
Farm Location	0.6937	0.4424	0.4107
Expiration Date	0.0843	0.0075**	0.0014**
Hormone Use	0.4994	0.1734	0.1409
Cloning	0.5711	0.2357	0.1875

[a] Two-tailed test of the hypothesis that the mean importance weights are the same between treatments and/or treatment subsections
[b] One asterisk (*) denotes statistical significance of 5%
[c] Two asterisks (**) denote statistical significance of 1%

Table 6. Anchor Test Correlations

	Treatment 1 Preference Learning + Feedback				Treatment 2 Preference Learning			Treatment 3 Control
	All Responses†[a]	Unrevised†	Revised†	$W_{iP} > 0$†	All Responses††[b]	All Responses†	$W_{iP} > 0$†	All Responses††
Pearson	0.0240 (0.6537)[c]	0.0270 (0.7092)	0.0203 (0.8001)	0.0175 (0.7596)	0.1020 (0.0548)	0.1019 (0.0547)	0.1093 (0.0501)	0.2623 (<.0001)
	[-0.0809, 0.1283][b]	[-0.1147, 0.1676]	[-0.1363, 0.1758]	[-0.0946, 0.1292]	[-0.0022, 0.2038]	[-0.0022, 0.2036]	[-0.0002, 0.2158]	[0.1617, 0.3568]
Spearman	0.0560 (0.2959)	0.1382 (0.0554)	-0.0462 (0.5642)	0.0191 (0.7393)	0.1850 (0.0005)	0.1025 (0.0532)	0.1317 (0.0181)	0.1857 (0.0005)
	[-0.0489, 0.1598]	[-0.0032, 0.2740]	[-0.2003, 0.1113]	[-0.0929, 0.1309]	[0.0822, 0.2833]	[-0.0016, 0.2042]	[0.0225, 0.2374]	[0.0823, 0.2846]

[a] † Denotes that correlations were estimated using WTP values that were calculated using preference learning data and equation 3

[b] †† Denotes that correlations were estimated using WTP values were manually entered by survey participants

[c] Numbers in parentheses ( ) are *p*-values

[d] Numbers in brackets [ ] are 95% confidence limits

Table 7. Holdout Choice Predictions

	Treatment 1 Preference Learning + Feedback				Treatment 2 Preference Learning			Treatment 3 Control
	All Responses†[a]	Unrevised†	Revised†	$W_{iP} > 0$†	All Responses††[b]	All Responses†	$W_{iP} > 0$†	All Responses††
Hit Rate, Top Choice[c]	47.01%	47.15%	46.84%	48.53%	58.43%	49.72%	50.93%	58.40%
Pearson Correlation	0.3847	0.3305	0.4509	0.4228	0.5046	0.4212	0.4446	0.4950
	[0.3265, 0.4428][d]	[0.2486, 0.4125]	[0.3701, 0.5316]	[0.3627, 0.4830]	[0.4532, 0.5550]	[0.3634, 0.4790]	[0.3851, 0.5041]	[0.4423, 0.5477]

[a] † Denotes that predictions were made using WTP values that were calculated using preference learning data and equation 3. For participants who entered "zero" as the importance weight for price ($W_{iP} = 0$), we calculated WTP using $W_{iP} = 0.01$.

[b] †† Denotes that predictions were made using WTP values that were manually entered by survey participants

[c] The hypothesis that the Hit Rates for All Responses are independent of Treatment is rejected at the $p = 0.01$ level of significance according to a Chi-Square test for independence

[d] Numbers in brackets [ ] are 95% confidence limits

Table 8. Top Choice from Holdout Profiles

	Treatment 1 Preference Learning + Feedback				Treatment 2 Preference Learning		Treatment 3 Control
	All Responses	Unrevised	Revised	$W_{iP} > 0$	All Responses	$W_{iP} > 0$	All Responses
Milk C	47%	50%	43%	47%	45%	46%	45%
Milk D	21%	20%	22%	21%	22%	21%	18%
Milk E	24%	24%	25%	25%	25%	26%	31%
Milk F	8%	7%	10%	8%	8%	8%	6%

Table 9. Average Product Rank from Holdout Profiles

	Treatment 1 Preference Learning + Feedback				Treatment 2 Preference Learning		Treatment 3 Control
	All Responses	Unrevised	Revised	$W_{iP} > 0$	All Responses	$W_{iP} > 0$	All Responses
Milk C	1.83	1.78	1.89	1.84	1.91	1.89	1.90
Milk D	2.42	2.39	2.46	2.43	2.45	2.49	2.56
Milk E	2.49	2.53	2.44	2.48	2.39	2.35	2.32
Milk F	3.25	3.30	3.18	3.24	3.25	3.26	3.22

Table 10. Average Product Rank *t*-test[a] *p*-values

	TR 1 / TR 1	TR 1 / TR 2	TR 1 / TR 2	TR 1 / TR 3	TR 1 / TR 3	TR 2 / TR 3
	Unrevised / Revised	All Responses	Revised / All Responses	All Responses	Revised / All Responses	All Responses
Milk C	0.2663	0.2822	0.8559	0.3605	0.9583	0.8695
Milk D	0.5520	0.7290	0.9302	0.0707*[b]	0.2937	0.1491
Milk E	0.4156	0.1888	0.6013	0.0281**[c]	0.2186	0.3326
Milk F	0.2202	0.9437	0.4241	0.7099	0.6318	0.6561

[a] Two-tailed test of the hypothesis that the mean product ranks are the same between treatments and/or treatment subsections

[b] One asterisk (*) denotes statistical significance of 5%

[c] Two asterisks (**) denote statistical significance of 1%

Table 11. Marginal WTP [a] Values (per gallon) for Select Milk Product Attributes – Treatment 1

	All Responses[b]		$W_{iP} > 0$	
Change	Mean	Median	Mean	Median
Brand				
Borden vs. Great Value	\$1.45	\$0.00	\$0.58	\$0.00
Organic				
Organic vs. Not Organic	\$2.86	\$0.00	\$2.00	\$0.00
Expiration Date				
Expires within 14 days vs. 3 days	\$3.18	\$1.22	\$2.86	\$1.41
Expires within 14 days vs. 5 days	\$2.86	\$0.81	\$2.40	\$0.94
Expires within 14 days vs.7 days	\$1.55	\$0.19	\$1.19	\$0.33
Farm Location				
25 miles vs. 50 miles from store	\$1.26	\$0.00	\$0.46	\$0.00
25 miles vs. 100 miles from store	\$1.60	\$0.00	\$0.76	\$0.00
25 miles vs. 500 miles from store	\$2.49	\$0.03	\$1.29	\$0.07
25 miles vs. > 500 miles from store	\$2.80	\$0.04	\$1.55	\$0.08
Hormone Use				
Without artificial growth hormones vs. With artificial growth hormones	\$3.98	\$0.15	\$2.99	\$0.16
Cloning				
Not from cloned cows or offspring of cloned cows vs. From offspring of cloned cows	\$2.66	\$0.00	\$1.77	\$0.00
Not from cloned cows or offspring of cloned cows vs. From cloned cows	\$2.53	\$0.00	\$1.82	\$0.00

[a] Note: WTP values are bounded between -\$30 and \$30.

[b] For participants who entered "zero" as the importance weight for price ($W_{iP} = 0$), we calculated WTP using $W_{iP} = 0.01$

Table 12. Marginal WTP [a] Values (per gallon) for Select Milk Product Attributes – Treatment 2

	All Responses[b]		$W_{iP} > 0$	
Change	Mean	Median	Mean	Median
Brand				
Borden vs. Great Value	$1.48	$0.00	$1.10	$0.00
Organic				
Organic vs. Not Organic	$2.78	$0.00	$1.93	$0.00
Expiration Date				
Expires within 14 days vs. 3 days	$5.37	$2.70	$5.14	$2.86
Expires within 14 days vs. 5 days	$4.28	$1.92	$4.01	$2.00
Expires within 14 days vs.7 days	$2.58	$0.82	$2.04	$0.83
Farm Location				
25 miles vs. 50 miles from store	$0.82	$0.00	$0.47	$0.00
25 miles vs. 100 miles from store	$1.50	$0.00	$1.00	$0.00
25 miles vs. 500 miles from store	$2.20	$0.07	$1.68	$0.15
25 miles vs. > 500 miles from store	$2.52	$0.15	$2.09	$0.25
Hormone Use				
Without artificial growth hormones vs. With artificial growth hormones	$4.90	$0.19	$4.18	$0.21
Cloning				
Not from cloned cows or offspring of cloned cows vs. From offspring of cloned cows	$2.79	$0.00	$2.50	$0.00
Not from cloned cows or offspring of cloned cows vs. From cloned cows	$2.80	$0.00	$2.54	$0.00

[a] Note: WTP values are bounded between -$30 and $30.

[b] For participants who entered “zero” as the importance weight for price ($W_{iP} = 0$), we calculated WTP using $W_{iP} = 0.01$

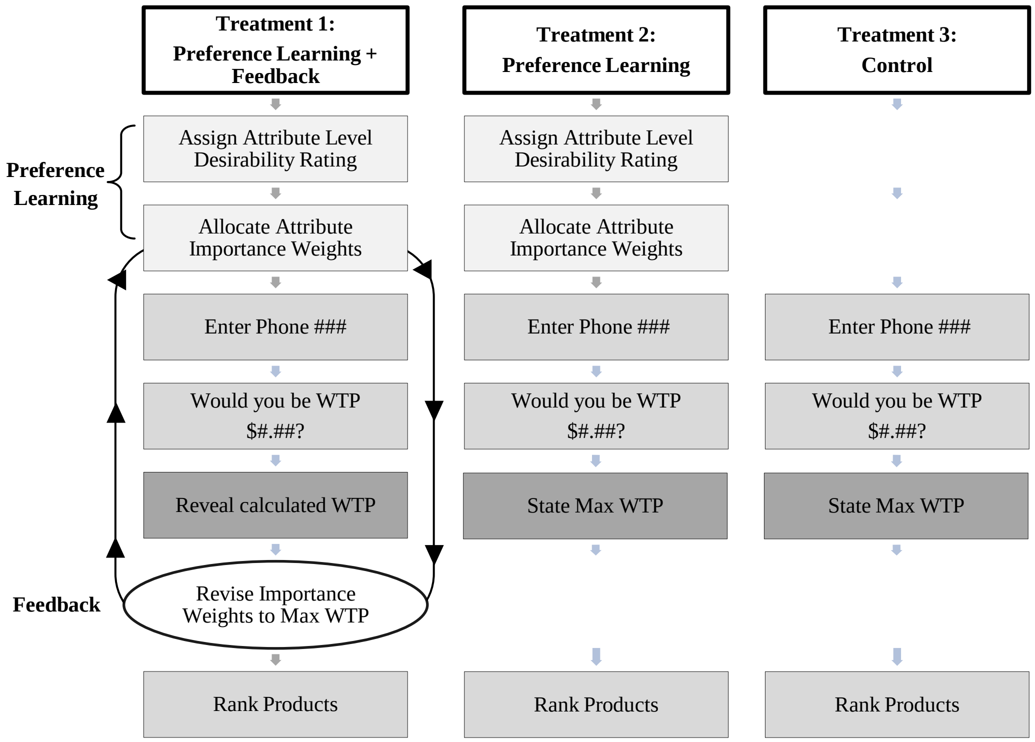
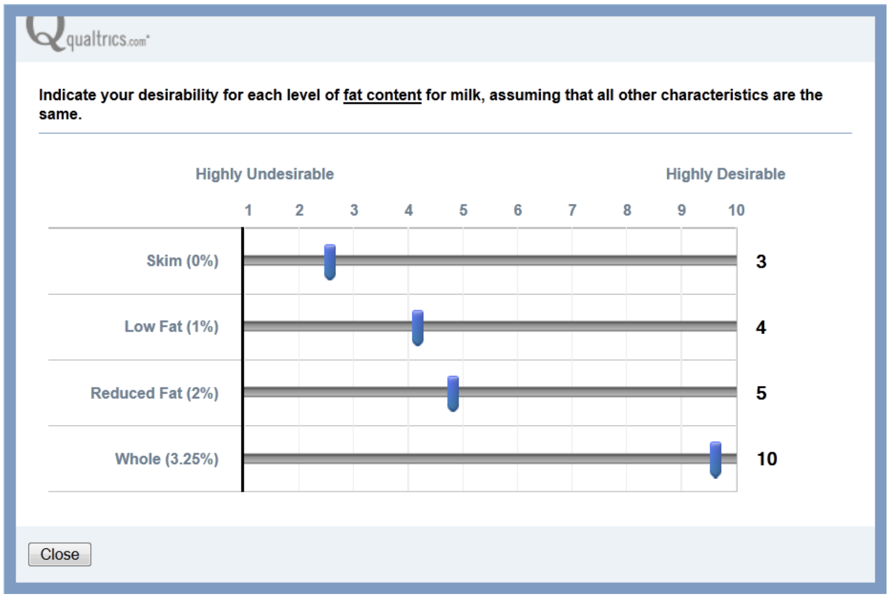


Figure 1. Order of Steps in Survey Treatments

a) Step 1: Rate the desirability of attribute levels



b) Step 2: Indicate the relative importance of each attribute

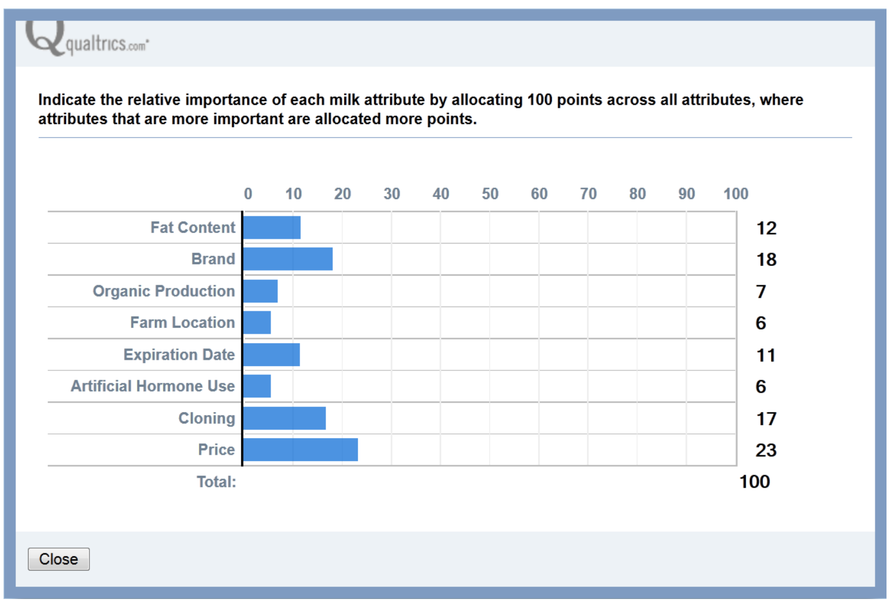
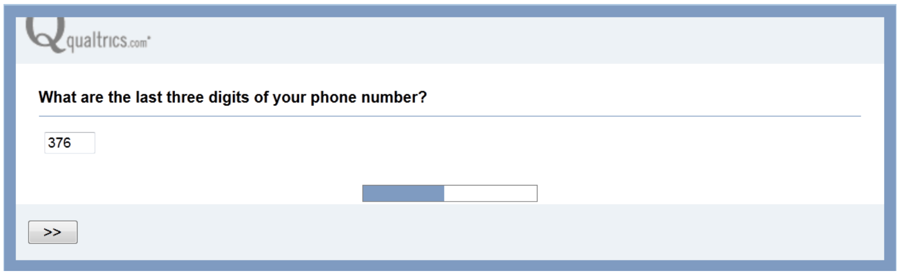


Figure 2. Steps in Preference Learning (Treatments 1 and 2)

a) Step 1: Enter last three digits of phone number (###)

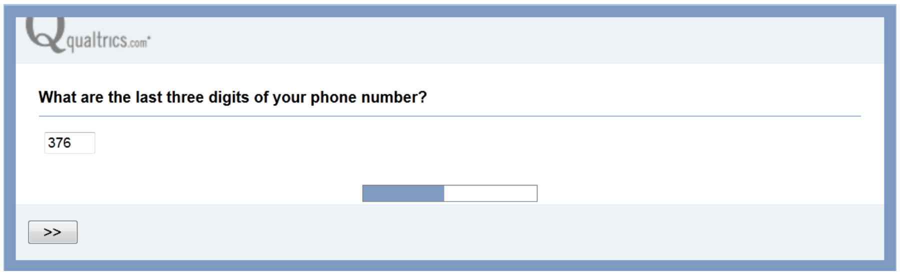


b) Step 2: Would you be willing to pay $#.##…?

c) Step 3: Indicate whether calculated WTP is correct

Figure 3. Steps in Anchor Test (Treatment 1)

a) **Step 1:** Enter last three digits of phone number (###)



b) **Step 2:** Would you be willing to pay $#.##…?

c) **Step 3**: What is the maximum that you would be WTP…?

Figure 4. Steps in Anchor Test (Treatment 2 and Treatment 3)

a) Step 1 (Treatment 2, 3): How much would you be willing to pay…?

b) Step 2 (Treatments 1, 2, 3): Rank products from 1-4…

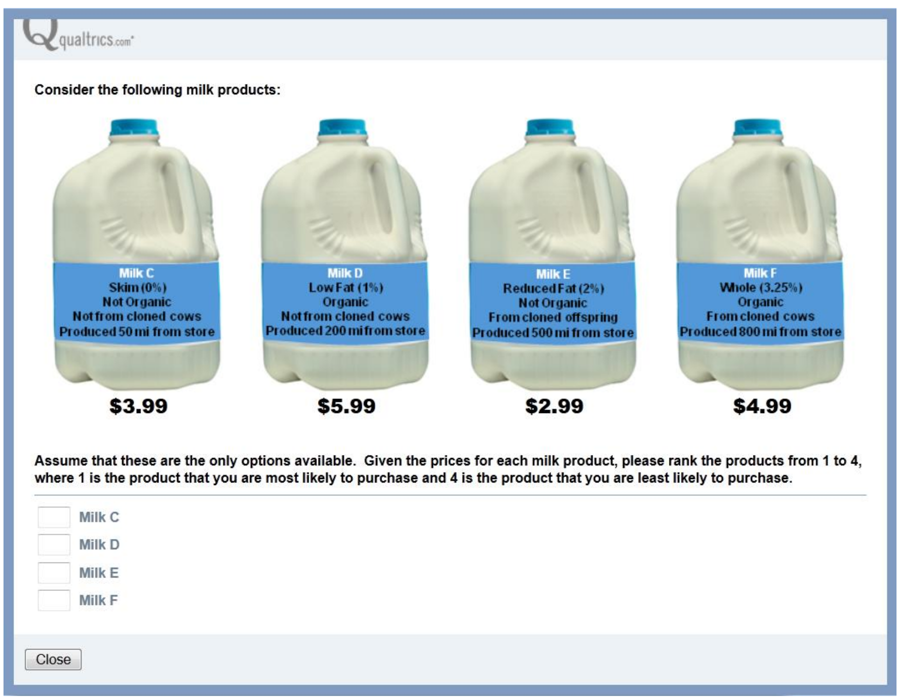


Figure 5. Steps in Holdout Choice

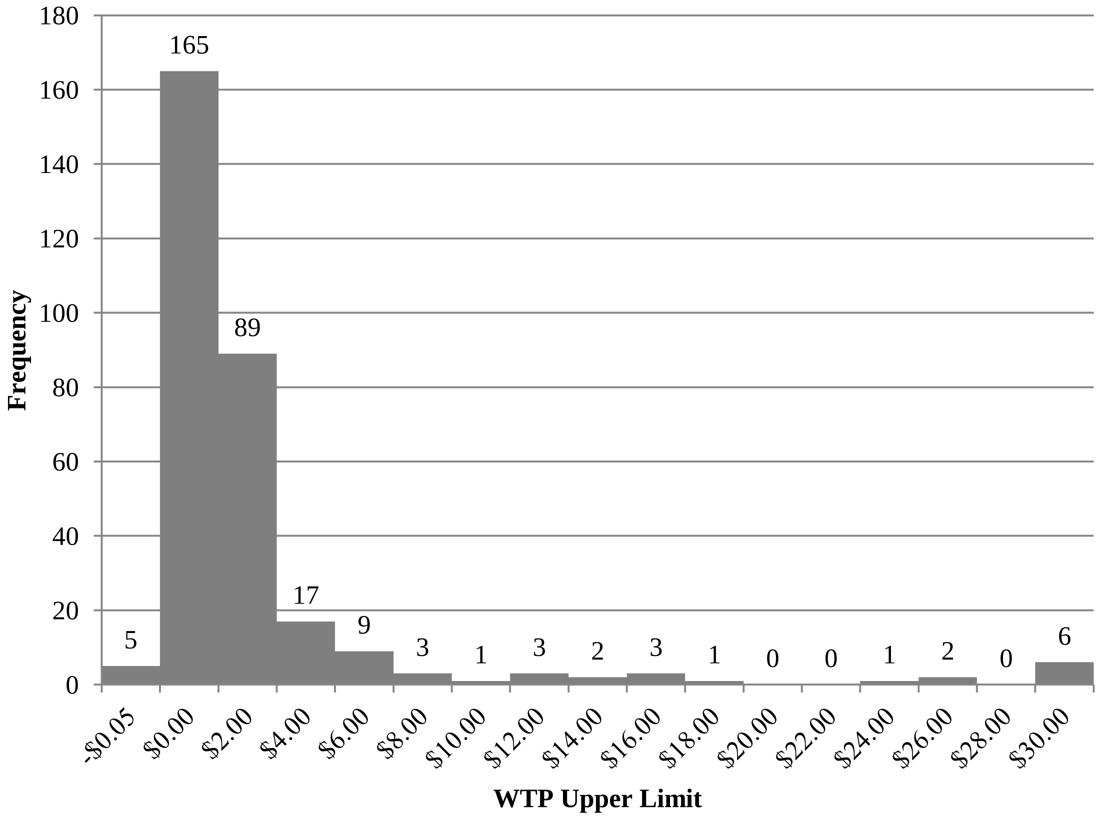


Figure 6. Distribution of Marginal WTP to Avoid Milk from Cloned Cows – Treatment 1

APPENDICES

Appendix A – Oklahoma State University Institutional Review Board Approval

Oklahoma State University Institutional Review Board

Date:	Wednesday, June 15, 2011
IRB Application No	AG1134
Proposal Title:	Consumer Willingness to Pay for Milk Products
Reviewed and Processed as:	Exempt

Status Recommended by Reviewer(s): Approved Protocol Expires: 6/14/2012

Principal Investigator(s):

Kristen Kovalsky	Jayson Lusk
421D Ag Hall	411 Ag Hall
Stillwater, OK 74078	Stillwater, OK 74078

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

☒ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 219 Cordell North (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,

Shelia M. Kennison

Shelia Kennison, Chair
Institutional Review Board

VITA

Kristen Lynne Kovalsky

Candidate for the Degree of

Master of Science

Thesis: DO CONSUMERS REALLY KNOW HOW MUCH THEY ARE WILLING TO PAY?

Major Field: Agricultural Economics

Biographical:

Education:
Completed the requirements for the Master of Science in Agricultural Economics at Oklahoma State University, Stillwater, Oklahoma in May, 2012.

Completed the requirements for the Bachelor of Science in Food & Resource Economics at University of Florida, Gainesville, Florida in May, 2010.

Professional Memberships:
Graduate and Professional Student Government Association, August 2010 to May 2012; Agricultural Economics Graduate Student Association, August 2010 to May 2012; Toastmasters International, November 2010 to Present.

Name: Kristen Kovalsky Date of Degree: May, 2012

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: DO CONSUMERS REALLY KNOW HOW MUCH THEY ARE WILLING TO PAY?

Pages in Study: 48 Candidate for the Degree of Master of Science

Major Field: Agricultural Economics

Scope and Method of Study:

Economists often assume that consumers know exactly what they want and how much they are willing to pay for it, but this assumption is perhaps unjustified. Because people lack stable preferences, estimates of willingness-to-pay (WTP) obtained through surveys and experimental settings are often bias, irrational, and sometimes outright arbitrary. This paper examines data collected from a split-sample online survey of 1,058 participants to determine whether consumer preferences can be elicited in a systematic way that provides a rational estimation of WTP and a better prediction of consumer choice. Specifically, we facilitate the preference-discovery process by providing survey participants with the opportunity to observe the tradeoffs of their choices and the subsequent consequences on their stated WTP for fluid milk products. By using this approach, we force internal consistency between statements of value and preference orderings. We test the rationality of the results with an anchor test and we test predictive validity with a holdout choice.

Findings and Conclusions:

Our results indicate that the imposition of internal consistency yields more rational estimates of WTP, however there is no evidence to support the hypothesis that this method provides WTP estimates that lead to a higher degree of predictive validity.

ADVISER'S APPROVAL: Dr. Jayson Lusk