It's All Relative: Consistent Marginal Effects with Willingness to Pay and Willingness to Accept Framing in Experimental Auctions

Abstract

When eliciting consumer preferences for controversial products—an increasing number of which exist due to increasing demographic diversity and political polarization—conventional assumptions that all individuals derive positive marginal utility from consumption are challenged. It is relatively easy to adjust hypothetical stated preference questionnaires to include negative willingness to pay (WTP), but few studies on controversial products investigate how individuals behave using incentive-compatible revealed preference techniques. Using a framed field experiment with 292 adult subjects, we fill this gap by comparing the differences and similarities between a set of results that arise from the Becker-DeGroot-Marschak (BDM) mechanism between WTP versus willingness to accept (WTA) elicitation methods. This study has two main findings. First, in economic experiments eliciting preferences for controversial products, neither the WTP nor the WTA method fully discovers the true valuation range across all participants. Second, despite framing effects that give rise to different bid distributions, relative revealed preferences for the examined products are consistent under various interventions, indicating that WTP and WTA estimates have consistent policy implications.

Keywords: Controversial products, Framing, Product Labeling, Willingness to Pay, Negative Willingness to Pay

JEL Codes: D12, C91, Q18

Introduction

While experimental economics studies on consumer products usually elicit consumer preferences using willingness to pay (WTP) measures (Corrigan and Rousu 2008; Liaukonyte et al. 2012; Costanigro and Lusk 2014), one may argue that WTP is a poor measure when considering issues related to controversial products such as food with perceived safety concerns or cultural conflicts; thus, willingness to accept (WTA) (sometimes referred to as negative WTP) may be more suitable (Bass, McFadden and Messer 2021). Hypothetical stated preference studies have noted that WTP and WTA are good substitutes for each other in some contexts (Johnston et al., 2017), and it is relatively easy to expand a hypothetical valuation range when incentive compatibility is not implemented. However, little methodological attention in experimental economics has been paid to controversial products (e.g., single-use plastics, red meat, and near-expiry products), the number of which has been rapidly growing in the 21st century due to technological innovation and political polarization (McCluskey et al., 2015). Given the rise of products that are appealing to some consumers but repelling to others, what is the implication on elicitation methods? This study examines this question by comparing findings that arise from the WTP versus WTA Becker-DeGroot-Marschak (BDM) mechanisms on a set of near-expiry and expired yogurt smoothie drinks.

For decades, economic studies have focused on inconsistencies between WTP and WTA with elicited WTP values generally being smaller than elicited WTA values (e.g., Coursey et al., 1987; Horowitz and McConnell, 2002; Knetsch and Sinden, 1984). Some studies have attributed it to endowment effects and to consumers viewing selling an item as a loss (Fehr, Hakimov and Kübler 2015; Corrigan and Rousu 2008; Kahneman, Knetsch and Thaler 1991). Other studies have

pointed to framing effects and elements of experiment designs (Plott and Zeiler 2005; Coursey et al. 1987).

To date, WTP and WTA studies in the literature commonly have bids censored at $0.00 in the lower limit under the free disposal assumption, while the comparisons between the two methods are based on the assumption that preferences for a product is uni-directional for all individuals, such that participants bid on WTP for acquiring a product and bid on WTA for foregoing the same product (Kahneman et al. 1991; Plott and Zeiler 2005; Drichoutis, Lusk and Pappa 2016). The plausibility of both assumptions is challenged when participants make choices on controversial products. To this end, if some consumers are willing to pay a positive amount for ingesting a product while others require a positive compensation for the same activity, both the conventional WTP and WTA methods are expected to cover only a certain range of the true bids.

The purpose of this study is to further investigate the differences and similarities between WTP and WTA in the context of a perishable food product with varying age and, hence, potential food safety concerns that may be viewed heterogeneously by different individuals. We focus our analysis on whether this gap exists for marginal as well as average effects with respect to how age and food labels change participants' WTP and WTA bids. Our main finding is that while the anticipated gap exists between average WTP and WTA bids, the two measures are consistent—in terms of sign and statistical significance—from a marginal effects perspective. This result is important because it indicates that WTP and WTA are good substitutes for each other in some contexts. Moreover, this result is significant because consistency in marginal effects estimated from WTP and WTA measures is often more relevant than consistency in point estimates meant to gauge consumers' true valuations in the contexts of marketing strategies and policy making.

A related goal of this study is to advocate for the inclusion of WTA measures in addition to WTP measures in consumer preference studies, which rarely elicit WTA even when some consumers may have negative WTP for the examined products. WTA is particularly valuable in cases where WTP values are censored. Many WTP studies have found significant censoring of bids at zero when some of the subjects have negative consumer reactions to the food/beverage product being auctioned. For example, one would expect to see a significant number of zero WTP bids for a food experiment featuring food safety issues (e.g., such as bidding on hamburgers after receiving information about the impacts of mad cow disease; Messer et al. 2011). In such studies, WTA would be a better measure to zero bidding subjects since their true valuation may be negative due to the potential negative health impacts rather than zero WTP. To fully investigate this concern, our context of a yogurt-based drink with a range of days since production allows us to measure both WTP and WTA in a single experiment, and including expired products allows us to consider consumers who potentially have negative WTP.

Our analysis of WTP versus WTA is based on a framed field experiment involving purchasing and accepting a perishable product that has a potential food safety concern due to the age of the products. WTP is elicited based on the conventional BDM mechanism, while WTA is elicited based on a reverse BDM mechanism. We estimate and compare the WTP and WTA bid distributions for a yogurt-based drink of different ages accompanied by different expiration date labels, as well as the signs and statistical significance of marginal effects with respect to how age and food labels change participants' bids.

2. Literature Review

The tendency of auction participants to have smaller values of WTP (the maximum amount a participant is willing to pay for an item) than WTA (the minimum amount a participant will accept to compensation for giving up the same item) has been documented in a variety of contexts, but the cause of the gap has been the subject of debate in the literature (e.g., Boyce et al., 1992). Typically, the gap is greatest in hypothetical surveys (Knetsch 1989) and when auctioning difficult-to-value or unfamiliar items such as public goods (Heberlein and Bishop 1986). For example, Brookshire and Coursey (1987) found that the gap in valuations of trees in a park, while present, was less than the gap in valuations of hypothetical trees. However, the gap has also been measured in contexts that should rule out potential explanations for the inconsistency. The presence of both loss aversion (an endowment effect) and reference-dependent preferences (anchoring or framing) has been confirmed using experimental protocols (Bateman et al. 1997; Tversky and Kahneman 1974).

Traditionally, the WTP-WTA gap has been justified more by loss aversion than by framing effects (Ericson and Fuster 2014). When responses to a BDM mechanism and multiple-price list were compared, the gap was greater under the multiple-price list, perhaps indicating that the gap is not due to participant misconceptions (Brebner and Sonnemans 2018). In an experiment involving a verbal protocol (subjects describing their thought processes), researchers found that the WTP-WTA gap was due primarily to subject beliefs that selling an item for less than the market price represented a type of loss (Brown 2005).

On the other hand, Plott and Zeiler (2007; 2005) rejected the influence of the endowment effect as driving WTP-WTA inconsistency, finding instead that the gap was caused by subject misperceptions related to framing. When subjects clearly understood what they were being asked

to decide, the WTP-WTA gap disappeared (these results have been replicated by Kovalchik et al. (2005) and others). These findings are consistent with the observation that the gap is greatest when the item under consideration is unfamiliar or is difficult to value. Further confirmation comes from List (2011; 2004; 2003), who demonstrated that inexperienced traders were more reluctant to trade than experienced traders. In addition, studies have shown that repeat rounds attenuate the gap associated with one-shot experiments that do not allow for learning (e.g., Coursey et al., 1987). This explanation is also verified by the evidence of large gaps in risky lotteries (Isoni, Loomes and Sugden 2011). On the other hand, no gap in WTP-WTA was found in a study of social influences and voting for public goods that used induced values (Messer et al. 2010). Consequently, the literature suggests that experimental design can influence the presence or absence of WTP-WTA gaps, but no firm conclusion has been drawn about the source of these gaps.

In addition to experimental studies, several meta-analyses have addressed this gap issue. Tunçel and Hammitt (2014) found less inconsistency for private goods than for public goods. An earlier meta-analysis found that the gaps in experiments eliciting WTP and WTA were greater for lotteries than for ordinary private goods such as mugs and pens (Horowitz and McConnell 2002).

To further this literature, we use a unique experimental design that allows us to gather both WTP and WTA from subjects (between-subjects treatments) and valuations for multiple production dates (within-subjects treatments). This design does not fall prey to endowment effects because participants in the between-subjects design face the same choice of bidding on how much they are willing to pay or how much do they need to be compensated in order to ingest the same set of drink; participants in neither treatment were endowed with the product. Framing effects are expected to persist because WTP and WTA methods can change participants' reference points. Nevertheless, the main contribution of this study is to compare the signs and statistical significance

of marginal effects with respect to how age and food labels change participants' bids under the two different mechanisms.

Our experimental design is unique compared to the existing literature and specifically offers guidance regarding studies on consumer preferences and behavior for controversial products that some individuals may find appealing but others repelling. Although previous research has been creative when designing experiments to measure both WTP and WTA for same products or tasks, the focus has been different. For example, Coursey et al. (1987) conducted one of the first experiment-based and innovative studies to measure both using a bitter tasting, but safe-to-ingest substance (sucrose octa-acetate) as the auctioned good. Similar comparison was used in studies about climate neutral foods (Drichoutis et al. 2016) and drinking water that contained a sanitized cockroach (Kecinski et al. 2018). In these studies, WTP to avoid a task was compared to WTA to perform it, while our design focuses on the controversy aspect of a product and enables us to elicit preferences when some individuals are willing to pay for ingesting it, but others request compensation for the same task. Valuation studies regarding food[1] have almost always used WTP measures (Corrigan and Rousu, 2008; D'amato et al., 2020; Dillaway et al., 2011; Li et al., 2018; Liaukonyte et al., 2012; Lusk et al., 2005). Only a handful have used WTA only (Costa-Font, Gil and Traill 2008; Heiman and Zilberman 2011; Li et al. 2020). In environmental contexts, researchers' preference for WTP studies is evident as 14 WTP studies were found for every one WTA study in the Environmental Valuation Reference Inventory (Lloyd-Smith and Adamowicz 2018). WTP is a standard measure of consumable products, but in contexts with potential for negative valuations WTA is a particularly valuable tool.

[1] For instance, see the meta-analysis of studies of valuations of genetically modified food in which 13 used only WTP, 10 used WTP and WTA, and 4 used only WTA (Lusk et al. 2005).

3. Experiment Design

We conducted a framed field experiment with 292 adult subjects who participated in either a conventional BDM eliciting WTP or a reverse BDM eliciting WTA (Becker, DeGroot and Marschak 1964) for yogurt-based drink labeled with various expiration date labels. The experiment was conducted on April 29, 2017, at an annual educational event held in the Mid-Atlantic United States. The experiment design was registered with the American Economics Association's Randomized Controlled Trials Registry (AEARCTR- #######, blocked for anonymity).

Participants were compensated $10 to participate in the experiment that took approximately 15 minutes. They were asked to either place WTP bids to consume the products or WTA bids to consume the products (depending on which treatment they were in)—in five rounds that varied by the drinks' production dates presented in random order. The allowable bid ranges from $0.00 to $10.00 in all treatments. Through written and oral instructions (see Appendix 1), participants were told that if their WTP/WTA bid was accepted, the participant needed to immediately consume the drink on-site. This feature was designed to prevent people from simply disposing of the item, which could have influenced the WTA bids. Therefore, participants in the WTP and WTA treatments were paying or being paid, respectively, to immediately consume the drink.

For each participant, once the five bidding rounds were completed, the experiment software (Python-Willow) randomly selected one round, $j \in [1,5]$, as binding and generated a random price for the product, p, from a uniform distribution between $0.00 and $10.00 ($p \sim U(0, 10)$). In the WTP treatment, participants who bid at least the randomly determined price consumed the drink and paid for the price. In the WTA treatment, participants who submitted an offer that was equal to or lower than the randomly determined compensation consumed the drink and received the randomly drawn compensation. Those who bid less than the random price in WTP treatment or

higher than the randomly drawn compensation in the WTA treatment received the full initial $10 payment for participation in the experiment and did not consume the drink.

The experiment incorporated the following two between-subjects treatments and one within-subjects treatment (table 1):

1. Between-subjects treatments (orthogonal to each other)
    a. WTP or WTA measure for consuming the drink.
    b. Wording of expiration date label: “BEST IF USED BY” or “PRODUCED ON.”
2. Within-subjects treatment presented in random order
    a. Otherwise-identical products (same label wording) that varied in age: seven days until expiration, one day until expiration, on the date of expiration, one day past expiration, and four days past expiration.

Following standard experimental economics protocol, no deception was used in the experiment. The research team has secured otherwise identical products with five different production dates. Two out of five tested products had passed their expiration dates on the day of this experiment, while various sources confirmed that they were safe to ingest. This was intentionally designed because individuals who are concerned about the expiration dates may have negative WTPs (positive WTAs) but individuals who are conscious about food waste and the scientific facts of expiration dates are expected to have positive WTPs. The expiration dates on the bottles were imprinted by the manufacturer in the format of MMM-DD-YYYY. Expiration date labels truthfully representing these dates were shown on the computer interface. The experiment was purposefully conducted on a single day so the number of days since production was the same for all participants.

Participants were given a short survey after they completed the five BDM rounds to collect demographic and attitudinal information. The experiment protocol provides unique insight into whether or not the WTP and WTA mechanisms result in comparable marginal as well as average effect of product characteristics—products' age (days since production) and accompanying expiration date labels—on participants' valuation.

4. Data and Descriptive Statistics

Descriptive statistics for the demographic and attitudinal characteristics of the participants are provided in Appendix 2. On average, the 292 subjects in the experiment were 31 years old, were college-educated, and 63% identified themselves as female. About 59% of the participants reported themselves as primary shoppers and 23% indicated that they had children in the household. Participants in the experiment had diverse but higher household income on average compared to the U.S. population. The average WTP and WTA bids for the various aged yogurt-based drink and the two expiration date labels are presented in Table . Across all production dates, subjects' average WTA to perform the task of consuming the drink was $3.80 while the average WTP to purchase the drink was $1.60. This ratio is consistent in all treatment groups.

The density of the bids under the WTP and WTA auctions is displayed in Figure 1. This figure clearly illustrates the presence of censoring for both WTP and WTA. The true value of the product to a subject is only partially known because allowable bids were restricted to between $0 and $10. Under the WTA treatment, 21% (159) bid $0, indicating that they did not require compensation to ingest the product. One can infer that only these participants had a zero or higher WTP, because the other 79% bid strictly positive WTA, hence indicating strictly negative WTP. However, under the WTP mechanism, only 32% of the bids (227) were $0 and the other 68% bids

were strictly positive. As will be discussed in more details in the next section, we reject the hypothesis that the proportion of WTA > 0 is indifferent from the proportion of WTP = 0 for all products at the 1% level.

Censoring of WTP and WTA values occurs in experimental settings when bids cannot be less than $0, or greater than a certain threshold amount. When subjects actually have negative WTP (positive WTA), censoring affects interpretation of the results by introducing bias and vice versa. As shown in the top panel of Figure 1, the WTA bid distribution has a high density at $0.00. We know that the true WTA for those observations are *at most* $0 and probably negative for at least some participants (they would pay to perform the task, thus positive WTP). As shown in the lower panel of Figure 1, the WTP distribution also has a high density at $0. In that case, the true WTP are *at most* $0 but could be negative (subjects would have to be paid to consume the drink). Therefore, neither the WTP nor WTA mechanism covered the full range of true valuations in this experiment. We expect that the WTP and WTA censoring for each production date will be proportional to the density of the average WTP and WTA bids, indicating that framing effects play a significant role.

5. Results and Discussion

We use a random-effects Tobit model to assess the impact of production date on WTP and WTA bids for the drink, which are censored at the lower ($0) and upper ($10) limits. This model is appropriate because the dependent variable, the participant's true valuation, cannot be observed in certain cases while the independent variable, drink production date, is observable.

$$bid_{ij}^{*} = \begin{cases} 0 & \text{if } bid_{ij} \leq 0 \\ \alpha + \sum_j \beta_j \, I[\text{production}_j] + \epsilon_{ij} & \text{if } 0 < bid_{ij} < 10 \\ 10 & \text{if } bid_{ij} \geq 10 \end{cases} \quad (1)$$

In the model, bid_{ij}^{*} is the latent variable that represents bid_{ij}, the participant i's true valuation of the product with production date $j \in \{44, 49, 50, 51, 54\}$.[2] $I[\text{production}_j]$ is an indicator variable that takes value 1 for each production date. Table 3 presents the results of the random-effects Tobit specification regressing production date on the WTA and WTP offers when pooling the BEST IF USED BY and PRODUCE ON treatments. Under the WTP mechanism, the average bid significantly decreases relative to the base level of 44 days since production with additional days passed. In other words, subjects were less willing to pay for the product as it aged. Consistently, under the WTA mechanism, the average compensation required to consume the product increased as it aged. The reassuring result is, not only that the signs of product age's marginal effects are consistent, but also that the magnitudes of these marginal effects follow a consistent pattern comparing the WTA and WTP mechanisms. Specifically, as product's age increases, participants in the WTA treatment revealed monotonically increasing WTA (Tobit model coefficients as age increases: 0.783 < 1.221 < 1.830 < 3.036) while participants in the WTP treatment revealed monotonically decreasing WTP (Tobit model coefficients as age increases: -0.524 > -0.625 > -1.127 > -1.476) to consume the product.

Figure 2 further visualizes the results and illustrates these effects in margins plots in which the x-axis is the number of days since production. The two lines representing WTA and WTP are nearly mirror images of each other. In sum, we find that the effects of the two auction mechanisms on the bids for the drinks are proportional: as the yogurt-based drink ages, changes in the

[2] Uneven dates due to product availability at the time of experiment.

coefficients on WTA and WTP have similar magnitudes (but opposite signs: decreases in WTP and increases in WTA). When we separate the results for the BEST IF USED BY and PRODUCED ON treatments, the consistent effect of the drink's age is robust. In *Table 4*, columns 1 and 2 present the random-effects Tobit specification for WTA bids for the PRODUCED ON and BEST IF USED BY treatments. Notably, while the marginal effects of aging under WTA and WTP are consistent, responses to the BEST IF USED BY label were stronger than to the PRODUCED ON label. This is evident as the absolute values in column (2) are higher than the absolute values in columns (1) for each drink's age; similarly the absolute values in column (4) are higher than the absolute values in columns (3) for each drink's age. This is expected because PRODUCED ON does not provide clear information on when the drinks expire thus does not raise as much food safety concerns.

As illustrated in Figure 3, the drink's increasing age resulted in lower WTP bids and higher WTA bids, and the interaction effects are consistently stronger when participants were presented with the BEST IF USED BY label compared to the PRODUCED ON label. In sum, *Table 4* and Figure 3 further provide strong evidence that the marginal effects of the drink's age are consistent under WTA and WTP. Importantly, such consistency also holds with respect to the interaction effects between the drink's age and different expiration date labels.

We next evaluate the frequency of bids at the lower limit ($0.00) under the WTP and WTA treatments. In absence of the framing effect, we would expect to find similar proportions of $0 WTP bids and WTA bids greater than $0 for each product age. In other words, when WTP bids are censored, only a $0 bid may represent *negative WTP*, which equates to positive WTA. Our results, however, provide evidence of framing effects. As shown in Table 5, the frequency of $0 WTP bids is significantly lower ($p < 0.01$) than the frequency of positive WTA bids. When

participating under a WTA frame, the majority of participants required a compensation of more than $0 to consume a yogurt-based drink when the product has not yet passed its expiration date. In contrast, when participating under a WTP frame, the majority of participants again indicated that they were willing to pay more than $0 to consume the same drinks. This supports the literature that framing of the experiment has a significant impact on participants' behavior. Table presents the results of two-sample tests of these proportions and z-statistics for the differences and the consistently increasing trend in WTP and WTA.

To summarize, this study has three evidence-based findings. First, there exists consumable products for which some individuals are willing to pay while others require compensation to consume. In such case neither the conventional WTP nor WTA mechanism can cover the full range of consumers' true valuation. Second, incentive-compatible revealed preferences elicitation methods are subject to framing effects and the mean bids can be significantly different using the WTP versus WTA version of BDM. Third, despite the above two findings, the marginal effects of product characteristics as well as different labeling formats on consumer valuation are consistent between both elicitation mechanisms, indicating that WTP and WTA estimates have consistent policy implications.

6. Conclusions and Future Research

Consumer studies have often overlooked using WTA measures, which can be appropriate (even more valuable than WTP) when WTP values are censored. This is particularly important when looking at products that have potential negative consumer reactions, for example, due to safety, health, or ethical concerns. Our context of a yogurt-based drink with a range of days since

production allows us to measure WTP and WTA in a single experiment because including expired products allows us to consider consumers who potentially have negative WTP.

In this research, we conducted a framed field experiment that employed a within-subject treatment of days since the drink was produced and two dimensions of between-subject treatments: WTP versus WTA mechanisms and the wording of label indicating whether the drink had exceeded its expiration date (BEST IF USED BY and PRODUCED ON). This design allows us to measure the marginal effects of expiration date and labeling language under both WTP and WTA mechanisms.

We find that the marginal effect of the drink's increasing age under WTP and WTA mechanisms are consistent in terms of sign and statistical significance. This result is robust under various model specifications and the interaction effects between the drink's age and BEST IF USED BY versus PRODUCED ON labeling are also consistent. This result provides insights into consumer preferences experiments regarding controversial products. When studying products that some participants are willing to pay for, but others may derive negative utility from, some researchers may strongly prefer the WTP mechanism while others advocate for WTA. It is evident from this experiment that both WTA and WTP mechanisms can result in consistent marginal effects estimates and thus provide consistent marketing and policy implications.

Nevertheless, consistent with the existing literature, we find that BDM mechanisms are subject to framing effects and the mean bids can be significantly different using the WTP versus WTA version. The framing effects are quite persistence, indicated by the fact that the majority of our participants were willing to pay a positive amount in order to consumer the drinks under the WTP framing but, dramatically, the majority of participants requested some compensation for drinking the exact same products under the WTA framing.

Summarizing, in the context of controversial products, this study finds that both the WTA and WTP versions of the BDM mechanism result in consistent marginal effects estimates with respect to between-subjects and within-subjects treatments. Neither the WTA nor the WTP mechanism was able to discover the full range of participants' true values in our study, and we identified strong framing effects. Therefore, this study suggests that either mechanism can provide valid marketing and policy implications while neither is perfect.

Future experimental studies could consider a mechanism in which participants can report both positive WTP and negative WTP. This approach could be implemented by relaxing the range of bids and allowing bids to be either positive or negative. Such an auction likely would require providing participants with significant training since it would not be a familiar mechanism, but it could produce valuable results.

Tables

Table 1 Treatment Assignments

	WTP		WTA	
	A	B	C	D
Product Age	BEST IF USED BY	PRODUCED ON	BEST IF USED BY	PRODUCED ON
1) 44 days	A1	B1	C1	D1
2) 49 days	A2	B2	C2	D2
3) 50 days	A3	B3	C3	D3
4) 51 days	A4	B4	C4	D4
5) 54 days	A5	B5	C5	D5

Note: Columns denote the between-subjects treatments and rows denote the within-subject treatments

Table 2 Descriptive Statistics for Bids by Treatment

WTA					WTP				
Produced On					Produced On				
Age (Days)	Mean	Std. Dev.	Min.	Max.	Age (Days)	Mean	Std. Dev.	Min.	Max.
44	$3.33	3.26	$0.00	$10.00	44	$2.07	2.03	$0.00	$7.91
49	$3.73	3.22	$0.00	$10.00	49	$1.88	2.13	$0.00	$8.13
50	$3.98	3.30	$0.00	$10.00	50	$1.97	2.26	$0.00	$8.24
51	$4.16	3.45	$0.00	$10.00	51	$1.78	2.09	$0.00	$8.49
54	$4.41	3.42	$0.00	$10.00	54	$1.74	2.18	$0.00	$8.40
N	375				N	360			
Best If Used By					Best If Used By				
Age (Days)	Mean	Std. Dev.	Min.	Max.	Age (Days)	Mean	Std. Dev.	Min.	Max.
44	$2.36	2.89	$0.00	$10.00	44	$2.21	2.05	$0.00	$10.00
49	$3.05	3.15	$0.00	$10.00	49	$1.65	1.74	$0.00	$9.02
50	$3.38	3.24	$0.00	$10.00	50	$1.37	1.62	$0.00	$8.90
51	$4.02	3.29	$0.00	$10.00	51	$0.94	1.47	$0.00	$8.29
54	$5.59	3.38	$0.00	$10.00	54	$0.67	1.51	$0.00	$9.56
N	385				N	340			
Overall WTA Mean	$3.80	3.35	$0.00	$10.00	Overall WTP Mean	$1.64	1.98	$0.00	$10.00
Ratio of mean WTA/WTP			2.4						
Number of Observations (N)			1460						

Table 3 Random Effects Tobit Specification: Impact of Production Date on Offer for Pooled BEST IF USED BY and PRODUCED ON under WTP and WTA Mechanisms

		(1)	(2)
		WTA	WTP
Bid			
Base Level: Days after production date = 44			
Days after production date = 49		0.783***	-0.524***
		(0.219)	(0.140)
Days after production date = 50		1.221***	-0.625***
		(0.219)	(0.140)
Days after production date = 51		1.830***	-1.127***
		(0.219)	(0.144)
Days after production date = 54		3.036***	-1.476***
		(0.220)	(0.148)
Constant		2.023***	1.789***
		(0.390)	(0.231)
Sigma_*u* constant		4.321***	2.426***
		(0.299)	(0.175)
Sigma_*e* constant		1.699***	1.071***
		(0.0596)	(0.0397)
Observations		760	700
Left-censored observations		159	227
Right-censored observations		73	1
Standard errors are shown in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.			

Table 4 Random-effects Tobit Specification: Impact of Production Date on Bids for Separate BEST IF USED BY and PRODUCED ON Treatments under WTP and WTA Mechanisms (Baseline: Days after production date=44 days)

	(1)	(2)	(3)	(4)
	WTA		**WTP**	
	PRODUCED ON	BEST IF USED BY	PRODUCED ON	BEST IF USED BY
Bid				
Days after production date				
49	0.503*	1.085***	-0.286	-0.756***
	(0.265)	(0.322)	(0.174)	(0.201)
50	0.873***	1.591***	-0.200	-1.046***
	(0.265)	(0.321)	(0.175)	(0.202)
51	1.139***	2.530***	-0.436**	-1.869***
	(0.266)	(0.322)	(0.176)	(0.214)
54	1.468***	4.572***	-0.540***	-2.589***
	(0.265)	(0.327)	(0.177)	(0.233)
Constant	2.598***	1.424***	1.626***	1.952***
	(0.594)	(0.529)	(0.352)	(0.295)
Sigma_*u* constant	4.516***	4.119***	2.651***	2.111***
	(0.455)	(0.391)	(0.264)	(0.221)
Sigma_*e* constant	1.441***	1.760***	0.947***	1.085***
	(0.0699)	(0.0888)	(0.0462)	(0.0611)
Observations	375	385	360	340
Left-censored observations	76	83	99	128
Right-censored observations	32	41	0	1

Standard errors in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 5 Two-sample Test of Proportions Using Indicator Variables WTA = 0 and WTP > 0 by Product Age

Product Age	Proportion of WTA > 0	Proportion of WTP = 0	z Statistic of Difference	
44 days	69%	19%	8.54	***
49 days	77%	27%	8.52	***
50 days	80%	27%	8.99	***
51 days	83%	39%	7.78	***
54 days	87%	50%	6.81	***

Figures

Figure 1 Distribution of Bids by Pooled WTA and WTP Treatments

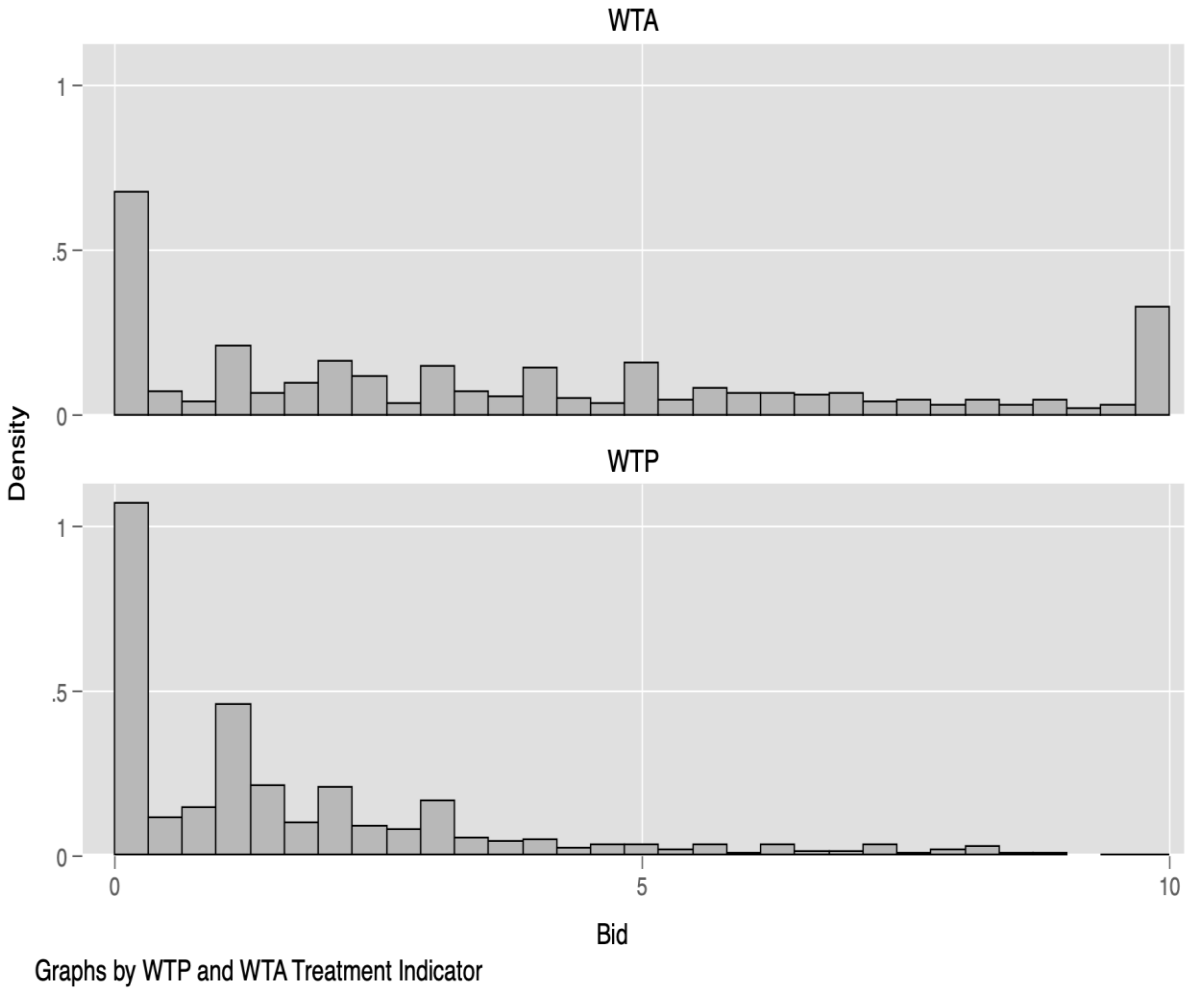


Figure 2 Predictive Margins for Pooled BEST IF USED BY and PRODUCED ON Treatments

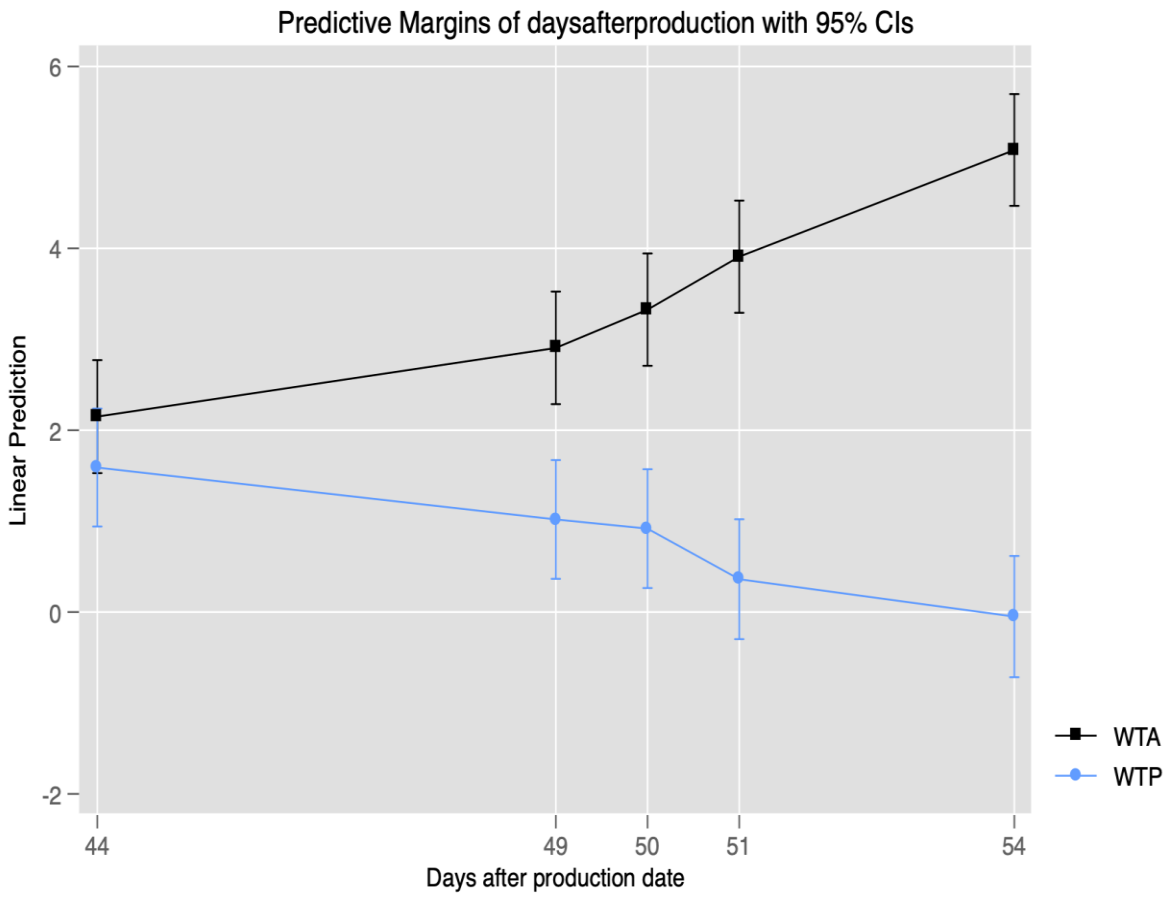
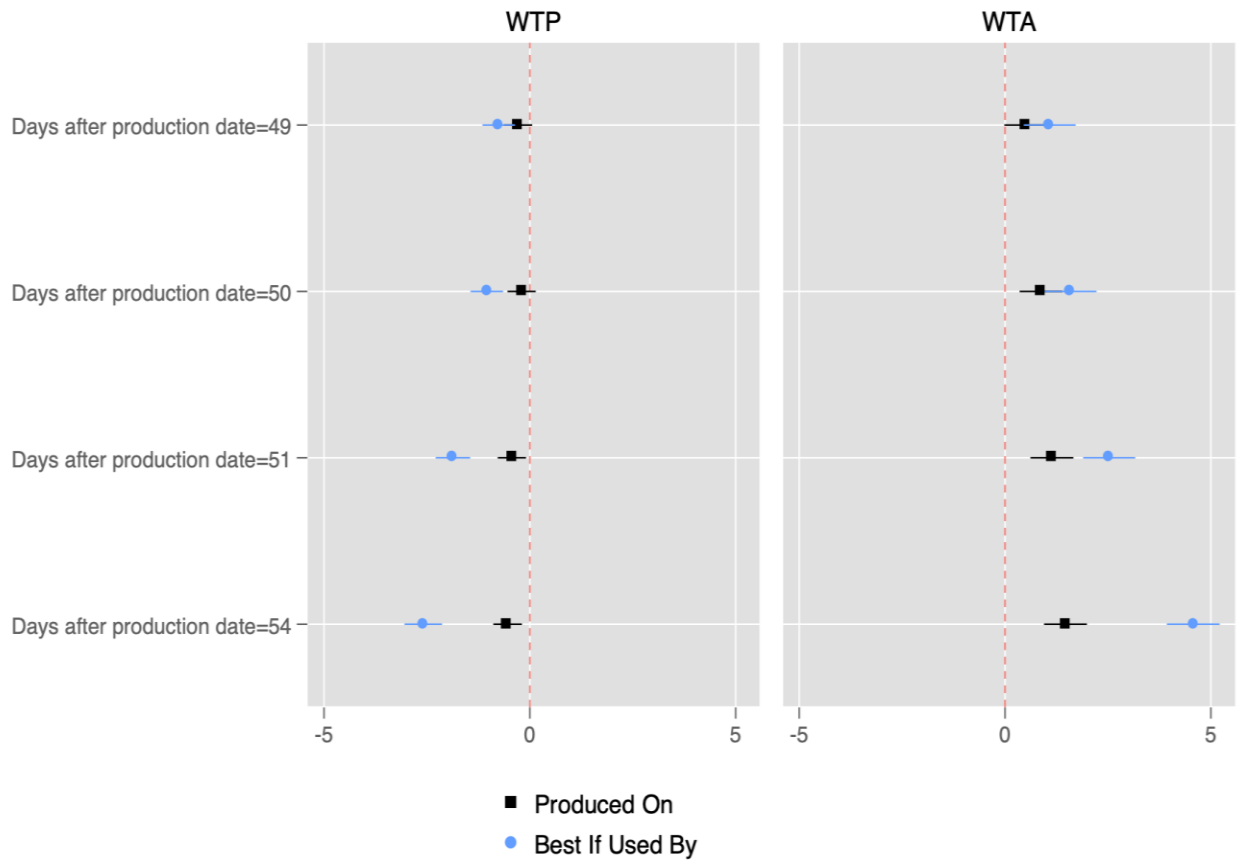


Figure 3 Coefficient Plot of WTP and WTA Treatments for BEST IF USED BY and PRODUCED ON Treatments (Baseline: Days after production date=44 days)



References

Bass, D.A., B.R. McFadden, and K.D. Messer. 2021. "A case for measuring negative willingness to pay for consumer goods." Food Policy 104 (2021): 102126.

Bateman, I., A. Munro, B. Rhodes, C. Starmer, and R. Sugden. 1997. "A Test of the Theory of Reference-Dependent Preferences." *The Quarterly Journal of Economics* 112(2):479–505.

Becker, G.M., M.H. DeGroot, and J. Marschak. 1964. "Measuring utility by a single-response sequential method." *Behavioral science* 9(3):226–232. Available at: http://doi.wiley.com/10.1002/bs.3830090304 [Accessed July 22, 2020].

Boyce, R.R., T.C. Brown, G.H. McClelland, G.L. Peterson, and W.D. Schulze. 1992. "An experimental examination of intrinsic values as a source of the WTA-WTP disparity." *American Economic Review* 82(5):1366–1373.

Brebner, S., and J. Sonnemans. 2018. "Does the elicitation method impact the WTA/WTP disparity?" *Journal of Behavioral and Experimental Economics* 73:40–45. Available at: https://doi.org/10.1016/j.socec.2018.01.006 [Accessed December 17, 2020].

Brookshire, D.S., and D.L. Coursey. 1987. *Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures*.

Brown, T.C. 2005. "Loss aversion without the endowment effect, and other explanations for the WTA-WTP disparity." *Journal of Economic Behavior and Organization* 57(3):367–379.

Corrigan, J., and M.C. Rousu. 2008. "Testing whether field auction experiments are demand revealing in practice." *Journal of Agricultural and Resource Economics* 33(2):290–301.

Costa-Font, M., J.M. Gil, and W.B. Traill. 2008. "Consumer acceptance, valuation of and attitudes towards genetically modified food: Review and implications for food policy." *Food Policy* 33(2):99–111.

Costanigro, M., and J.L. Lusk. 2014. "The signaling effect of mandatory labels on genetically engineered food." *Food Policy* 49(P1):259–267.

Coursey, D.L., J.L. Hovis, and W.D. Schulze. 1987. "The Disparity Between Willingness to Accept and Willingness to Pay Measures of Value." *The Quarterly Journal of Economics* 102(3):679–690. Available at: http://www.jstor.org/stable/1884223 [Accessed October 11, 2020].

D'amato, A., T. Goeschl, L. Lorè, and M. Zoli. 2020. "Date Marks, Valuation, and Food Waste: Three In-Store 'Eggsperiments.'" Available at: https://ec.europa.eu/food/safety/food_waste/eu_actions/date_marking_en [Accessed November 14, 2020].

Dillaway, R., K.D. Messer, J.C. Bernard, and H.M. Kaiser. 2011. "Do Consumer Responses to Media Food Safety Information Last?" *Applied Economic Perspectives and Policy* 33(3):363–383.

Drichoutis, A.C., J.L. Lusk, and V. Pappa. 2016. "Elicitation formats and the WTA/WTP gap: A study of climate neutral foods." *Food Policy* 61:141–155.

Ericson, K., and A. Fuster. 2014. "The Endowment Effect." *Annual Review of Economics*

6(1):555–579. Available at: www.annualreviews.org [Accessed May 5, 2020].

Fehr, D., R. Hakimov, and D. Kübler. 2015. “The willingness to pay-willingness to accept gap: A failed replication of Plott and Zeiler.” *European Economic Review* 78:120–128. Available at: http://dx.doi.org/10.1016/j.euroecorev.2015.05.006.

Heberlein, T.A., and R.C. Bishop. 1986. “Assessing the validity of contingent valuation: Three field experiments.” *Science of the Total Environment, The* 56(C):99–107.

Heiman, A., and D. Zilberman. 2011. “The Effects of Framing on Consumers’ Choice of GM Foods.” *AgBioForum* 14(3):171–179. Available at: https://mospace.umsystem.edu/xmlui/bitstream/handle/10355/12464/EffectsFramingConsumersChoice.pdf?sequence=1&isAllowed=y [Accessed November 22, 2020].

Horowitz, J.K., and K.E. McConnell. 2002. “A review of WTA/WTP studies.” *Journal of Environmental Economics and Management* 44(3):426–447.

Isoni, A., G. Loomes, and R. Sugden. 2011. “The willingness to pay - Willingness to accept gap, the ‘endowment effect,’ subject misconceptions, and experimental procedures for eliciting valuations: Comment.” *American Economic Review* 101(2):991–1011. Available at: http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.2.991 [Accessed November 22, 2020].

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

Kahneman, D., J.L. Knetsch, and R.H. Thaler. 1991. “Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias.” Journal of Economic perspectives 5.1 (1991): 193-206.

Kecinski, M., D.K. Keisner, K.D. Messer, and W.D. Schulze. 2018. “Measuring Stigma: The Behavioral Implications of Disgust.” *Environmental and Resource Economics* 70(1):131–146. Available at: https://doi.org/10.1007/s10640-017-0113-z [Accessed October 16, 2020].

Knetsch, J.L. 1989. “The Endowment Effect and Evidence of Nonreversible Indifference Curves.” *The American Economic Review* 79(5):1277–1284.

Knetsch, J.L., and J.A. Sinden. 1984. “Willingness to Pay and Compensation Demanded: Experimental Evidence of an Unexpected Disparity in Measures of Value.” *The Quarterly Journal of Economics* 99(3):507. Available at: https://academic.oup.com/qje/article-lookup/doi/10.2307/1885962 [Accessed August 13, 2020].

Kovalchik, S., C.F. Camerer, D.M. Grether, C.R. Plott, and J.M. Allman. 2005. “Aging and decision making: A comparison between neurologically healthy elderly and young individuals.” *Journal of Economic Behavior and Organization* 58(1):79–94.

Li, T., K.D. Messer, and H.M. Kaiser. 2020. “The impact of expiration dates labels on hedonic markets for perishable products.” *Food Policy* (February):101894. Available at: https://doi.org/10.1016/j.foodpol.2020.101894.

Liaukonyte, J., B.J. Rickard, H.M. Kaiser, A.M. Okrent, and T.J. Richards. 2012. “Economic and Health Effects of Fruit and Vegetable Advertising: Evidence from Lab Experiments.” *Food Policy* 37(5):543–553.

List, J.A. 2011. “Does Market Experience Eliminate Market Anomalies ? The Case of Exogenous

Market Experience." *The American Economic Review* 101(3):313–317.

List, J.A. 2003. "Does Market Experience Eliminate Market Anomalies?" *The Quarterly Journal of Economics* 118(1):41–71. Available at: https://academic.oup.com/qje/article-lookup/doi/10.1162/00335530360535144 [Accessed November 22, 2020].

List, J.A. 2004. "Neoclassical theory versus prospect theory: Evidence from the marketplace." *Econometrica* 72(2):615–625. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1468-0262.2004.00502.x [Accessed November 22, 2020].

Lloyd-Smith, P., and W. Adamowicz. 2018. "Can stated measures of willingness-to-accept be valid? Evidence from laboratory experiments." *Journal of Environmental Economics and Management* 91:133–149. Available at: https://doi.org/10.1016/j.jeem.2018.07.003 [Accessed November 5, 2020].

Lusk, J.L., M. Jamal, L. Kurlander, M. Roucan, and L. Taulman. 2005. *A meta-analysis of genetically modified food valuation studies*.

McCluskey, J. J., J. Swinnen, and T. Vandemoortele. 2015. "You get what you want: A note on the economics of bad news." *Information Economics and Policy* 30: 1–5.

Messer, K.D., H.M. Kaiser, C. Payne, and B. Wansink. 2011. "Can generic advertising alleviate consumer concerns over food scares?" *Applied Economics* 43(12):1535–1549. Available at: https://www.tandfonline.com/action/journalInformation?journalCode=raec20 [Accessed June 18, 2021].

Messer, K.D., G.L. Poe, D. Rondeau, W.D. Schulze, and C.A. Vossler. 2010. "Social preferences and voting: An exploration using a novel preference revealing mechanism." *Journal of Public Economics* 94(3–4):308–317.

Plott, C.R., and K. Zeiler. 2007. "Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory?" *American Economic Review* 97(4):1449–1466.

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.

Tunçel, T., and J.K. Hammitt. 2014. "A new meta-analysis on the WTP/WTA disparity." *Journal of Environmental Economics and Management* 68(1):175–187. Available at: http://dx.doi.org/10.1016/j.jeem.2014.06.001.

Tversky, A., and D. Kahneman. 1974. "Judgment under uncertainty: Heuristics and biases." *science* 185(4157):1124–1131.

Appendix 1 Experimental Instructions

Instructions for Choice Activity (WTA):

You will receive $10 for completing this study. We want to know about your views and your preferences for foods with varying production dates. In these instructions, we are going to describe some Stonyfield organic strawberry yogurt smoothies that can be found in most grocery stores.

For each smoothie, we will ask if you are willing to drink it for cash compensation. You will indicate the minimum amount that we would have to pay you to drink each smoothie. The amount will be referred to as your bid. After you submit your bid for each smoothie, the system will randomly select **one** of them. For that smoothie, we will randomly draw a "market price." **If** your bid is below the "market price", you will drink the selected smoothie and we will pay you the "market price." Otherwise you will not be eligible to drink the smoothie. If you are eligible, you will be given the smoothie to drink at the end of the study and the compensation for drinking the smoothie will be added to your participation fee of $10.

Note that only one smoothie option will be binding. This option will determine whether you drink the smoothie and how much cash you receive. The binding decision will be determined randomly after you have made all of your decisions.

Also note that you cannot influence the randomly drawn price with your bid. Therefore, it is in your best interest to submit a bid equal to the minimum amount of money you would have to be paid for drinking each smoothie. For example, if you really want to drink the smoothie, you should bid zero. If you do not want to drink the smoothie, you should bid a positive amount reflecting the minimum amount to compensate you for having to drink it.

*New Page
After you have submitted your bids, only one smoothie will be randomly selected. The price of the selected smoothie will be determined randomly. Then there are two possible outcomes:

You DRINK the selected smoothie. This happens when your bid is lower than or equal to the randomly determined price. In this case, you will drink the smoothie and receive the compensation plus the $10 participation fee.

You DO NOT DRINK the selected smoothie. This happens when your bid is higher than the randomly determined price. In this case, you will not drink any smoothie nor get the compensation, but instead receive the $10 participation fee.

*Practice round

*New page
Now you will bid on five Stonyfield organic strawberry yogurt smoothies. Keep in mind that you will get to drink at most one smoothie at the end of this study. Note that your bids will no longer be hypothetical.

Instructions for Choice Activity (WTP):

You will receive $10 for completing this study. We want to know about your views and your preferences for foods with varying production dates. In these instructions, we are going to describe some Stonyfield organic strawberry yogurt smoothies that can be found in most grocery stores.

For each smoothie, we will ask if you are willing to pay to drink it. You will indicate the maximum amount that you would be willing to pay to drink each smoothie. The amount will be referred to as your bid. After you submit your bid for each smoothie, the system will randomly select **one** of them. For that smoothie, we will randomly draw a "market price." **If** your bid is above the "market price", you will drink the selected smoothie and pay the "market price." Otherwise you will not be eligible to drink the smoothie. If you are eligible, you will be given the smoothie to drink at the end of the study and the cost of drinking the smoothie will be deducted from your participation fee of $10.

Note that only one smoothie option will be binding. This option will determine whether you drink the smoothie and how much you pay. The binding decision will be determined randomly after you have made all of your decisions.

Also note that you cannot influence the randomly drawn price with your bid. Therefore, it is in your best interest to submit a bid equal to the maximum amount of money you would pay to drink each smoothie. For example, if you really do not want to drink the smoothie, you should bid zero. If you want to drink the smoothie, you should bid a positive amount reflecting the maximum amount of money you would pay to drink it.

*New Page
After you have submitted your bids, only one smoothie will be randomly selected. The price of the selected smoothie will be determined randomly. Then there are two possible outcomes:

You DRINK the selected smoothie. This happens when your bid is greater than or equal to the randomly determined price. In this case, you will drink the smoothie and receive the remaining balance from your $10 participation fee after the price has been paid.

You DO NOT DRINK the selected smoothie. This happens when your bid is lower than the randomly determined price. In this case, you will not drink any smoothie nor pay the price, but instead receive the $10 participation fee.

*Practice Round

*New Page
Now you will bid on five Stonyfield organic strawberry yogurt smoothies. Keep in mind that you will drink at most one smoothie at the end of this study. Note that your bids will no longer be hypothetical.

Appendix 2. Summary Statistics for Demographic Variables

Number of Respondents	292
Average age (years)	31
Variable	Percentage of Respondents
Indicator if Female	63%
Primary Shopper	59%
Indicator if Children in Household	23%
Education (highest level)	
Some high school	1%
High school graduate	9%
Some college/Associate	38%
Bachelor's degree	26%
Graduate degree/Professional	27%
Household Income (in 2016)	
Less than $10000	15%
$10000-$14999	3%
$15000-$24999	5%
$25000-$34999	5%
$35000-$49999	10%
$50000-$74999	13%
$75000-$99999	16%
$100000-$149999	19%
$150000-$199999	8%
$200000-$249999	3%
$250000 and above	4%

Appendix 3. Tobit with Demographic Controls

	(1)	(2)
	WTA-pooled	WTP-pooled
Bid		
Days since production		
49	0.782***	-0.524***
	(0.219)	(0.140)
50	1.221***	-0.624***
	(0.219)	(0.140)
51	1.830***	-1.126***
	(0.219)	(0.144)
54	3.036***	-1.474***
	(0.220)	(0.148)
Indicator whether female = 1	0.168	1.256***
	(0.754)	(0.454)
Indicator of college+ education = 1	0.520	-0.0870
	(1.549)	(0.614)
Age in years	-0.0145	-0.0130
	(0.0299)	(0.0161)
Primary shopper = 1	-0.260	-0.651
	(0.791)	(0.466)
Indicator whether children in household=1	1.204	0.0940
	(0.846)	(0.512)
Constant	1.725	1.822**
	(1.703)	(0.782)
Sigma_*u* constant	4.281***	2.316***
	(0.296)	(0.167)
Sigma_*e* constant	1.699***	1.070***
	(0.0596)	(0.0396)
Observations	760	700

Standard errors in parentheses. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Appendix 4: Tobit Specification: Expired Indicator

	(1)	(2)
	WTA-pooled	WTP-pooled
Bid		
Indicator if Expired	1.772***	-0.922***
	(0.150)	(0.0985)
Constant	2.691***	1.411***
	(0.369)	(0.218)
sigma_u		
Constant	4.313***	2.419***
	(0.300)	(0.175)
sigma_e		
Constant	1.810***	1.104***
	(0.0635)	(0.0409)
Observations	760	700

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$