Food Policy 104 (2021) 102126

Contents lists available at ScienceDirect

Food Policy

journal homepage: www.elsevier.com/locate/foodpol

A case for measuring negative willingness to pay for consumer goods

Daniel A. Bass, Brandon R. McFadden *, Kent D. Messer

Department of Applied Economics and Statistics at University of Delaware, 531 S. College Avenue, Rm 224, Newark, DE 19716, United States

ARTICLE INFO

Keywords:
WTP
Willingness to accept
Negative willingness to pay
Market goods

ABSTRACT

Negative willingness to pay (WTP) has received significant attention in the environmental economics literature but generally has been ignored by economists interested in measuring consumer preferences for food and beverages. Most consumer preference studies have focused solely on WTP measures and, in most cases, have treated negative consumer responses to a product as equivalent to consumers simply not wanting to purchase it since both sentiments have a WTP estimate of $0. However, for some goods such as bottled water, this approach can be an oversimplification that fails to reveal consumers' true sentiments, a problem that is relevant in policy contexts. To test this question, we conduct a randomized controlled study involving 1384 adult consumers that tests how three unconventional elicitation approaches affect estimates of consumer preferences for nearly identical goods: bottled water and tap water. We find that stated values for both types of water differ by treatment and that ignoring negative WTP values upwardly biases the premium consumers are willing to pay for bottled water. Our findings show that allowing negative values of WTP provides more-accurate valuations and measures of differences in valuations for two goods, something that is particularly important for policymaking since consumer valuations can reflect their support for efforts to address externalities associated with consumer goods or may reflect other consumer sentiments, such as food safety concerns.

1. Introduction

Willingness to pay (WTP) and willingness to accept (WTA) have historically been collected separately and treated as different concepts based on property rights. WTP is the maximum amount an individual will pay to acquire the rights to something while WTA is the minimum amount of compensation an individual requires to sell something (Shogren et al., 1994). Thus, for example, WTP can be used to value improved environmental quality while WTA can be used to measure compensation required for reduced environmental quality (Ahlheim and Buchholz, 2000). Expected-utility theory posits that an individual's WTP and WTA are equal for any good, but empirical studies have determined that those values often are not equal (Knetsch and Sinden, 1984; Kahneman et al., 1991; Knetsch, 1995; Horowitz and Kenneth, 2000; Plott and Zeiler, 2005; Kecinski et al., 2018). Historically, differences in individual WTP and WTA have been attributed to a variety of theoretical explanations, such as loss aversion (Kahneman et al., 1991), an endowment effect (Thaler, 1980), status quo bias (Samuelson and Zeckhauser, 1988), prospect theory (Kahneman and Tversky, 1979), income and substitution effects (Hanemann, 1991), and uncertainty (Zhao and Kling, 2001).

Treating WTP and WTA as equivalent also does not account for scenarios in which a consumer does not own a good but nevertheless could require compensation to consume it. This can arise when consumption negatively affects intrinsic valuations of individual characteristics, such as personal health and branding, and it has been argued that kinked indifference curves can occur when intrinsic values are included in WTA and excluded from WTP (Boyce et al., 1992). For example, plastic use and pollution associated with bottled water are common environmental concerns (Saylor et al., 2011). Thus, some environmentalists may be strongly opposed to consumption of bottled water because of its associated negative externalities (large-scale plastic waste) and consequently would not accept a bottle of water even at no cost. Additionally, it is possible that some consumers may have safety concerns related to either tap or bottled water and may again not want to accept this water at no cost.

Since intrinsic values can cause WTP for a market good to be negative, typical studies that just measure WTP cannot accurately represent an environmentally-oriented or safety concern-oriented consumer's true value for bottled water because they assume there is effective demand at $0. Allowing consumers to express negative WTP for market goods allows researchers to differentiate between consumers who have effective

* Corresponding author.
E-mail address: foodecon@udel.edu (B.R. McFadden).

https://doi.org/10.1016/j.foodpol.2021.102126
Received 17 December 2020; Received in revised form 24 May 2021; Accepted 5 July 2021
Available online 14 July 2021

demand at $0 and those who have negative demand and must be compensated to consume the good. Measuring this negative WTP can also be valuable when evaluating consumer support for policy efforts to address negative externalities associated with market goods.

The contingent valuation literature has long explored the concept of negative WTP in the context of environmental goods, finding in some cases that respondents truly have negative WTP (Bohara et al., 2001) and, in other cases, that the respondents likely did not understand the environmental good being valued (Loomis and Ekstrand, 1998). Several studies questioned whether negative WTP should be eliminated using econometric methods (Haab and McConnell, 1997, 1998). However, there is empirical evidence of the existence of negative WTP for non-market (e.g., environmental) goods (Bohara et al., 2001). It typically is observed when individuals value the status quo more than a potential policy. For example, negative WTP has been observed when individuals valued continued cattle grazing more than environmental preservation (Lockwood et al., 1994) and when they valued increased access to public spaces (Campbell et al., 2014; Lundhede et al., 2013; Jacobsen et al., 2011). Moreover, Hanley et al. (2009) concluded that ignoring negative WTP would upwardly bias estimates of a woodland planting project by as much as 44%. These negative WTP values typically were explained by individual preferences regarding protecting habitats and biodiversity (Campbell et al., 2014; Lundhede et al., 2013; Jacobsen et al., 2011).

The literature measuring negative WTP for food, drinks, and other market goods is sparse. In fact, one issue with using contingent valuation to measure the value of a market good is the assumption that respondents are interested in purchasing the good and thus derive utility from purchasing it. However, goods exist that do not contribute utility or contribute negatively to utility (Kriström, 1997), such as bottled water for environmentalists and animal-based meat for vegans. As a result, collecting only positive WTP values bounded at $0 for market goods can upwardly bias estimates of effective demand, which would also bias welfare estimates that can impact the policy evaluation.

Many econometric methods used to estimate WTP, such as log-logistic and Weibull models, assume that all respondents have positive WTP (Kriström, 1997). Normalizing a model by the price coefficient, sometimes called 'fixing' the price coefficient, allows a researcher to estimate a model in WTP space and directly interpret coefficient estimates as WTP values. An additional benefit of Modelling in WTP space is sign preservation so that the incidence of a negative WTP estimates is decreased, which is a desirable property because a negative WTP is not generally assumed to be plausible (Hensher et al., 2015). Random parameter logit, or mixed logit, models are often estimated to examine choice experiment data, which allows negative preferences and WTP values for products, attributes, or attribute levels (McKendree et al., 2013; Tonsor et al., 2005; Lusk et al., 2003). Although, any negative WTP is relative to another product, attribute, or attribute level. For example, a WTP value of -$0.70 for bottle packaging in Staples et al. (2020) also indicates a $0.70 premium for aluminum cans. It is common in these models to force the price coefficient to be negative, which assumes increasing prices are associated with marginal disutility and is reasonable given the law of demand. However, forcing WTP to be positive for a product assumes that it provides marginal utility, which may not always be reasonable.

Other econometric methods focus on how to treat zero WTP values when estimating the effects of product attributes or individual characteristics on WTP. Tobit models are typically estimated to left-censor WTP at zero (Jackman and Lorde, 2014; Halstead et al., 1991), as ordinary least squares estimates may be biased due to censored values, but Tobit models can also be right-censored for logarithm WTP values after removing the undefined zero bids (Oduor et al., 2016). While a researcher could easily left-censor a Tobit model at a negative value, to the best of our knowledge this has not generally been done using negative WTP values. The underlying distribution of Poisson models, or the more general negative binomial, assumes non-negativity. Poisson models are commonly used for contingent valuation data and zero-inflated Poisson are relied upon when the data have a high proportion of zero WTP values (Cameron and Englin, 1997). Furthermore, hurdle and zero-inflated models, like the zero-inflated ordered probit, are particularly useful when a researcher is interested in estimating differences between non-consumers, potential consumers, and current consumers (Jiang et al., 2017). These are excellent techniques when dealing with zero WTP values, but have either not been implanted or necessarily useful for negative values.

To investigate the potential effects of negative WTP for market goods, we used an experimental approach in which respondents were randomly assigned to one of four contingent valuation treatments: (1) WTP Only; (2) WTA Only; (3) WTP/WTA Choice; and (4) WTP/WTA Scale. These treatments are described in greater detail in Section 2.2.

The experiment elicited stated values for 16 oz of bottled water and tap water. This design allowed us to compare values for bottled water and tap water across treatments and differences in individual values for bottled and tap water. We find that the elicited values vary by treatment and that the overall estimated mean values were greatest for the WTP Only treatment and least for the WTA Only treatment. However, when we examined differences in individual valuations of the two types of water and analyzed positive and negative values separately, we found important variations associated with the elicitation mechanisms.

2. Methodology

2.1. Survey overview

This study was approved by the Institutional Review Board at the University of Delaware. Between April 1 and April 8, 2019, 1384 responses were collected from U.S. adults using an online panel maintained by Qualtrics. Quota-based sampling was used to ensure that respondent characteristics matched the U.S. population based on age, education, sex, and income. Respondents provided informed consent before answering any questions. Respondents were randomly assigned to a contingent valuation treatment and answered various questions, used as covariates, to complete the study; both the treatments and covariates are described in more detail in the following subsections.

2.2. Willingness-to-pay and willingness-to-accept treatments

To study the effect of elicitation methods on WTP and WTA values, we randomly assigned respondents to one of four treatments. All four treatments used a cheap talk strategy. The treatments were:

(1) **WTP Only** – respondents were asked a standard question to elicit their WTP values for bottled and tap water bounded between $0 and $4;
(2) **WTA Only** – respondents were asked a standard question to elicit WTA values for bottled and tap water bounded between -$4 and $0;
(3) **WTP/WTA Choice** – respondents chose either the WTP Only or the WTA Only treatment by first stating whether they would pay or would have to be compensated to consume bottled and tap water; and
(4) **WTP/WTA Scale** – respondents were offered both positive and negative values, bounded between –$4 and $4, to consume bottled water and tap water rather than choosing to self-select the WTP Only or WTA Only treatments.

Screenshots of the scripts used for WTP and WTA elicitation are presented in Appendix Figs. 1–6. The WTP only treatment script is in Appendix Fig. A1, WTA Only is in Appendix Fig. A2, WTP/WTA Choice is in Appendix Figs. A3-A5 because there were multiple steps required for this treatment, and WTP/WTA Scale is in Appendix Fig. A6.

Since studies have typically only elicited WTP to acquire a market good, the WTP Only treatment can be thought of as a control. The WTA

Table 1
Respondent characteristics by treatment.

Question	Response Option	WTP Only	WTA Only	WTP/ WTA Choice	WTP/ WTA Scale
Age	18–24	12.57	13.25	11.87	13.73
	25–34	16.20	14.46	17.21	18.77
	35–44	19.27	17.77	15.13	15.97
	45–54	18.16	17.17	18.10	14.57
	55–64	16.48	17.47	18.69	16.53
	65+	17.32	19.88	18.99	20.45
Education	Less than high school degree	1.96	0.60	0.59	1.40
	High school graduate (including GED)	17.04	14.16	17.21	20.45
	Some college but no degree	20.67	28.31	24.93	21.57
	Associate degree in college (2-year)	12.29	13.86	10.98	10.92
	Bachelor's degree in college (4-year)	29.89	29.82	29.97	28.01
	Master's degree	15.64	10.54	11.57	14.01
	Doctoral degree	0.84	1.51	2.08	1.12
	Professional degree (JD, MD)	1.68	1.20	2.67	2.52
Sex	Male	48.32	48.80	51.63	46.22
	Female	51.68	51.20	48.37	53.78
Income	Less than $25,000	14.80	17.47	18.69	18.21
	$25,000 to $50,000	24.58	24.40	21.96	22.69
	$50,001 to $75,000	22.07	17.17	18.10	16.25
	$75,001 to $100,000	10.61	15.06	16.32	14.57
	$100,001 to $150,000	16.48	15.96	14.54	14.85
	$150,001to $200,000	5.87	3.31	4.75	8.12
	$200,001+	5.59	6.63	5.64	5.32
Density	Urban	27.65	27.71	26.11	27.17
	Suburban	52.51	50.00	52.82	48.74
	Rural	19.83	22.29	21.07	24.09
Employment Status	Unemployed	8.66	9.64	7.72	10.08
	Part time	10.06	12.05	9.5	9.52
	Full time	45.81	41.57	46.88	41.46
	Student	5.59	6.02	5.34	8.68
	Retired	21.79	25.60	21.66	24.65
	Not currently seeking employment	8.10	5.12	8.90	5.6
Race/ Ethnicity*	White	77.93	77.11	80.12	78.71
	Black or African American	6.98	8.43	6.23	7.28
	Hispanic, Latino, or Spanish Origin	3.91	3.31	3.56	2.52
	Other	11.18	11.15	10.09	11.49
Cultural Cognition Thesis	Individualistic	23.45	22.70	22.67	22.94
	Hierarchical	18.56	18.57	18.45	19.00
Percent of Water Consumed that is Bottled versus Tap		55.02	55.02	52.75	56.10
Healthier	Bottled	54.19	51.51	56.38	53.50
	Tap	12.57	12.35	10.09	12.04
	Equally	33.24	36.14	33.53	34.45
Tastier	Bottled	66.76	63.25	69.73	65.55
	Tap	12.57	16.57	11.57	14.85
	Equally	20.67	20.18	18.69	19.61
More Sustainable	Bottled	30.17	26.51	30.56	25.77
	Tap	42.74	49.40	44.51	49.86
	Equally	27.09	24.10	24.93	24.37
n		358	332	337	357

Note: A MANOVA using 1384 observations was estimated to ensure randomization across treatments, and the null hypothesis of no difference between treatments was not rejected [Wilks' Lambda F(3, 1,380) = 0.984, P-value < = 0.998]. *Races with proportions less than 3% were combined to compare randomization across treatments.

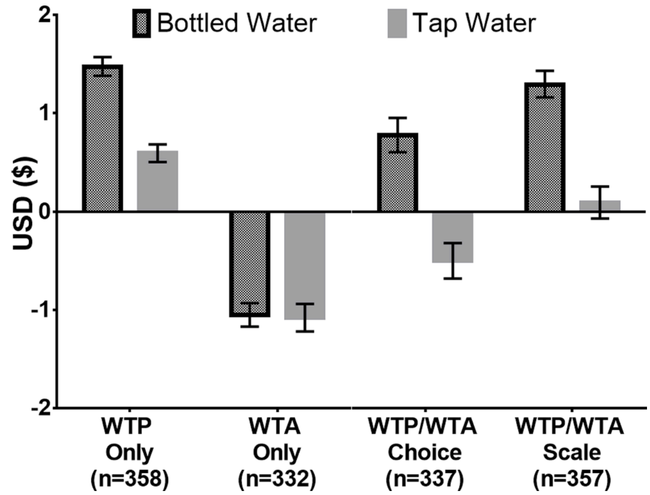

Fig. 1. WTP and WTA Means by Treatment. Note: The error bars represent upper and lower limits of a 95% confidence interval.

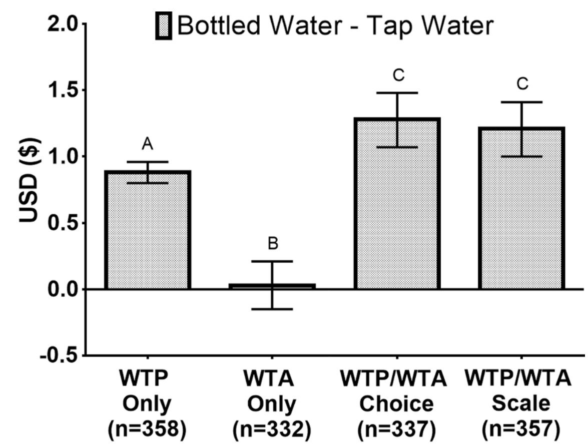

Fig. 2. Mean Difference Between values for Bottled Water Minus Tap Water by Treatment. Note: The error bars represent upper and lower limits of a 95% confidence interval. The letters above the error bars represent groupings based on pairwise comparisons using T-tests with Bonferroni corrected P-values.

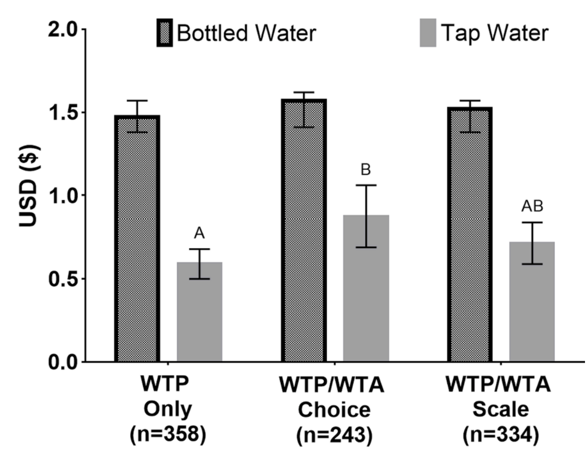

Fig. 3. Positive WTP Means by Treatment with 95% Confidence Intervals. Note: The error bars represent upper and lower limits of a 95% confidence interval. The letters above the error bars represent groupings based on pairwise comparisons using T-tests with Bonferroni corrected P-values.

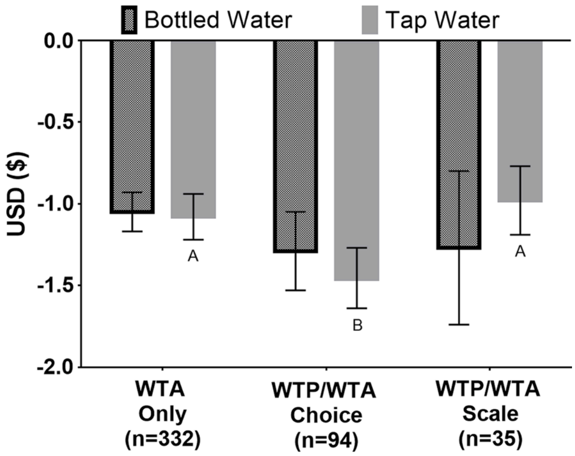

Fig. 4. Negative WTA Means by Treatment with 95% Confidence Intervals. Note: The error bars represent upper and lower limits of a 95% confidence interval. The letters below the error bars represent groupings based on pairwise comparisons using T-tests with Bonferroni corrected P-values.

Only treatment is perhaps the least conventional and tests whether respondents offer negative values to consume a good. Though $0 was an option in both treatments, they forced respondents into choice sets that were restricted to either non-negative or non-positive values. The WTP/WTA Choice treatment, on the other hand, allowed respondents to self-select into positive (WTP Only) or negative values (WTA Only), and the WTP/WTA Scale treatment allowed them to state positive or negative values without the first stage of self-selection.

We study bottled water and tap water because of consumer stigma associated with tap water in terms of taste and quality (Anadu and Harding, 2000; Jakus et al., 2009; Hu, Morton and Mahler, 2011; McSpirit and Reid, 2011). Also, because of the extremely low cost associated with consuming tap water, we expected that WTP for it would be close to or even less than $0. By allowing negative WTP, we can better determine monetary differences in consumers' demand for bottled water and tap water. In addition, bottled water is a unique good for which some segments of consumers (e.g., environmentalists) are likely to have negative WTP.[1]

2.3. Covariates

Beyond the questions used for sampling (i.e., age, education, sex, and income), we also asked respondents about the density of their community (i.e., urban, suburban, or rural), employment status, race, and ethnicity. Respondents also completed the Cultural Cognition Quiz, which is a two-dimensional scale (i.e., individualist versus solidarist and hierarchical versus egalitarian) used to determine cultural worldviews (Kahan, 2012; Kahan et al., 2009, 2011). Respondents who are more individualistic tend to be committed to the autonomy of markets and those who are more hierarchical tend to view warnings of environmental risks as an affront to the competence of higher-ranked officials or professionals, and those who are either more individualistic or hierarchical are typically less concerned about environmental risks and accepting of climate change action (Bray, 2015; Kahan et al., 2005). Respondents were also asked about the proportion of personal consumption that is bottled versus tap water, and questions that measured subjective performance of bottled versus tap water across the product attributes of health, taste, and sustainability. The subjective performance questions included a response option that allowed respondents to indicate that the waters had equal performance for an attribute.

2.4. Statistical methods

We analyzed the treatment effects in several ways. First, we identified any differences in mean values for bottled water and tap water between treatments by testing the null hypothesis of no difference across treatments. A one-way analysis of variance (ANOVA) was estimated for both types of water (i) to test the null hypothesis:

$$H_{0i}: \mu_{i\_WTPOnly} = \mu_{i\_WTAOnly} = \mu_{i\_WTP/WTAChoice} = \mu_{i\_WTP/WTAScale} \tag{1}$$

ANOVA is often used to determine treatment effects (Kaufmann and Schering, 2014; Stahle and Wold, 1989), and we expect to reject this null hypothesis for both types of water because respondents were limited to non-negative and non-positive values for the WTP Only and WTA Only treatments, respectively. Nevertheless, it is good practice to determine if there is a 'global' difference in groups means prior to making multiple comparisons between various treatments (Kim, 2015). Thus, when a null hypothesis of no difference across all treatments was rejected, post-hoc pairwise comparisons were estimated to test null hypotheses of no difference between various treatments (e.g., $H_{0i}: \mu_{i\_WTPOnly} = \mu_{i\_WTP/WTAChoice}$). Multiple comparisons increase the probability of type I error, which is the probability falsely rejecting a null hypothesis; therefore, we estimated T-tests with Bonferroni corrected P-values to test post-hoc pairwise comparisons between various treatments (Lee and Lee, 2018).

We then completed a similar analysis on the differences in elicited values for bottled water and tap water. This allowed us to test the null hypothesis:

$$H_{0i}: (\mu_{b\_WTPOnly} - \mu_{t\_WTPOnly}) = (\mu_{b\_WTAOnly} - \mu_{t\_WTAOnly}) = (\mu_{b\_WTP/WTAChoice} - \mu_{t\_WTP/WTAChoice}) = (\mu_{b\_WTP/WTAScale} - \mu_{t\_WTP/WTAScale}) \tag{2}$$

where b and t denote elicited values for bottled and tap water, respectively. When eliciting values, examining the relative difference between products can provide additional insight. For example, while the presence of hypothetical bias is well documented, particularly for contingent valuation (Carson, 1997), it has been shown that marginal WTP between products is generally the same for hypothetical and non-hypothetical settings (Lusk and Schroeder, 2004). Thus, it is plausible that differences observed in stated values for both types of water across treatments is diminished when examining the difference in values for bottled and tap water.

To further explore the effect of negative WTP, we extracted positive values in the WTP/WTA Choice and WTP/WTA Scale treatments to compare with values in the WTP Only treatment and extracted negative values to compare with values in the WTA Only treatment. This allowed us to compare mean WTP values in the Choice and Scale treatments with mean WTP values in the WTP Only treatment and WTA values with the WTA Only treatment, and test the null hypotheses for both types of water:

$$H_{0i}: \mu_{i\_WTPOnly} = \mu_{i\_WTPChoice} = \mu_{i\_WTPScale} \tag{3}$$

and

$$H_{0i}: \mu_{i\_WTAOnly} = \mu_{i\_WTAChoice} = \mu_{i\_WTAScale} \tag{4}$$

Respondents who reported values of $0 in the scale treatment presented a problem since it was not clear whether those values should be compared to the WTP Only or WTA Only treatment values. Thus, we included responses of $0 in the scale treatment in both tests. Again, we used one-way ANOVA estimates to test the null hypotheses of no difference in means across treatments and used T-tests with Bonferroni corrected P-values when the null hypothesis was rejected.

Next, we explored heterogeneity across respondents and elicited

[1] We also acknowledge that some individuals consider tap water to be unsafe and therefore may have a negative value for it, either to protect their own health or perhaps to protect the health of others.

Table 2
Probability of choosing WTA and probability of stating a negative value.

	Prob of Choosing WTA		Prob of Stating a Negative Value			
	WTP/WTA Choice		WTP/WTA Choice	WTP/WTA Scale	WTP/WTA Choice	WTP/WTA Scale
	Bottled	Tap	Bottled		Tap	
Age	0.09	−0.04	0.19*	0.15	−0.04	0.08
	(0.09)	(0.09)	(0.10)	(0.16)	(0.09)	(0.08)
Education	−0.15	−0.11	−0.23*	−0.01	−0.16*	−0.13
	(0.11)	(0.10)	(0.12)	(0.21)	(0.10)	(0.11)
Female	−0.26	0.21	−0.28	0.08	0.29	0.08
	(0.31)	(0.27)	(0.32)	(0.52)	(0.27)	(0.29)
Income	−0.11	−0.27***	−0.16	−0.25	−0.24***	−0.02
	(0.11)	(0.09)	(0.11)	(0.19)	(0.09)	(0.09)
Urban	−0.79*	0.16	−0.93**	1.64**	0.14	−0.50
	(0.44)	(0.38)	(0.45)	(0.73)	(0.39)	(0.43)
Suburban	−0.32	0.36	−0.53	0.81	0.24	−0.42
	(0.36)	(0.33)	(0.37)	(0.69)	(0.33)	(0.35)
Employed	−0.24	−0.11	−0.04	−0.46	−0.19	−0.35
	(0.32)	(0.29)	(0.33)	(0.59)	(0.30)	(0.31)
White	−0.64	0.12	−0.78	−1.29*	0.25	−0.36
	(0.47)	(0.43)	(0.49)	(0.75)	(0.43)	(0.42)
Black	0.34	0.35	0.15	0.00	0.31	−0.35
	(0.69)	(0.70)	(0.73)	(omitted)	(0.69)	(0.63)
Hispanic	−0.60	0.10	−1.37	0.00	0.07	0.00
	(0.95)	(0.76)	(1.19)	(omitted)	(0.74)	(omitted)
Individualistic	0.01	0.05*	0.00	0.06	0.06**	−0.01
	(0.03)	(0.03)	(0.03)	(0.06)	(0.03)	(0.03)
Hierarchical	−0.02	−0.02	−0.03	0.01	−0.06***	0.02
	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)	(0.02)
Bottled Water Consumption	−0.02***	0.01***	−0.02***	−0.01	0.01***	0.02***
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
Bottled Water Healthier	−0.75*	0.65*	−1.03**	−0.91	0.62*	0.89**
	(0.39)	(0.36)	(0.41)	(0.81)	(0.35)	(0.38)
Tap Water Healthier	0.10	−0.49	−0.28	1.14	−0.28	0.15
	(0.50)	(0.49)	(0.52)	(0.76)	(0.52)	(0.67)
Bottled Water Tastier	−0.42	0.44	−0.42	−0.07	0.63	0.35
	(0.41)	(0.40)	(0.43)	(0.84)	(0.41)	(0.53)
Tap Water Tastier	0.21	0.68	−0.27	1.29	0.53	0.35
	(0.52)	(0.51)	(0.54)	(0.85)	(0.53)	(0.68)
Bottled Water More Sustainable	0.02	−1.19***	0.05	−0.37	−0.97***	−0.18
	(0.44)	(0.38)	(0.47)	(0.98)	(0.37)	(0.38)
Tap Water More Sustainable	0.01	−0.49	−0.02	−0.55	−0.49	−0.13
	(0.37)	(0.34)	(0.39)	(0.65)	(0.33)	(0.36)
Constant	2.37**	−0.05	2.97***	−3.29	−0.08	−1.94
	(1.04)	(0.92)	(1.10)	(2.12)	(0.91)	(1.20)
n	337	337	337	322	337	348
Log-Likelihood	−157	−193	−144	−61	−190	−169

Note: These are estimated coefficients from binary logit models and standard errors are in parentheses. *, **, and *** denote significance level at 0.10, 0.05, and 0.01, respectively. There were 35 observations omitted in the Bottled – WTP/WTA Scale model because none of the 26 Black or 9 Hispanic respondents stated a negative value, and 9 observations were omitted in the Tap – WTP/WTA Scale model because none of the Hispanic respondents stated a negative value.

values by estimating models that allowed us to determine characteristics associated with the: 1) probability of choosing WTA in the WTP/WTA Choice treatment, 2) probability of stating a negative value in the WTP/WTA Choice and WTP/WTA Scale treatments, 3) probability of stating $0 in all treatments, and 4) magnitude of non-zero WTP and WTA values for both types of water.

Characteristics associated with the probability of choosing WTA in the WTP/WTA Choice treatment were examined by estimating binary logit models that can be mathematically defined by:

$$Prob_{ij}(ChoseWTAinWTP\_WTA\_Choice) = \alpha_{1i} + \boldsymbol{X}_j\boldsymbol{\beta} \tag{5}$$

where $\boldsymbol{X}_j$ is a vector of characteristics for respondent j, $\boldsymbol{\beta}$ are coefficients to be estimated for respondent characteristics, and α_{1i} is a constant to be estimated. Respondent characteristics included in estimation were age, education, an indicator variable for females, income, indicator variables for urban and suburban, an indicator variable for respondents employed full or part time, indicator variables for White, Black, and Hispanic respondents, Individualistic and Hierarchical score from the Cultural Cognition Quiz, personal consumption that is bottled versus tap water, and indicator variables for subjective performance of bottled and tap water across the product attributes of health, taste, and sustainability with 'equally' used as a base.

Conducting the exact same analysis for the WTP/WTA Choice treatment was impossible because, again, it was not clear what to do with $0 values. Thus, we also estimated binary logit models to examine characteristics associated with the probability of stating a negative value in the WTP/WTA Choice and WTP/WTA Scale treatments (k) that can be mathematically defined by:

$$Prob_{ijk}(Stating\ \ a\ \ negative\ \ value) = \alpha_{2ik} + \boldsymbol{X}_j\boldsymbol{\beta} \tag{6}$$

where the same independent variables described in the previous paragraph were also used in these models.

Lastly, we examined characteristics associated heterogeneity across the magnitude of elicited values and the nature of this data allows for a unique analysis. We first estimated binary logit models to determine characteristics associated with the probability of stating $0 in all treatments that can be mathematically defined by:

$$Prob_{ij}(Stating\ \$0) = \alpha_{3i} + \beta_1 WTA\_Only + \beta_2 WTP\_WTA\_Choice + \beta_3 WTP\_WTA\_Scale + \boldsymbol{X}_j\boldsymbol{\beta} \tag{7}$$

Table 3
Probability of stating $0 and censored non-zero WTP and WTA.

	Prob of Stating a $0 WTP/WTA		Bottled		Tap	
	Bottled	Tap	Nonzero WTP	Nonzero WTA	Nonzero WTP	Nonzero WTA
WTA Only	−2.67***	−0.78***				
	(0.32)	(0.18)				
WTP/WTA Choice	−0.10	−1.28***	−0.01	0.07	0.17	0.08
	(0.42)	(0.18)	(0.07)	(0.14)	(0.11)	(0.19)
WTP/WTA Scale	0.01	−0.07	0.06	0.42*	0.21**	0.06
	(0.42)	(0.19)	(0.06)	(0.25)	(0.09)	(0.18)
Age	0.05	−0.01	−0.13***	−0.07*	−0.06**	0.02
	(0.07)	(0.04)	(0.02)	(0.04)	(0.03)	(0.05)
Education	0.00	0.04	0.01	−0.02	−0.01	−0.06
	(0.08)	(0.05)	(0.02)	(0.05)	(0.03)	(0.06)
Female	−0.23	−0.21*	0.01	−0.06	−0.10	0.29**
	(0.21)	(0.13)	(0.06)	(0.12)	(0.08)	(0.15)
Income	−0.08	−0.01	0.11***	0.04	0.05**	−0.11**
	(0.07)	(0.04)	(0.02)	(0.04)	(0.03)	(0.05)
Urban	−0.16	0.20	0.35***	0.32*	0.52***	0.22
	(0.31)	(0.19)	(0.09)	(0.17)	(0.12)	(0.21)
Suburban	−0.12	−0.05	−0.05	0.05	0.00	0.17
	(0.26)	(0.16)	(0.07)	(0.15)	(0.10)	(0.18)
Employed	0.16	−0.08	0.01	−0.24*	−0.01	−0.09
	(0.23)	(0.14)	(0.06)	(0.14)	(0.09)	(0.16)
White	−0.44	−0.01	0.22**	−0.13	0.07	0.47**
	(0.36)	(0.21)	(0.09)	(0.19)	(0.13)	(0.22)
Black	0.36	−0.43	0.25*	0.31	0.06	0.42
	(0.60)	(0.29)	(0.13)	(0.27)	(0.21)	(0.32)
Hispanic	−0.74	0.02	0.24	0.00	0.34	0.09
	(0.63)	(0.40)	(0.17)	(0.40)	(0.23)	(0.49)
Individualistic	−0.03	−0.04***	−0.02***	−0.01	−0.01	0.01
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Hierarchical	0.00	0.02**	0.00	−0.01	0.01**	−0.01
	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.02)
Bottled Water Consumption	0.00	0.00	0.00	0.00	0.00	−0.01
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Bottled Water Healthier	0.38	−0.10	0.18**	0.07	−0.06	0.42**
	(0.28)	(0.17)	(0.08)	(0.18)	(0.11)	(0.20)
Tap Water Healthier	−0.16	0.31	0.07	0.40*	0.20	0.38
	(0.35)	(0.24)	(0.11)	(0.21)	(0.14)	(0.31)
Bottled Water Tastier	−0.41	−0.23	0.01	−0.10	−0.09	−0.05
	(0.30)	(0.20)	(0.09)	(0.19)	(0.12)	(0.24)
Tap Water Tastier	0.41	−0.25	0.11	0.11	0.32**	0.29
	(0.38)	(0.25)	(0.12)	(0.21)	(0.15)	(0.30)
Bottled Water More Sustainable	0.74**	0.49***	0.34***	0.19	0.35***	0.07
	(0.34)	(0.19)	(0.08)	(0.19)	(0.12)	(0.20)
Tap Water More Sustainable	−0.13	−0.12	−0.16**	−0.18	−0.27***	−0.42**
	(0.25)	(0.16)	(0.07)	(0.16)	(0.10)	(0.19)
Constant	4.36***	2.30***	1.64***	2.25***	0.77**	1.40**
	(0.87)	(0.50)	(0.22)	(0.48)	(0.31)	(0.60)
n	1,384	1,384	910	335	586	385
Left-censored at 0			0	0	0	0
Right-censored at 4			37	21	13	51
Log-Likelihood	−344	−786	−1,115	−492	−789	−631

where *WTA_Only*, *WTP_WTA_Choice*, and *WTP_WTA_Scale* are indicator variables equal to one if a respondent was randomized to that treatment. WTP Only is the base, so β_1, β_2, and β_3 indicate if the probability of stating a $0 value for either type of water was affected by random treatment assignment. Respondent characteristics were also included as independent variables. This first step can be thought of as a 'hurdle' or likelihood of participation (Wei, Guan, and Zhu, 2016).

Respondents who cleared the zero-value hurdle were then separated to examine characteristics associated with heterogeneity in the magnitude of WTP (positive stated-values) and WTA (negative stated-values). Recall that the WTP values were bounded between $0 and $4 for WTP Only and for respondents who chose WTP in the WTP/WTA Choice treatment, -$4 and $0 for WTA Only and for respondents who chose WTA in the WTP/WTA Choice treatment, and -$4 and $4 for WTP/WTA Scale. We then estimated Tobit models, which are used for contingent valuation data (Whitehead and Haab, 2001), that can be mathematically defined by:

$$WTP_i^* = \alpha_{4i} + \beta_4 WTP\_WTA\_Choice + \beta_5 WTP\_WTA\_Scale + \boldsymbol{X}_j\boldsymbol{\beta} \tag{8}$$

and

$$-(WTA_i^*) = \alpha_{5i} + \beta_6 WTP\_WTA\_Choice + \beta_7 WTP\_WTA\_Scale + \boldsymbol{X}_j\boldsymbol{\beta} \tag{9}$$

where WTP_i^* and WTA_i^* are latent values. The negative sign before WTA_i^* allows us to compare estimated coefficients more easily across models, so that an increase in an independent variable with a positive sign is associated with an increase in WTP or WTA. The independent variables are the same as described in Eq. (7); however, the *WTA_Only* indicator variable has been removed because only positive values were included in Eq. (8) and it is the base for Eq. (9) which excludes the WTP Only treatment. The latent values were left-censored at $0 and right-censored at $4. Although, we should not have any left-censored observations in these models; recall that $0 did not clear the hurdle.

All statistical analyses were performed using Stata/IC 16.1.

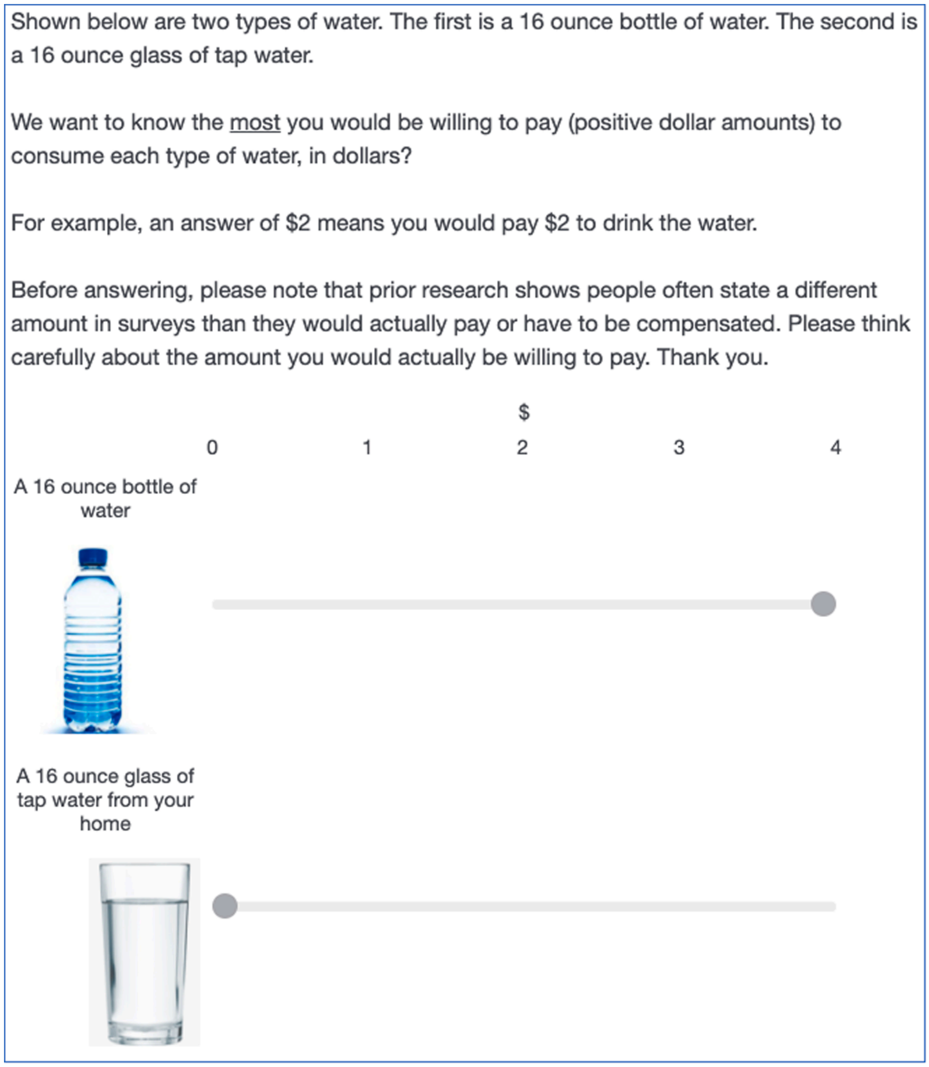

Fig. A1. Screenshot of WTP Only Treatment Questions.

2.5. Respondent characteristics

For the sample, median age was 47.43 years, median income was $71,941.46, and 50.69% of the respondents were female. These demographic proportions are consistent with quotas set to ensure that the sample matched the characteristics of the U.S. population overall. Further description of the sample's demographic characteristics, by contingent valuation treatment, is presented in Table 1 along with mean scores for the Cultural Cognition Quiz, personal consumption, and subjective performance of attributes. Randomization across treatments was checked by estimating a multivariate analysis of variance (MANOVA) using and all the covariates in Table 1, and the null hypothesis off a 'global' difference between treatments was not rejected [Wilks' Lambda F(3, 1,380) = 0.984, P-value < = 0.998].

3. Results

The mean values elicited for bottled water and tap water by treatment are illustrated in Fig. 1 and presented in Appendix Table A1. For bottled water, the mean values are $1.47 for WTP Only, -$1.05 for WTA Only, $0.77 for WTP/WTA Choice, and $1.29 for WTP/WTA Scale. We find an overall significant difference in these means [F(3, 1,380) = 282.26, P-value < 0.01], and the pairwise comparisons are significantly different (P-value < 0.01) for all but the WTP Only and WTP/WTA Scale comparison (P-value = 0.353). For tap water, the mean values by treatment are $0.59 for WTP Only, -$1.08 for WTA Only, -$0.50 for WTP/WTA Choice, and $0.09 for WTP/WTA Scale. These means are also significantly different overall [F(3, 1,380) = 93.33, P-value < 0.01], as are all the pairwise comparisons (P-value < 0.01). Approximately 28% and 59% of the 337 respondents randomized to the WTP/WTA Choice treatment selected WTA Only, instead of WTP only, for bottled water and tap water, respectively. The proportion of respondents choosing a negative value in the WTP/WTA Scale treatment was lower, with approximately 6% and 25% of the 357 respondents selecting a value below $0 for bottled water and tap water, respectively. These results confirm that WTP/WTA values vary by elicitation method, and the fact that the WTP Only values are significantly greater for every treatment except WTP/WTA Scale for bottled water shows that preventing respondents from providing negative WTP likely upwardly biases WTP estimates.

As shown in Fig. 2 and Appendix Table A2, however, some nuances are observed when analyzing differences in the values elicited for (WTP/WTA for bottled water minus WTP/WTA for tap water. While the test determined a significant difference between elicitation treatments overall [F(3, 1,230) = 41.47, P-value < 0.01], there were no significant differences in elicited values for bottled water versus tap water using the WTP/WTA Choice and WTP/WTA Scale treatments (P-value = 1.0). Moreover, the difference in value for bottled water and tap water in the WTP Only treatment was less than the values elicited using WTP/WTA Choice (P-value = 0.009) and WTP/WTA Scale (P-value = 0.047). The

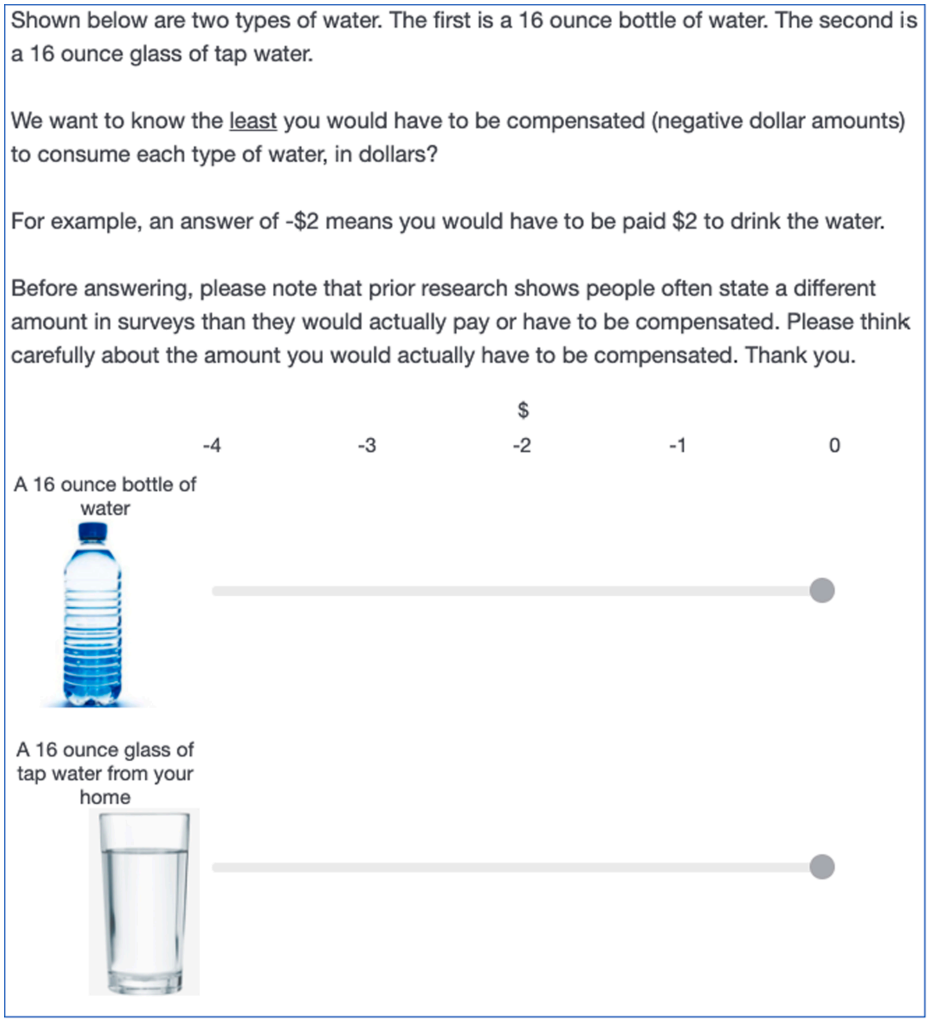

Fig. A2. Screenshot of WTA Only Treatment Questions.

Would you be willing to pay to drink bottled water or would you require compensation to drink bottled water?

- I would pay
- I would require compensation

Would you be willing to pay to drink tap water or would you require compensation to drink tap water?

- I would pay
- I would require compensation

Fig. A3. Screenshot of WTP/WTA Choice Treatment's Initial Questions.

fact that the difference in values for the two types of water is significantly greater when positive and negative WTP are allowed reinforces the importance of allowing negative WTP when comparing multiple goods.

Fig. 3 and Appendix Table A3 compare the means of positive values elicited from respondents in the WTP/WTA Choice and WTP/WTA Scale treatments to the mean of positive values elicited in the WTP Only treatment. For bottled water, we find no significant differences in elicited positive values across the treatments or for the $0 value in the for WTP/WTA Scale treatment [F(2, 932) = 0.75, P-value = 0.475]. However, there is a significant difference in the mean value for tap water under the WTP/WTA Choice treatment [F(2, 762) = 4.32, P-value = 0.014] versus the mean value for tap water under the WTP Only treatment (P-value = 0.012).

Interestingly, the results of comparing negative elicited values are similar. As shown in Fig. 4 and Appendix Table A4, there is no significant difference in mean WTA values elicited for bottled water [F(2, 458) = 2.02, P-value = 0.135]. However, there is a significant difference in mean values elicited for tap water [F(2, 691) = 7.03, P-value < 0.01]. As with the positive elicited values, the mean negative elicited value for tap

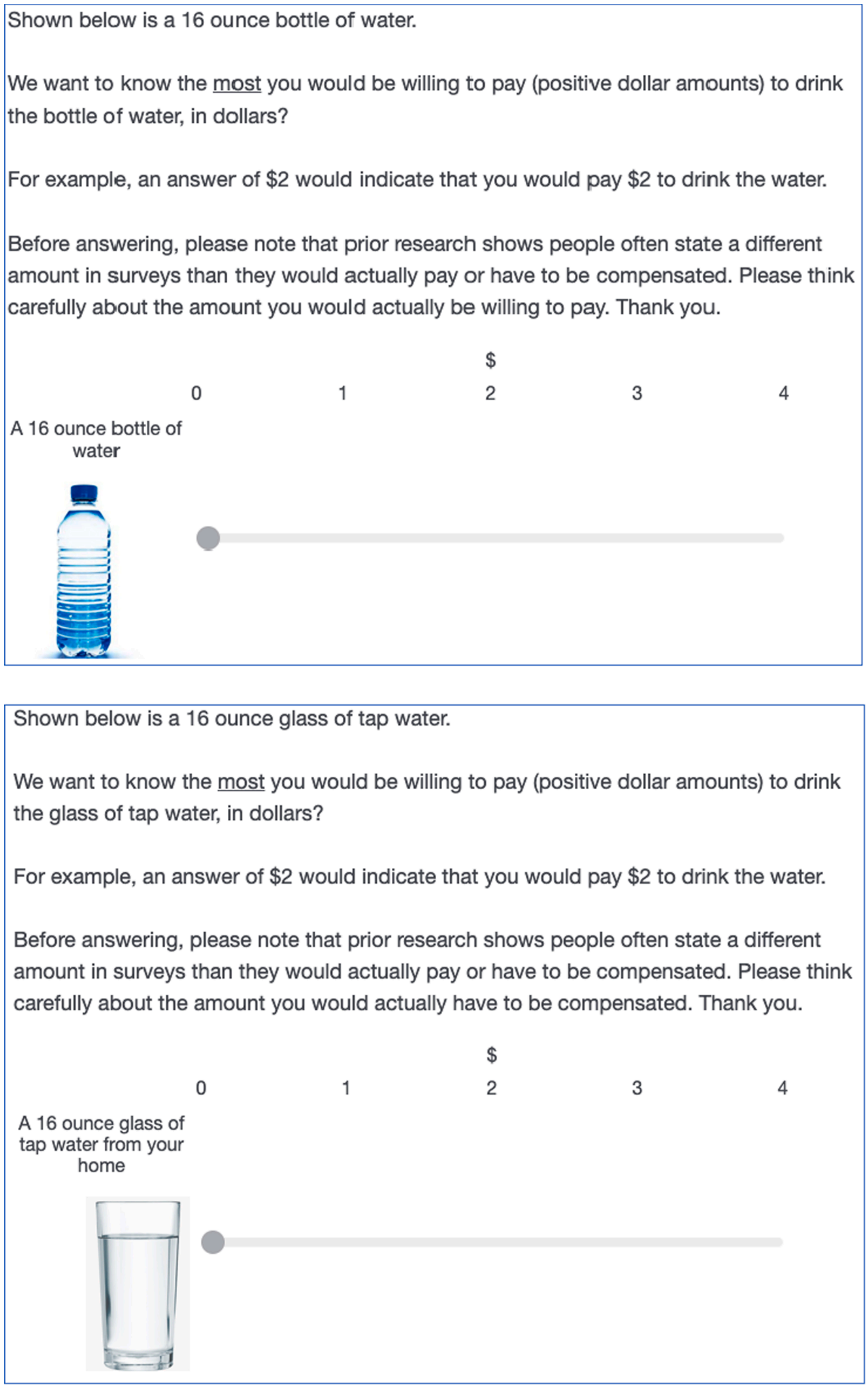

Fig. A4. Screenshot of WTP/WTA Choice Treatment's Follow-Up WTP Questions.

water under the WTP/WTA Choice treatment is more extreme than the mean value under the WTA Only treatment (P-value = 0.002). The scale-treatment mean is not as negative as the WTP/WTA Choice treatment mean or the WTA Only mean (P-values = 0.005). These results suggest that respondents, when given the option to report negative values, will report greater positive values on average than when they can only report positive WTP. Along the same lines, their negative values in the WTP/WTA Choice treatment will be more negative than their WTA Only values.

Table 2 presents the estimated coefficients for Eqs. (5) and (6). Bottled Water Consumption is significant in all columns, except for stating a negative value in the WTP/WTA scale treatment. A similar trend is noticed for Bottled Water Healthier; although, the association is weaker. The probability of choosing WTA for bottled water decreases as the proportion of bottled water consumed increases, and the opposite is true for the probability of choosing WTA for tap water. Also, the probability of choosing WTA for tap water and the probability of stating a negative value for tap water in the WTP/WTA Choice treatment both

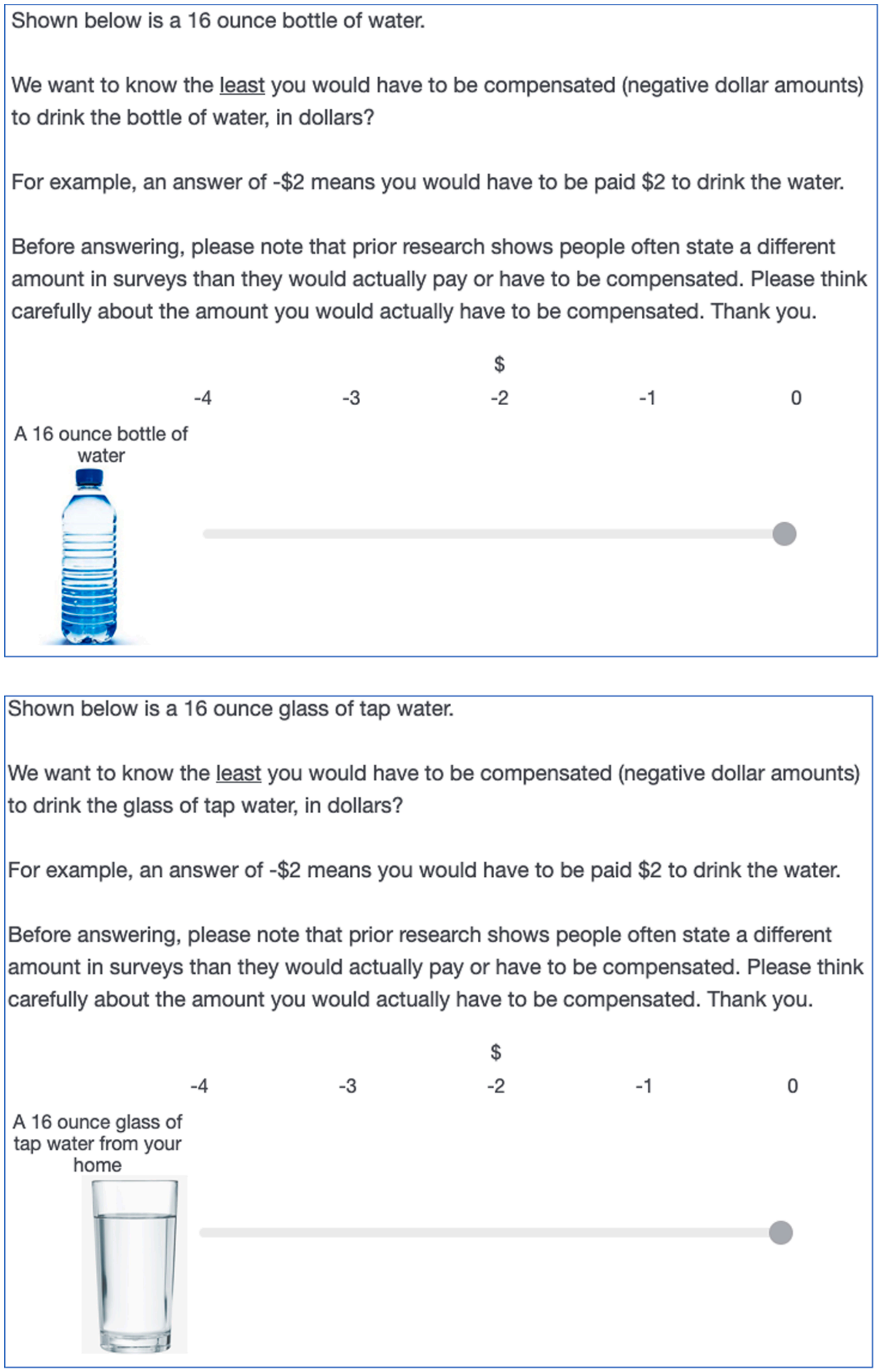

Fig. A5. Screenshot of WTP/WTA Choice Treatment's Follow-Up WTA Questions.

decrease as income increases. Respondents who perceived bottled water to be more sustainable were also less likely to choose WTA for tap water, which is counterintuitive, and a similar relationship is observed for the probability of stating a negative value for tap water in the WTP/WTA Choice treatment. Urban respondents in the WTP/WTA Choice treatment were less likely to state a negative value for bottled water in the WTP/WTA Choice treatment; however, an opposite relationship is observed for bottled water in the WTP/WTA Scale treatment. Additionally, the signs of estimated coefficients for Individualistic and Hierarchical variables are in different directions for the probability of stating a negative value for tap water in the WTP/WTA Choice treatment. Intuitively, these signs should both be positive as respondents who are either more individualistic or hierarchical are thought to be less concerned about environmental issues (Bray, 2015; Kahan et al., 2005). Although, the negative sign for hierarchical could indicate a trust in the source of tap water.

It is important to note that these results should be interpreted with care. As shown in Table 2, Hispanic was omitted in two models and Black was omitted in one model because there were no observations. Although an experiment may be designed with enough power to identify treatment effects, that does not indicate that there is enough power to identify heterogeneity across groups of individuals within treatments.

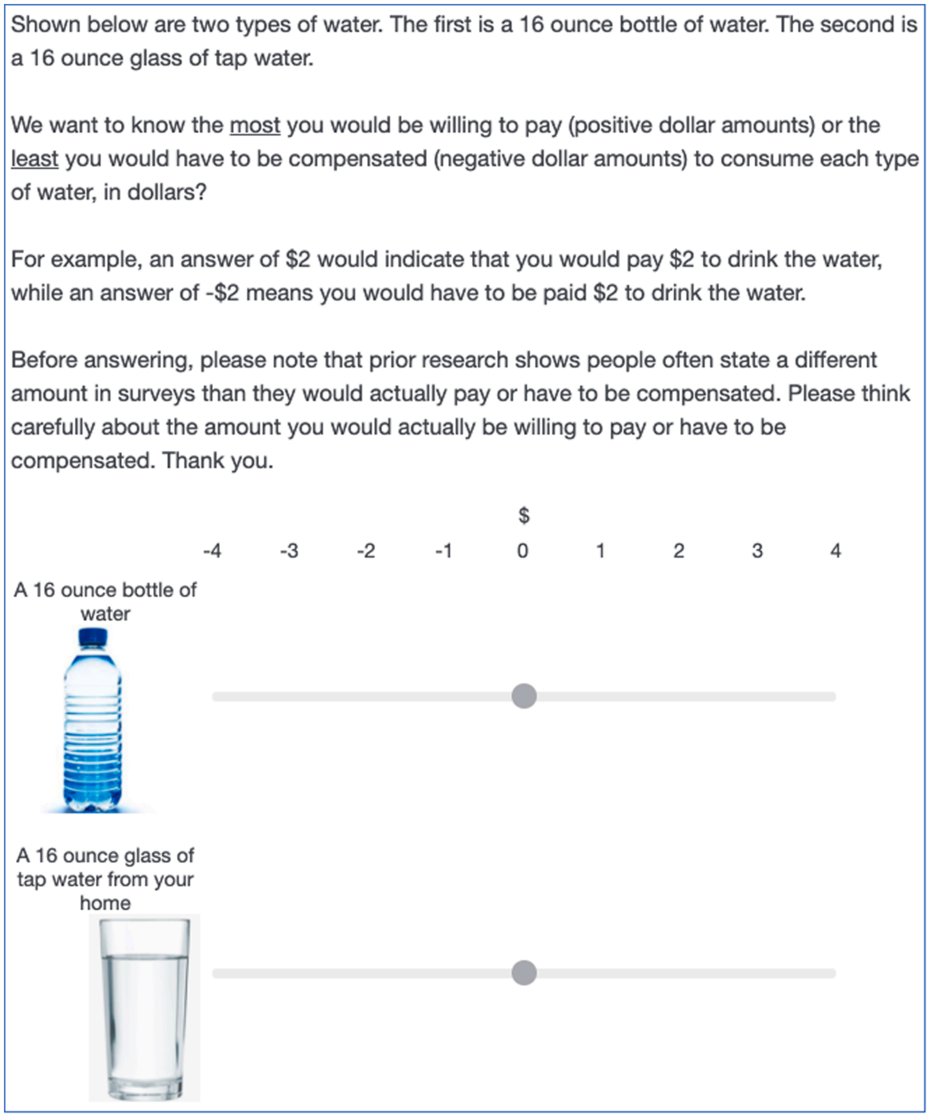

Fig. A6. Screenshot of WTP/WTA Scale Treatment Questions.

Table A1
One-way ANOVA results separately comparing WTP/WTA treatments.

		Bottled			Tap		
Variable	n	Mean	95% CI	Within Grouping	Mean	95% CI	Within Grouping
WTP Only	358	1.47	(1.38, 1.57)	A	0.59	(0.50, 0.68)	A
WTP/WTA Scale	357	1.29	(1.16, 1.43)	A	0.09	(−0.07, 0.25)	B
WTP/WTA Choice	337	0.77	(0.60, 0.95)	B	−0.50	(−0.68, −0.32)	C
WTA Only	332	−1.05	(−1.17, −0.93)	C	−1.08	(−1.22, −0.94)	D

Note: Results for the one-way ANOVA show a significant difference in WTP/WTA values across treatments for bottled and tap water. Bottled: [F(3, 1,380) = 282.26, P-value < 0.01]; tap: [F(3, 1,380) = 93.33, P-value < 0.01]. Pairwise comparisons of means between each treatment conducted using T-tests with Bonferroni corrected P-values found that the treatments for each type of water were all significantly different (P-value < 0.01) except WTP Only compared to WTP/WTA Scale for bottled water (P-value = 0.353).

Moreover, recall that only 6% of respondents randomized to the WTP/WTA Scale treatment selected WTA Only, and this is also communicated by relatively lower Log-Likelihood in Table 2.

Table 3 presents the estimated coefficients for Eqs. (7)–(9). Compared to the WTP only treatment, respondents in the WTA Only treatment were less likely to state a $0 value for both types of water. Respondents in the WTP/WTA Choice treatment were also less likely to state a $0 value for tap water. Females and respondents who are more individualistic were less likely to state a $0 value for tap water. Again, Individualistic and Hierarchical have opposite signs which further supports our speculation that more hierarchical respondents may also be more trusting in government bodies that supply tap water and thus the mechanism of choice may have nothing to do with the environment. Perceiving bottled water to be more sustainable was associated with being more likely to state a $0 value for both types of water, which could represent an overall higher likelihood of stating a $0 value. Now we will discuss the coefficients for the Tobit models; however, it is important to keep the results from these 'hurdles' in mind while interpreting the Tobit

Table A2
One-way ANOVA results separately comparing the difference between WTP/WTA values for bottled minus tap by treatment.

Variable	n	Mean	95% CI	Within Grouping
WTP/WTA Choice	337	1.28	(1.07, 1.48)	A
WTP/WTA Scale	357	1.21	(1.00, 1.41)	A
WTP Only	358	0.88	(0.80, 0.96)	B
WTA Only	332	0.03	(−0.15, 0.21)	C

Note: Results for the one-way ANOVA show a significant difference in values between bottled and tap water across treatments. Bottled-Tap: [F(3, 1,380) = 41.47, P-value < 0.01]. All tests are significantly different (P-value < 0.01) except WTP Only compared to WTP/WTA Scale (P-value = 0.046) and WTP/WTA Choice compared to WTP/WTA Scale (P-value = 1.0).

models.

For bottled water, respondents in the WTP/WTA Scale treatment stated higher WTA values compared to those in the WTA Only treatment, and they also stated higher WTP values for tap water compared to those in the WTP Only treatment. Older respondents stated lower WTP values for both types of water, while respondents with higher incomes and who live in an urban community stated higher WTP values for both types of water. Also, perceiving bottled water to be more sustainable was associated with higher WTP values for both types of water, and perceiving tap water to be more sustainable was associated with lower WTP values for both types of water and a lower WTA for tap water. While some of the signs of these coefficients may seem nonintuitive, e.g., respondents who perceive bottled water to be more sustainable stating higher values for tap water than respondents who perceive tap water to be more sustainable, the signs are consistent across models. The observation that some groups of respondents consistently state higher or lower values may reflect heterogeneity in reference dependence (Caputo et al., 2020.) White respondents and those who perceived bottled water was healthier stated higher WTP values for bottled water and higher WTA values for tap water. Respondents who perceived tap water to taste better stated higher WTP values for tap water. Higher individualistic scores were associated with lower WTP values for bottled water, which is consistent with claims that these individuals have less concern about environmental issues. Respondents with higher Hierarchical scores stated higher WTP values for tap water, and again seem to be relatively supportive of tap water.

4. Policy implications

Welfare economics has used potential Pareto improvements as a measure of efficiency (Gowdy, 2005), and valuation studies often determine social efficiency using a cost-benefit framework to estimate whether gains are sufficient to compensate losses (Nguyen, Knetsch, and Mahasuweerachai, 2021). However, assumptions and decisions made when designing valuation methods to determine potential Pareto improvements may result in biased empirical estimates of welfare changes (Gowdy, 2004). For example, deciding the elicitation method, either WTP or WTA, can bias welfare estimates (Nguyen, Knetsch, and Mahasuweerachai, 2021). Moreover, WTP and WTA may not equal compensating variation and equivalent variation, used by Hicksian welfare theory, to estimate welfare effects, and the divergence may be partially caused by endogenous preferences (Kim, Kling, and Zhao, 2015). Relaxing the assumption that WTP is nonnegative may reduce biases in estimating gains and losses and allow for endogenous preferences like intrinsic values; thus, reducing biases in welfare estimates. If some individuals truly have negative WTP, possibly even a relatively small number of individuals depending on the population size, then mean welfare estimates derived from WTP bounded at $0 will be upwardly bias, while median welfare estimates will be less sensitive to the assumption of non-negative WTP and could be equal to welfare estimates that allow negative WTP (Hanemann, 1989).

Contingent valuation studies have found evidence of negative WTP for non-market goods (Campbell et al., 2014; Lundhede et al., 2013; Jacobsen et al., 2011; Hanley et al., 2009; Lockwood et al., 1994). However, WTP for market goods in these studies typically has had a lower bound of zero and therefore has been forced to be non-negative (Kriström, 1997). Bounding WTP at zero does not account for intrinsic values such as personal health concerns and valuations attached to branding. For example, a vegan would have to be compensated to eat meat and a brand valued for being environmentally friendly could be damaged if applied to bottled water. In these examples, consumers would require compensation for their losses of intrinsic value. Therefore, omitting negative WTP implies effective demand at price $0 that does not necessarily exist, leading to upwardly biased WTP estimates. We find consistent evidence that this is the case.

The benefit of allowing positive and negative WTP in an elicitation is further exemplified by the results of our comparison of values for bottled water and tap water by treatment. The differences in the WTP/WTA

Table A3
One-way ANOVA results separately comparing WTP treatments.

		Bottled				Tap		
Variable	n	Mean	95% CI	Within Grouping	n	Mean	95% CI	Within Grouping
WTP Choice	243	1.57	(1.45, 1.69)	A	138	0.87	(0.69, 1.06)	A
WTP Scale	334	1.52	(1.41, 1.62)	A	269	0.71	(0.59, 0.84)	AB
WTP Only	358	1.47	(1.38 1.57)	A	358	0.59	(0.50, 0.68)	B

Note: Results for the one-way ANOVA show no significant difference in WTP values across treatments for bottled water [F(2, 932) = 0.75, P-value = 0.475] but a significant difference for tap water [F(2, 762) = 4.32, P-value = 0.014]. Pairwise comparisons of means between each treatment conducted using T-tests with Bonferroni corrected P-values found that, for tap water, WTP Only and WTP Choice were significantly different (P-value = 0.012).

Table A4
One-way ANOVA results separately comparing WTA treatments.

		Bottled				Tap		
Variable	n	Mean	95% CI	Within Grouping	n	Mean	95% CI	Within Grouping
WTA Choice	94	−1.29	(−1.53, −1.05)	A	199	−1.46	(−1.64, −1.27)	A
WTA Scale	35	−1.27	(−1.74, −0.80)	A	163	−0.98	(−1.19, −0.77)	B
WTA Only	332	−1.05	(−1.17–0.93)	A	332	−1.08	(−1.22, −0.94)	B

Note: Results for the one-way ANOVA show no significant difference in WTA values for bottled water [F(2, 458) = 2.02, P-value = 0.135] but a significant difference for tap water: F(2, 691) = 7.03, P-value = 0.001]. Pairwise comparisons of means between each treatment conducted using T-tests with Bonferroni corrected P-values found that, for tap water, WTA Only was significantly different than WTA Choice (P-value = 0.005) and WTA Choice was significantly different than WTA Scale (P-value = 0.02).

Choice and WTP/WTA Scale treatments are significantly greater than the differences in the WTP Only treatment. These results show that allowing negative values can reveal the true magnitude of premiums associated with a preference for bottled water over tap water. These magnitudes can be especially important when comparing goods that are relatively similar in nature. Not only does allowing negative WTP help identify greater price discrepancies that can exist between goods, but it can also provide important information to guide policy by measuring the intensity of consumer concern about the negative externalities associated with the good.

5. Conclusions

In this study involving 1384 adult consumers, we test four WTP/WTA elicitation approaches to determine how each approach affects estimates of consumer preferences for two types of a nearly identical good, bottled and tap water. The results of our analysis show that elicitation mechanism matters and that WTP Only approaches have significant limitations. We find that stated values for both types of water differ by elicitation treatment and that the differences extend beyond simple WTP Only versus WTA Only comparisons. For bottled water, the values elicited by the WTP Only treatment are significantly different from the values elicited using the WTP/WTA Choice treatment and values elicited for tap water using the WTP/WTA Choice and WTP/WTA Scale treatments. This occurs because the WTP Only treatment does not allow respondents to report negative WTP, forcing them to assign a value equal to or greater than zero. The fact that the results of the WTP Only treatment are significantly different from the results of the other three treatments overall and when comparing positive to positive and negative to negative values provides strong evidence that WTP Only approaches have upwardly biased prior estimates.

In addition to examining all WTP and WTA values together, we separately compared the WTP Only values to the WTA Only values for each type of water and found implications for marketers and policymakers there as well. Our elicitation of WTP values resulted in no significant differences in WTP for bottled water by treatment and a difference only between the WTP Choice and WTP Only treatments for tap water. The same is true for WTA Only values. We found no differences by treatment for bottled water and greater differences in the choice elicitation than in the WTA Only and WTP/WTA Scale elicitations. Thus, we find that the differences in WTP across the treatments come from allowing negative WTP values (i.e., WTA).

In a final step, we compared differences in stated WTP values for bottled water to stated WTP values for tap water and stated WTA values for bottled water to stated WTA values for tap water. In this case, we found no significant differences for the elicitation treatments. Thus, we find that differences in values for two goods elicited using a WTP Only mechanism or a WTA Only mechanism are consistent, but values elicited using WTP for one good and WTA for the other can lead to inaccurate results. That is, treatment matters. Our results clearly identify the benefits of allowing for negative WTP when valuing market goods: greater accuracy when measuring consumer values and differences in consumer value for two goods.

This study has limitations that could be addressed in future research. Our valuation measures are hypothetical; however, the between-treatment design and analysis hopefully minimize the effect of hypothetical bias on valuation estimates. Conducting economic experiments that involve actual purchases of actual products would be a logical next step for future research. Moreover, we use two goods that are substitutes but are generally very different in price. Additional studies comparing goods that are more similar in value than bottled water and tap water could be informative. Also, intrinsic values likely vary by the type of good, and thus so would the effect of elicitation method on valuation and could be another opportunity for fruitful research.

Funding

This research was funded through the U.S. Department of Agriculture National Institute of Food and Agriculture, 201668007250064.

CRediT authorship contribution statement

Daniel A. Bass: Writing - original draft, Methodology, Visualization. **Brandon R. McFadden:** Conceptualization, Data curation, Writing - review & editing, Methodology, Visualization. **Kent D. Messer:** Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Daniel A. Bass and Brandon R. McFadden share senior authorship of this work. A special thanks to the editor, Sven Anders, and four anonymous referees.

Appendix A

See Figs. A1–A6 and Tables A1–A4.

References

Ahlheim, M., Buchholz, W., 2000. WTP or WTA – is that the question? Zeitschrift für Umweltpolitik & Umweltrecht 23 (2), 253–271.
Anadu, E.C., Harding, A.K., 2000. Risk perception and bottled water use. Am. Water Works Assoc. 92 (11), 82–92.
Bohara, A.K., Kerkvliet, J., Berrens, R.P., 2001. Addressing negative willingness to pay in dichotomous choice contingent valuation. Environ. Resour. Econ. 20 (3), 173–195.
Boyce, R.R., Brown, T.C., McClelland, G.H., Peterson, G.L., Schulze, W.D., 1992. An experimental examination of intrinsic values as a source of the WTA-WTP disparity. Am. Econ. Rev. 82 (5), 1366–1373.
Bray, J., 2015. Cultural Cognition and Climate Change: Communicating climate science across potential divides in Sweden. Available at: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A820440&dswid=-5567.
Cameron, T.A., Englin, J., 1997. Respondent experience and contingent valuation of environmental goods. J. Environ. Econ. Manage. 33 (3), 296–313.
Campbell, D., Vedel, S.E., Thorsen, B.J., Jacobsen, J.B., 2014. Heterogeneity in the WTP for recreational access: distributional aspects. J. Environ. Plann. Manage. 57 (8), 1200–1219.
Caputo, V., Lusk, J.L., Nayga, R.M., 2020. Am i getting a good deal? Reference-dependent decision making when the reference price is uncertain. Am. J. Agric. Econ. 102 (1), 132–153.
Carson, Richard T., 1997. Contingent valuation: theoretical advances and empirical tests since the NOAA panel. Am. J. Agric. Econ. 79 (5), 1501–1507.
Haab, T.C., McConnell, K.E., 1998. Referendum models and economic values: theoretical, intuitive, and practical bounds on willingness to pay. Land Econ. 74 (2), 216–229.
Gowdy, J.M., 2004. The revolution in welfare economics and its implications for environmental valuation and policy. Land Econ. 80 (2), 239–257.
Gowdy, J., 2005. Toward a new welfare economics for sustainability. Ecol. Econ. 53 (2), 211–222.
Haab, T.C., McConnell, K.E., 1997. Referendum models and negative willingness to pay: alternative solutions. J. Environ. Econ. Manage. 32 (2), 251–270.
Halstead, J.M., Lindsay, B.E., Brown, C.M., 1991. Use of the Tobit model in contingent valuation: experimental evidence from the Pemigewasset Wilderness Area. J. Environ. Manage. 33 (1), 79–89.
Hanemann, W.M., 1989. Welfare evaluations in contingent valuation experiments with discrete response data: reply. Amer. J. Agricult. Econ. 71 (4), 1057–1061.
Hanemann, W.M., 1991. Willingness to pay and willingness to accept: how much can they differ? Am. Econ. Rev. 81 (3), 635–647.
Hanley, N., Colombo, S., Kriström, B., Watson, F., 2009. Accounting for negative, zero and positive willingness to pay for landscape change in a National Park. J. Agric. Econ. 60 (1), 1–16.
Hensher, D.A., Rose, J.M., Greene, W.H., 2015. Applied Choice Analysis, Second Edi. ed.
Horowitz, J.K., Kenneth, M.E., 2000. A Review of WTA/WTP Studies. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=257336.
Hu, Z., Morton, L.W., Mahler, R.L., 2011. Bottled water: United States consumers and their perceptions of water quality. Int. J. Environ. Res. Public Health 8 (2), 565–578.

Jackman, M., Lorde, T., 2014. Why buy when we can pirate? The role of intentions and willingness to pay in predicting piracy behavior. Int. J. Social Econ.
Jacobsen, J.B., Lundhede, T.H., Martinsen, L., Hasler, B., Thorsen, B.J., 2011. Embedding effects in choice experiment valuations of environmental preservation projects. Ecol. Econ. 70 (6), 1170–1177.
Jakus, P.M., Shaw, W.D., Nguyen, T.N., Walker, M., 2009. Risk perceptions of arsenic in tap water and consumption of bottled water. Water Resour. Res. 45 (5), 2.
Jiang, Y., House, L.A., Kim, H., Percival, S.S., 2017. Zero-inflated ordered probit approach to modeling mushroom consumption in the United States. Int. Food Agribusiness Manage. Rev. 20 (1030–2018-019), 655–672.
Kahan, D.M., Braman, D., Gastil, J., Slovic, P., Mertz, C.K., 2005. Gender, race, and risk perception: the influence of cultural status anxiety. Available at SSRN 723762.
Kahan, Dan M., 2012. Ideology, motivated reasoning, and cognitive reflection: an experimental study. Judgment Decision Making 8, 407–424.
Kahan, Dan M., Braman, Donald, Slovic, Paul, Gastil, John, Cohen, Geoffrey, 2009. Cultural cognition of the risks and benefits of nanotechnology. Nat. Nanotechnol. 4 (2), 87.
Kahan, Dan M., Jenkins-Smith, Hank, Braman, Donald, 2011. Cultural cognition of scientific consensus. J. Risk Res. 14 (2), 147–174.
Kahneman, D., Knetsch, J.L., Thaler, R., 1991. Anomalies: the endowment effect, loss aversion, and status quo bias. J. Econ. Perspect. 5 (1).
Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47 (2).
Kaufmann, J., Schering, A.G., 2014. Analysis of variance ANOVA. Wiley StatsRef: Statistics Reference Online.
Kecinski, M., Kerley, D., Messer, K.D., Schulze, W.D., 2018. Measuring stigma: the behavioral response to drinking cockroach contaminated water. Environ. Resour. Econ. 70 (1), 131–146.
Kim, H.Y., 2015. Statistical notes for clinical researchers: post-hoc multiple comparisons. Restorative Dentistry Endodont. 40 (2), 172.
Kim, Y., Kling, Zhao, J., 2015. Understanding behavioral explanations of the WTP-WTA divergence through a neoclassical lens: Implications for environmental policy. Annu. Rev. Resour. Econ. 7 (1), 169–187.
Knetsch, J.L., 1995. Asymmetric valuation of gains and losses and preference order assumptions. Econ. Inquiry; Huntington Beach 33 (1), 134–141.
Knetsch, J.L., Sinden, J.A., 1984. Willingness to pay and compensation demanded: experimental evidence of an unexpected disparity in measures of value. Q. J. Econ. 99 (3), 507–521.
Kriström, B., 1997. Spike models in contingent valuation. Am. J. Agric. Econ. 79 (3), 1013–1023.
Lee, S., Lee, D.K., 2018. What is the proper way to apply the multiple comparison test? Korean J. Anesthesiol. 71 (5), 353.
Lockwood, M., Loomis, J., De Lacy, T., 1994. The relative unimportance of a nonmarket willingness to pay for timber harvesting. Ecol. Econ. 9 (2), 145–152.
Loomis, J., Ekstrand, E., 1998. Alternative approaches for incorporating respondent uncertainty when estimating willingness to pay: the case of the Mexican Spotted Owl. Ecol. Econ. 27 (1), 29–41.
Lundhede, T., Bille, T., Hasler, B., 2013. Exploring preferences and non-use values for hidden archaeological artefacts: a case from Denmark. Int. J. Cultural Policy 19 (4), 501–530.
Lusk, J.L., Schroeder, T.C., 2004. Are choice experiments incentive compatible? A test with quality differentiated beef steaks. Am. J. Agric. Econ. 86 (2), 467–482.
Lusk, J.L., Roosen, J., Fox, J.A., 2003. Demand for beef from cattle administered growth hormones or fed genetically modified corn: a comparison of consumers in France, Germany, the United Kingdom, and the United States. Am. J. Agric. Econ. 85 (1), 16–29.
McKendree, M.G., Widmar, N.O., Ortega, D.L., Foster, K.A., 2013. Consumer preferences for verified pork-rearing practices in the production of ham products. J. Agric. Resour. Econ. 397–417.
McSpirit, S., Reid, C., 2011. Residents' perceptions of tap water and decisions to purchase bottled water: a survey analysis from the Appalachian, big sandy coal mining Region of West Virginia. Society Nat. Resour. 24 (5), 511–520.
Nguyen, K.T., Knetsch, J.L., Mahasuweerachai, P., 2021. WTP or WTA: A Means of Determining the Appropriate Welfare Measure of Positive and Negative Changes When Preferences are Reference Dependent. Environ. Res. Econ. 78 (4), 615–633.
Oduor, F.O., Raburu, P.O., Mwakubo, S., 2016. Estimation of willingness to pay for conservation of Nyando Wetlands, Kenya: a contingent valuation approach. Adv. Ecol. Environ. Res. 1, 1–16.
Plott, C.R., Zeiler, K., 2005. The willingness to pay – willingness to accept gap, the 'endowment effect', subject misconceptions, and experimental procedures for eliciting valuations. Am. Econ. Rev. 95 (3), 530–545.
Samuelson, W., Zeckhauser, R., 1988. Status quo bias in decision making. J. Risk Uncertainty 1 (1), 7–59.
Saylor, A., Prokopy, L.S., Amberg, S., 2011. What's wrong with the tap? Examining perceptions of tap water and bottled water at Purdue University. Environ. Manage. 48 (3), 588–601.
Shogren, J.F., Shin, S.Y., Hayes, D.J., Kliebenstein, J.B., 1994. Resolving differences in willingness to pay and willingness to accept. Am. Econ. Rev. 84 (1), 255–270.
Stahle, L., Wold, S., 1989. Analysis of variance (ANOVA). Chemometrics Intelligent Lab. Syst. 6 (4), 259–272.
Staples, A.J., Reeling, C.J., Widmar, N.J.O., Lusk, J.L., 2020. Consumer willingness to pay for sustainability attributes in beer: a choice experiment using eco-labels. Agribusiness 36 (4), 591–612.
Thaler, R., 1980. Toward a Positive Theory of Consumer Choice. 22...
Tonsor, G.T., Schroeder, T.C., Fox, J.A., Biere, A., 2005. European preferences for beef steak attributes. J. Agric. Resour. Econ. 367–380.
Wei, X., Guan, Z., Zhu, H., 2016. Farmer's willingness to participate in wetland restoration: a hurdle model approach. Agric. Econ. 47 (6), 719–727.
Whitehead, J.C., Haab, T.C., 2001. Analysis of contingent valuation data from the 1997-98 southeast economic add-on survey data. Available at: https://repository.library.noaa.gov/view/noaa/8532.
Zhao, J., Kling, C.L., 2001. A new explanation for the WTP/WTA disparity. Econ. Lett. 73 (3), 293–300.