A Calibrated Choice Experiment Method

Lauren Chenarides, Carola Grebitus, Jayson L. Lusk, and Iryna Printezis*

February 27, 2022

Abstract: Although choice experiments have emerged as the most popular stated preference method in applied economics, the method is not free from biases related to order and presentation effects. This paper introduces a new preference elicitation method referred to as a calibrated choice experiment, and we explore the ability of the new method to alleviate starting point bias. The new approach utilizes the distribution of preferences from a prior choice experiment to provide real-time feedback to respondents about our best guess of their willingness-to-pay for food attributes, and allows respondents to adjust and calibrate their values. The analysis utilizes data collected in 2017 in two U.S. cities, Phoenix and Detroit, on consumer preferences for local and organic tomatoes sold through supermarkets, urban farms, and farmers markets to establish a prior preference distribution. We re-conduct the survey in May 2020 and implement the calibrated choice experiment. Conventional analysis of the 2020 choice experiment data shows willingness-to-pay is strongly influenced by a starting point: the higher the initial price a respondent encountered, the higher the absolute value of their willingness-to-pay. Despite this bias, we show that when respondents have the opportunity to update their willingness-to-pay when presented with the best-guess, the resulting calibrated willingness-to-pay is much less influenced by the random starting point.

Key words: anchoring, calibrated choice experiment, local food, farmers market, starting point bias, urban agriculture

JEL Codes: Q11, Q51, C83

*Authors are Assistant and Associate Professor, Morrison School of Agribusiness, Arizona State University, Distinguished Professor and Head, Department of Agricultural Economics at Purdue University, and Clinical Assistant Professor, Department of Supply Chain Management, Arizona State University, respectively.

Contact: Jayson Lusk; 403 West State Street; Purdue University, West Lafayette, IN 47907; Phone: 765-494-1903; Email: jlusk@purdue.edu.

Introduction

Discrete choice experiments (CEs) have emerged as the most popular and widely utilized stated preference elicitation method in applied economics research. CEs have been found to exhibit reasonably high levels of external validity, particularly when combined with revealed preference data (e.g., Azevendo, Herriges, and Kling, 2003; Adamowicz, Louviere, and Williams, 1994; Brooks and Lusk, 2010; Chang, Lusk and Norwood, 2009; Hensher, Louviere, and Swait, 1998; Swait and Andrews, 2003). However, critics of stated-preference methods remain (e.g., Hausman, 2012), and there remain concerns about CEs related to hypothetical bias (e.g., Wüpper, Clemm, Wree 2019), presentation and order effects (e.g., Schmiess and Lusk, forthcoming), and, as we discuss in more detail below, starting point biases. Thus, there are continual efforts to improve the internal and external validity of the choice experiment method (e.g., Fang et al., 2020), and questions remain as to how the CE method can be further improved.

As indicated, estimated preferences from CEs have been found to be influenced by starting point biases. Day et al. (2012) conducted an extensive analysis of various order effects in CEs and found them to be pervasive. They suggested providing advanced instructions or disclosures to respondents about the repeated nature of the choice tasks and found that this practice can help mitigate order effects. Dekker, Koster and Brouwer (2014) found CEs suffered from anchoring effects, which were induced by giving people low or high price choice questions at the start of the survey. They found, however, that the effect of anchoring dissipated as people progressed through the choice tasks, which again highlights the role of experience in helping alleviate biases. Ladenburg and Olsen (2006) found similar results, and noted some gender-specific differences in starting point bias in CEs. Several studies have shown that willingness-to-pay estimates obtained from CEs are influenced by price levels used in the experimental design (e.g., Hanley, Adamowicz, and Wright, 2005; Mørkbak, Christensen, and GyrdHansen, 2010; Su et al., 2017) and by reference-point prices provided either by the researcher or the respondent (Caputo, Lusk, and Nayga, 2018, 2020). Thus, there is ample evidence that CEs are prone to various forms of starting-point, anchoring, and reference-point biases.

The purpose of this paper is to propose an extension to the typical CE method, an approach we refer to as a calibrated choice experiment (CCE). We aim to explore the extent to which the CCE can help alleviate starting-point bias that has been identified in previous research. In particular, we explore how the estimated preferences are affected by the price level that respondents were first presented with in the CE (the "first price"). We find that the results from the standard CE suffer from starting point bias. The higher the initial starting price (i.e., the anchor), the less price sensitive is the respondent. Thus, without correction, WTP from the CE tends to be higher in absolute value the higher the starting point anchor. Despite this result, we find the calibrated WTP from the CCE is less affected by the arbitrary anchor, providing evidence of the validity of the approach insofar as its ability to help mitigate this bias.

The CCE approach works as follows. First, an initial version of a CE is conducted with a sample of respondents. Second, Bayesian priors on the distribution preferences are obtained by fitting mixed logit or latent class logit models to this initial CE data. Third, the survey and CE are repeated, either with the same or new respondents.[1] Upon completion of the second CE, the priors obtained from the first CE are updated with the person-specific choices to provide a conditional, posterior estimate of each respondent's preferences. Fourth, the second survey respondents are provided with real-time feedback on the implications of their choices in the form

[1] Practitioners of CEs are accustomed to the idea of conducting two waves of surveys because optimal experimental designs require prior estimates of consumer preferences (e.g., see Scarpa, Campbell, and Hutchinson, 2007). Nonetheless, it should be noted that initial waves of a survey used to obtain priors for experimental designs utilize much smaller sample sizes than are likely needed to implement the CE.

of implied WTP for different attributes or choice probabilities for competing alternatives calculated from the conditional, posterior preference estimates – our "best guess" of respondents' WTP. Finally, respondents are allowed to calibrate their WTP by indicating whether the estimated values are too low or too high.

As is clear from the previous discussion, the proposed CCE makes use of "individual specific" conditional, posterior estimates from mixed logit or latent class models; these estimates have been used in a large number of prior studies (e.g., Hensher, Greene, and Rose, 2006; Hess and Hensher, 2010; Lusk and Briggeman, 2009; Ortega et al., 2020; Scarpa et al., 2013). Previous research has explored the conditions under which one can use the "individual specific" estimates derived from conditional posterior distributions to obtain unbiased and reliable estimates of individuals' true preferences (e.g., Revelt and Train, 200l; Sarrias and Daziano, 2018; Sarrias, 2020; Train, 2009). We approach this problem in a very different way. Rather than using econometric theory or Monte Carlo simulations to determine whether our conditional posterior distributions accurately reflect true preferences, we reveal our "best guess" estimates to respondents and ask them whether the estimates are, in fact, accurate. After all, outside theory or Monte Carlo experiments, true preferences can only be known to the individual who made the choices.

The rest of the paper is organized as follows. The next section provides background which situates the CCE method in the broader literature on preference elicitation methods and the discovered preference hypothesis. The following section describes our methods, including more detailed discussion of the CCE, and our application which relates to consumer preferences for tomatoes bought through different channels (supermarket, farmers market, or urban farm) and with different characteristics such as organic and local.[2] Results are presented in the penultimate section, and the last section concludes.

Background

CEs, and other preference elicitation methods, are often utilized under the standard economic assumption of complete and stable preferences. If preferences are, in fact, stable, they can be utilized to understand how consumers' choices will change in environments that differ from the setting in which they were elicited. Despite the advantage of this assumption, the rise of behavioral economics suggests, rather, that preferences are often malleable and depend on potentially arbitrary details, undermining the potential role of survey- and experimentally-elicited preferences for informing marketing and public policy decisions. Ariely et al. (2003), for example, showed that consumers' willingness-to-pay for items like keyboards, books, and wine was influenced by showing people the last two digits of their social security number – a number that, in principle, should have no relationship to peoples' valuations for these goods. In this study, we introduce a new preference elicitation method that we refer to as a calibrated choice experiment, and we explore the ability of the method to mitigate this sort of starting-point bias. In doing so, we reveal how respondents, when prompted with feedback on their choices, may adjust and calibrate their values to arrive at willingness-to-pay that is more consistent with their true underlying preferences and behaviors.

[2] There is a sizable literature on consumer preferences for local foods and food from farmers markets (e.g., Darby et al., 2008; Grebitus, Lusk, and Nayga, 2013; Meas et al., 2015; Printezis and Grebitus, 2018; Printezis, Grebitus and Hirsch, 2019; Richards et al., 2017; Taylor and Villas-Boas, 2016; Thilmany, Bond, and Bond, 2008; Toler et al., 2009), most suggesting consumers, on average, prefer local food relative to more distant or unlabeled food. It is less clear that consumers prefer farmers markets to grocery stores per se. For example, using purchase diary data, Taylor and Villas-Boas (2016) show grocery stores, and especially superstores, are preferred to farmers markets.

Despite the large number of studies showing the existence of behavioral anomalies in preference elicitation, the approach need not be abandoned if methods can be identified that can yield more consistent preference estimates that are less susceptible to bias. Previous research has shown that knowledge and experience with the product or decision-making environment has been found to reduce many behavioral biases and decision-making errors. For example, List (2003, 2004) showed that the endowment effect, which is an explanation for the gap between willingness-to-pay (WTP) and willingness-to-accept (WTA), declines as people have more experience with the marketplace in question. Relatedly, Plott and Zeiler (2004) found that the WTP-WTA gap dissipates when subjects have extensive training and instructions on the elicitation mechanism. Cherry et al. (2003) showed that subjects exposed to market-like arbitrage subsequently demonstrated more preference consistency in a non-market setting. Cason and Plott (2014) showed that subjects only bid as theoretically predicted in an auction-like mechanism after they had been exposed to prior decision-making errors. Dyer and Kagel (1996) showed that, although the winner's curse in common value auctions was commonly observed in abstract experimental settings, experienced construction managers in their field environment were often able to avoid the winner's curse. There is also some evidence that participating in a non-hypothetical market prior to a contingent valuation exercise might reduce hypothetical bias (Taylor, 1998). Thus, acknowledging the long and rich line of antecedents in the literature that demonstrate behavioral biases in estimates of consumer preference, we propose a technique to partially mitigate commonly accepted challenges in existing preference elicitation methods.

A challenge is that many people are unfamiliar with the goods economists wish to study, and moreover, people are inexperienced with the environment and mechanisms used to elicit preferences. In a searing critique of stated preference methods, Diamond and Hausman (1994) made a fundamental criticism that (p. 63) “the internal consistency problems come from an absence of preferences, not a flaw in survey methodology,” suggesting people need experience with goods in markets to form stable preference. Plott’s (1996) discovered preference hypothesis suggests that more systematic and stable preferences, consistent with economic theory, are formed only after repeated experience and feedback from market environments.

Braga and Starmer (2005) further developed these ideas and they highlight the role of “institutional learning,” in which people learn to avoid errors in a particular decision-making environment through experience and “value learning,” in which repeated experience and feedback allows people to learn about their own preferences. These insights suggest a need for developing new approaches that allow people to learn about their preferences and gain experiences with the decision-making context. As Braga and Starmer (2005) put it, “One interpretation of this evidence is that anomalies are errors in stated preference that can be expected to disappear in environments that foster certain kinds of learning. If this is the right way to interpret anomaly evidence, it suggests quite a different agenda for those seeking to develop methods of preference elicitation with a view to providing reliable data for input to public policy.”

Following this suggestion, prior research has focused on elicitation mechanisms that might create an environment in which people can learn about the elicitation mechanisms and their preferences. For example, Bateman et al. (2008) proposed a “learning design contingent valuation” approach in which people repeated multiple double-bounded dichotomous choice questions. Bateman et al. (2008) found that the starting point bias often observed in double-bounded dichotomous choice questions (e.g., Herriges and Shogren, 1996) disappears as people gain experience with the question format, a finding they argue provides support for the discovered preference hypothesis.

Norwood and Lusk (2011), motivated by the discovered preference hypothesis, proposed a method they referred to as a calibrated auction-conjoint valuation method, which has been subsequently used by researchers such as Avitia et al. (2015) and Lobasenko and Llerena (2017). The method takes the respondent's answers to a series of simple rating and ranking questions to form an estimate of the respondent's WTP, which is then revealed to participants for updating and adjustment. Kovalsky and Lusk (2013) showed that WTP estimates obtained from the calibrated auction-conjoint method are less influenced by the presentation of arbitrary anchors than traditional valuation approaches, suggesting the method has the potential to mitigate some behavioral biases.

One advantage of CEs, in relation to the discovered preference hypothesis is that, as a routine matter of practice, they involve people making repeated choices, which allows subjects to gain experience with the decision-making context and environment. Indeed, prior research utilizing CEs and conjoint analysis has shown that it is often the case that error variance is lower (and thus preference consistency is higher) in later choice tasks as compared to initial choice tasks (Carlsson, Mørkbak, and Olsen, 2012; Czajkowski, Giergiczny, and Greene, 2014; DeShazo and Fermo, 2002; Hu et al. 2006; Lusk, Fields, and Prevatt, 2008), a finding supporting the discovered preference hypothesis.[3] The CCE takes this one step further by asking follow-up questions at the conclusion of the CE.

[3]There is a related stream of literature exploring the temporal stability of preferences over time (e.g., Dillaway et al., 2011; Ito and Kuriyama, 2017; Ji, Keiser, and Kling, 2020; Lusk, 2017).

Methods

Choice Experiment

We make use of the data, survey, and CE designed by Printezis and Grebitus (2018); these data were collected in 2017. The decision task entailed asking respondents to choose which one of four fresh tomato options (or none) they preferred to buy. Each option consisted of 1 lb of tomatoes and was described by five attributes: price ($0.99, $2.99, or $4.99)[4], purchase location (grocery store, farmers market, or urban farm), travel time to purchase (5 minutes, 15 minutes, or 25 minutes), organic status (USDA organic label or unlabeled), and provenance (locally grown or unlabeled).[5] Given this set of attributes and attribute levels, there are $3^3 2^2 = 108$ possible tomato purchase options that could be constructed. Printezis and Grebitus (2018) conducted a preliminary survey with 21 respondents using an orthogonal design in order to obtain priors. Then, they used the priors to construct a Bayesian-efficient design to minimize D-error, which is related to the size of the standard errors in a multinomial logit model. They specified their design so that main-effects for all attribute levels and two-way interaction effects between local, organic, and purchase location were identified. The resulting design consisted of 36 choice sets. The 36 choices were allocated to four blocks, each with nine choice questions. Respondents were randomly assigned to one of the four blocks, and the order of the questions was randomized within each block across respondents. An example choice task is shown in figure 1.

[4] Rather than utilizing three discrete price levels, an alternative design strategy would be to minimize D-error given a wider range of possible prices. While such an approach might improve D-error, it is also the case that it would result in price having many more levels than other attributes, which might increase the attention people pay to this attribute compared to others (Wittink et al., 1992). Moreover, in principle, only two price levels are needed to estimate the marginal utility of price changes (i.e., two points make a line). Ultimately, the estimated price effect is highly significant indicating the design is sufficiently efficient to identify this effect.

[5] To facilitate a uniform understanding of "urban farm," participants received the following text before making choices: "Urban agriculture or urban farming is the practice of growing, processing, and distributing food and other products through intensive plant cultivation and animal husbandry in and around cities. Examples include backyard, roof-top and balcony gardening, community gardening in vacant lots and parks, community supported agriculture based in urban areas, family farms located in metropolitan greenbelts, and roadside urban fringe agriculture."

First CE Conducted in 2017

The first component of our dataset consists of data collected in the summer of 2017, which was previously analyzed and discussed in Printezis and Grebitus (2018). This 2017 dataset consists of 524 individuals in Phoenix and 522 individuals in Detroit, each of whom answered nine CE questions. The survey was programmed in Qualtrics and was administered to samples of respondents maintained by Qualtrics.

Second CE Conducted in 2020

The exact same CE as in Printezis and Grebitus (2018) was repeated in 2020 with a different sample of respondents. The 2020 survey was also programmed in Qualtrics (meaning the presentation and "look and feel" of choices was identical to the 2017 survey), and the CE used the same experimental design, same blocking, etc. as the 2017 survey. As in the 2017 implementation, each of the respondents was randomly assigned to one of the four blocks, each of which consisted of nine CE questions. Again, residents of Phoenix and Detroit were surveyed. The only substantive difference between the two implementations was that a different panel provider, Dynata, was used in 2020. The survey was in the field from May 13 to May 20, 2020. The final 2020 data set consists of responses from 415 individuals in Phoenix and 449 individuals in Detroit.

This paper primarily focuses on the methodological issues associated with the CCE rather than differences in the WTP between 2017 and 2020 *per se*. Thus, in this paper, we largely refrain from comparing estimated WTP from 2017 to that from 2020. Rather, to implement the CCE, the 2017 data are used to estimate a distribution of preferences, and the CCE uses choices in 2020 to locate respondents' preferences in that prior distribution. Even if the distributions of preferences in 2017 and 2020 are not identical, this need not be problematic for the CCE approach so long as the initial sample has enough preference variation to encompass preferences exhibited in 2020. Even if the 2020 posterior distributions conditioned on 2017 preferences are biased, the whole purpose of the CCE approach is that it allows the 2020 respondents to indicate if and how the estimates are incorrect. Nonetheless, to help reduce potential differences between the 2017 and 2020 samples, we created weights to force our 2020 samples to match the 2017 distribution of demographic characteristics in the respective locations in terms of gender, age, education, presence of children in the household, race, and income. For reasons that are not entirely clear (although perhaps due to the different panel providers), the samples collected in 2020 tended to be older, more highly educated, and higher-income in both Detroit and Phoenix than the samples in 2017 (see appendix table A1 for the demographic characteristics of the samples and table A2 for characteristics of the populations of the locations according to the Census). In practice, the weights were constructed using iterative proportional fitting techniques following Izrael et al. (2000). By construction, when weights are applied to the data, the means of these five categories of demographic variables from the 2020 samples match the 2017 samples in each location.[6]

Calibrated Choice Experiment (CCE)

The CCE works in the following steps.

[6] An alternative approach to adjust for demographic differences is to interact demographic variables with the attribute coefficients. Our CE model has 7 coefficients associated with the mean attribute and attribute levels. Our weighting procedure uses 11 demographic dummies associated with five demographic categories. Thus, using an approach with demographic interactions would result in 11*7 = 77 new coefficients. Another approach might be propensity score matching; however, this approach requires judgments about sufficient degrees of overlap and balance. Thus, we chose weighting, which is common in survey and polling research, as it accomplishes the objective of yielding more comparable samples in a far more parsimonious way.

Step 1. **Conduct Initial Survey**. The first step consists of carrying out an initial CE. This step was accomplished by Printezis and Grebitus (2018) in their 2017 study, where each respondent completed nine choices like the one shown in figure 1.

Step 2. **Estimate Preference Distribution from First Sample**. The second step of the CCE entails using the data obtained in the initial CE to estimate the distribution of consumer preferences in the population; estimates which can be used as priors in a follow-up survey. It is possible to fit any number of models, but in our case, we utilized the data collected in 2017 to estimate a latent-class model (LCM) fit to the data pooled across locations. We chose to estimate an LCM for two main reasons. First, prior research suggests that, for an equivalent number of choices, the LCM often provides more accurate conditional estimates of individual preferences than the random parameter or mixed logit (Sarrias and Daziano, 2018; Sarrias, 2020). Second, and more pragmatically, the CCE approach entails providing survey respondents with an estimate of their preferences "on the fly," while they are taking the survey, and the LCM is computationally easier to utilize in our survey software (Qualtrics) compared to a random parameter model, which would have required evaluating integrals through methods such as simulation.

The prior estimates are obtained by use of a random utility model where consumer *i* is assumed to derive the following utility from choice option *j*: $U_{ij} = V_{ij} + \varepsilon_{ij}$. If the ε_{ij} follow a Type I extreme value distribution and are independently and identically distributed across *i* and *j*, then the conventional multinomial logit model (MNL) is obtained:

(1) $$\text{Prob}(i \text{ chooses } j) = \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}}.$$

The systematic portion of the utility function is:

(2) $V_{ij} = \alpha_i p_j + \beta_{i1} FM_j + \beta_{i2} UF_j + \beta_{i3} Organic_j + \beta_{i4} Local_j + \beta_{i5} Time_j + \beta_{i6} None_j$

where p_j is the price of alternative *j*, α_i is the marginal utility of a price change, *FM* and *UF* are dummy variables indicating whether the tomatoes are from a farmers market or urban farm (vs. a grocery store), *Organic* takes the value of 1 if tomato option *j* is labeled organic and zero otherwise, *Local* takes the value of 1 if the tomato is labeled local and zero otherwise, *Time* is the travel time to make the purchase of option *j* in minutes, *None* is an indicator variable indicating whether alternative *j* is the "none of these" option, and the β's are the marginal utilities of the aforementioned attributes. To allow for preference heterogeneity and to provide a prior for subsequent respondent preference estimates, we estimate an LCM, where the probability of choice is given by:

(3) Prob(i chooses *j*) = $\sum_{c=1}^{C} P_c \frac{e^{V_{jc}}}{\sum_{k=1}^{J} e^{V_{kc}}}$.

where P_c is the estimated probability belonging to class *c*, and V_{jc} is the class-specific version of equation (2). AIC and BIC fit criteria were used to guide the selection of the number of classes. These measures each improved up to 4 classes, and a 5-class model would not converge. Appendix table A3 shows fit statistics associated with MNL and LCM models. The estimates associated with the final 4-class LCM model, which were used to establish the priors for the CCE, are in appendix table A4.

Various authors have proposed comparing the means and standard deviations from the unconditional distribution to those implied by the individual-specific conditional distributions as a measure of goodness of fit (Train, 2009; Sarrias and Daziano, 2018; Sarrias, 2020). It has also been argued that the closer are the estimates from the two distributions, the more likely it is that the conditional estimates truly approximate each individual's true underlying preferences. We find a very high level of concordance between our estimated LCM unconditional means and standard deviations and the means and standard deviations from the conditional distributions (see appendix table A5). This implies that we can have confidence proceeding with the CCE and using the conditional estimates as an inference of each individual's preference.

Step 3. **Repeat the Choice Experiment.** The third step consists of repeating the CE with a new sample of respondents or re-surveying the same respondents to the initial survey. This step was accomplished by our 2020 study, where a new sample of respondents was surveyed in each location. As previously indicated, the experiment design and CE questions were identical to those used in 2017.

Step 4. Provide Real-Time Feedback to Respondents. The key purpose of estimating the initial LCM in step 2 is to establish priors on the distribution of preferences. These priors can be combined with an individual's actual choices obtained in step 3 to obtain a Bayesian posterior estimate of the conditional mean of an individual's preferences. As shown by Greene and Hensher (2003) or Sarrias and Daziano (2018), once in possession of the LCM estimates from step 2, the posterior probability of individual *i* belonging to class *c* can be calculated as:

$$(4) \qquad \hat{\pi}_{ic} = \frac{\hat{P}_{ic} \prod_{t=1}^{T} \prod_{j=1}^{J} \left(\frac{e^{\hat{V}_{jc}}}{\sum_{k=1}^{J} e^{\hat{V}_{kc}}} \right)^{y_{ijt}}}{\sum_{c=1}^{C} \hat{P}_{ic} \prod_{t=1}^{T} \prod_{j=1}^{J} \left(\frac{e^{\hat{V}_{jc}}}{\sum_{k=1}^{J} e^{\hat{V}_{kc}}} \right)^{y_{ijt}}}$$

where $y_{ijt} = 1$ if individual *i* chose option *j* in choice task *t,* and zero otherwise.[7] An estimate of the individual's preference for the k^{th} attribute is provided by the conditional expectation:

[7] To be clear, in the CCE approach, the estimated preferences, $\hat{V}_{jc}$, are taken from the prior survey in 2017, whereas the choices, y_{ijt}, are made in 2020. Thus, we create a posterior estimate using the distribution of 2017 preferences as priors updated with 2020 choices. It is possible to utilize the same formula to make the more typical calculation: a posterior estimate using the distribution of 2020 preferences as priors updated with 2020 choices.

(5) $\hat{\hat{\beta}}_{ik} = \sum_{c=1}^{C} \hat{\beta}_{ikc}\,\hat{\pi}_{ic}$

where $\hat{\beta}_{ikc}$ are class-specific parameter estimates from the LCM model (in appendix table A4), and where $\hat{\pi}_{ic}$ is defined in equation (4).[8] It follows, that the conditional expectation of an individual's willingness-to-pay for attribute *k*, or our "best guess," is:

(6) $\widehat{WTP}_{ik} = -\hat{\hat{\beta}}_{ik}/\hat{\hat{\alpha}}_{i}.$

Immediately after the CE was completed, and in real time, we utilized equation (6) to calculate each individual's willingness-to-pay for product attributes (local and organic) and point of sale (urban farm). Figure 2 shows an example of a feedback question shown to an individual for which equation (6) equaled \$0.51 for organic. We simply asked respondents to indicate whether the "best guess" was accurate, and if not, whether they were willing to pay more or less for the attribute in question. If the estimated willingness-to-pay for an individual and attribute was negative, respondents were directed to a reframed version of the question to avoid potentially confusing respondents with a negative willingness-to-pay. For example, if a respondent's calculated willingness to pay for organic was negative, rather than stating, "we calculate you are willing to pay -\$0.25 *more* for organic," we rephrased the statement to more intuitively read, "we calculate you are willing to pay \$0.25 *less* for organic."

Step 5. Calibrate Willingness-to-Pay

[8] Sarrias and Daziano (2018) discuss the fact that the posterior mean, or the conditional expectation of individual-specific parameters, $\hat{\hat{\beta}}_{ik}$, converges to the true β_{ik} as the number of choices made by the individual approaches infinity (assuming the model is appropriately specified). In their simulations, Sarrias and Daziano (2018) calculate the bias in using $\hat{\hat{\beta}}_{ik}$ as an inference for β_{ik} under scenarios were each individual made 1, 5, 10, 20, and 50 choices. They found absolute bias roughly halves when going from 5 to 10 choices. Moreover, they find a substantial reduction in bias in larger sample sizes. All considered, our dataset consists of 1,046 individuals, each of whom made 9 choices, which suggests an adequate number of individual-specific choices and sample size to use posterior means to approximate true preferences.

By soliciting respondents' assessments of whether the "best guess" willingness-to-pay values are accurate, too high, or too low, the overall mean willingness-to-pay can be adjusted (or calibrated) using interval censored regressions (e.g., see Cameron, 1988). Define WTP_{ik}^* as a respondent's true willingness-to-pay for attribute *k*. Let L_{ik} be an indicator variable denoting whether individual *i* indicated the estimated value, $\widehat{WTP}_{ik}$, was too low, implying $\widehat{WTP}_{ik} < WTP_{ik}^*$. Let H_{ik} be a variable indicating individual *i* responded that the estimated value, $\widehat{WTP}_{ik}$, was too high, implying $\widehat{WTP}_{ik} > WTP_{ik}^*$. Finally, let $E_{ik} = 1$ if an individual responded that $\widehat{WTP}_{ik} = WTP_{ik}^*$. If WTP_{ik}^* is Normally distributed with standard deviation σ_k, then the following likelihood function can be formulated:

$$(7) \qquad LF_k = \prod_{i=1}^{N} \left(1 - \Phi\left(\frac{\widehat{WTP}_{ik} - \mu_k}{\sigma_k}\right)\right)^{L_{ik}} \left(\Phi\left(\frac{\widehat{WTP}_{ik} - \mu_k}{\sigma_k}\right)\right)^{H_{ik}} \left(\frac{1}{\sigma_k}\varphi\left(\frac{\widehat{WTP}_{ik} - \mu_k}{\sigma_k}\right)\right)^{E_{ik}}$$

where Φ and φ are the cumulative and probability density functions associated with the standard Normal distribution, respectively, and μ_k is the calibrated mean willingness-to-pay for attribute *k*, which can be estimated by maximizing (7). It is also possible to specify the model to allow the mean, μ_k, to vary with other variables, such as demographics. The likelihood function in equation (7) is a standard interval censored regression model, which similar to that in a two-limit tobit model. The last part of the likelihood, associated with the $E_{ik} = 1$ are the uncensored observations, in which the Normal probability density function applies. By contrast, the other two portions of the likelihood relate to those observations where we only know that a respondents' value is higher or lower than our best guess, in which case the cumulative distribution function applies.

In many ways, step 5 of the CCE can be viewed in the same light as much of the early contingent valuation literature in which respondents are presented a price/tax/bid and are asked if they would pay it. The key difference between the older contingent valuation dichotomous choice question and this calibration step of our CCE is: 1) in addition to yes or no, respondents can indicate their willingness-to-pay is exactly equal to the bid, and 2) more importantly, instead of showing people a randomly chosen price/tax/bid as in conventional contingent valuation, in the CCE, the presented price/tax/bid is our estimate of the respondent's WTP resulting from their answers to the CE (and the estimated priors). Viewed in this light, equation (7) can be re-interpreted as a random utility function expressed in WTP-space (see Cameron, 1988).

In conventional CEs, great efforts are made to allow for respondent heterogeneity via the use of mixed logit or latent class models. Likewise, the calibrated WTP resulting from a CCE is not assumed to be identical, but rather varies in the population according to a specified distribution. As shown in equation (7), the calibrated WTP for attribute *k* is assumed to be distributed Normal, with mean μ_k and standard deviation σ_k. Thus, WTP is not assumed fixed or constant, but rather varies in the population according to μ_k and σ_k. While these parameters are not derived from a mixed logit model, the CCE approach allows for preference heterogeneity as indicated by the σ_k parameter.

Standard Analysis of CE Data

Note that the WTP estimates from the CCE approach described above uses the choices made in the 2020 CE to create individual-specific estimates based on the 2017 prior distribution, but the WTP values from the CCE approach do *not* require re-estimating the attribute-based utility function using the 2020 CE choices. Nonetheless, it is instructive to explore how the results would compare to those from a standard application of a CE had the CCE not been conducted. As indicated above, we chose to implement the CCE by fitting a LCM to the 2017 data. The feedback given to respondents to the second survey is, thus, based on LCM estimates. However, once the data are obtained from the follow-up survey, any econometric model can be fit to the follow-up data if interest is in the more conventional CE analysis. Therefore, it becomes an empirical question as to which model best fits the subsequent CE data.

We begin by fitting MNL models (based on equation 1) to the 2020 data (see estimates in the appendix). Then, we consider how to model respondent heterogeneity. One possibility is the LCM that has been previously discussed, and another is the mixed logit, or random parameter logit model (RPL), that specifies a continuous distribution of preferences. With the RPL, one must decide how to specify the distribution of preferences. Assuming a Normal distribution is a natural choice for many variables, but for attributes like price and time traveled, economic theory would constrain the coefficients to be negative. Two distributional options that constrain the sign of the coefficient are the one-sided triangular and the log-normal. These various specifications were compared along several different fit criteria (see appendix table A6). The RPL model, assuming time and price were log-normally distributed, was the best fitting model according to BIC measure of fit in both locations, but the LCM performed better according to the percentage of correct predictions in sample. Given this ambiguity, we chose to report results associated with the RPL because it is the more parsimonious specification.

The RPL model is specified as in equations (1) and (2), where for the farmers market, urban farm, organic, and local, the coefficients for individual i are specified as, $\beta_{ik} = \bar{\beta}_k + \rho_k \lambda_{ik}$, where $\bar{\beta}_k$ is the mean preference parameter for attribute k, ρ_k is the standard deviation of preferences, and λ_{ik} draw for individual i and attribute k from a standard Normal distribution. For the price and time parameters, they are assumed to be distributed log-normal. Because the log-normal is constrained to be positive, we multiplied price and time by negative one. For example, the price coefficient is specified as: $\alpha_i = -1 * \exp(\bar{\alpha} + \rho_\alpha \tau_i)$, where τ_i is an individual-specific random term that is distributed N(0,1). The RPL is estimated via simulation using 1,000 Halton draws.

Testing for Anchoring Effects

Following Day et al. (2012), Dekker, Koster and Brouwer (2014), and Ladenburg and Olsen (2006), we explore whether the results from step 3 in our CCE (i.e., the CE conducted in 2020) are affected by an arbitrary anchor: the first price shown to participants in the CE. As previously described, each respondent answered nine choice questions (there were a total of 36 questions that were blocked into four sets of nine). We randomly varied the order of questions across respondents. Because each price attribute was varied at three levels ($0.99, $2.99, or $4.99), this implies that the price of option A in the first question a respondent received also randomly varied between these same three levels. As such, we essentially have three treatment groups consisting of respondents who first saw a low, medium, or high price level. Moreover, because these were randomly assigned, the underlying CE design is identical for each group. To determine whether this arbitrary anchor, or starting point, affects respondent choice behavior, we modified the aforementioned choice model by letting each coefficient depend on the randomly assigned starting point. For example, we re-specify the price coefficient as: $\boldsymbol{\alpha_i = -1 * \exp(\bar{\alpha} + \gamma SP_i + \rho_\alpha \tau_i)}$, where $\boldsymbol{SP_i}$ is the starting point, representing the first price presented to respondents in Question 1, Option A, and $\boldsymbol{\gamma}$ represents the effects of the starting point on estimated price sensitivity. If $\boldsymbol{\gamma}$ is statistically significant, it implies price sensitivity, and thus WTP is affected by the starting point, $\boldsymbol{SP_i}$.

Results

We begin by discussing the conventional analysis of the CE conducted in 2020. These are the results that would be obtained had the CCE not been conducted, and they provide a baseline against which the subsequent CCE can be compared. In discussing these conventional CE results, we also explore the extent to which estimates are influenced by starting point bias. We then move to an exploration of the CCE approach, first by exploring the distribution of "best guesses" presented to respondents. Then we determine the extent to which respondents thought these "best guesses" were too high or too low, and the factors that determine the self-reported accuracy of the best guesses. We wrap-up by estimating the calibrated WTP values, and determine the extent to which these calibrated WTP values are more or less influenced by the starting-point biases observed in the conventional CE approach.

Standard Choice Experiment (CE) Results

Table 1 reports the results of RPL models fit to the 2020 choice data by location. Two specifications are presented for each location: one in which anchoring effects are ignored (Models 1 and 3) and another specification in which the mean of each coefficient is allowed to vary with the price starting point (Models 2 and 4).

The RPL models show sizable standard deviation estimates, suggesting significant preference heterogeneity across respondents. In fact, in many cases, standard deviations are larger than the means for farmers market, urban farm, and organic, suggesting people with divergent preferences. For example, in Model 1, the estimated mean preference for farmers market vs. grocery store in Phoenix was -0.242 with a standard deviation of 1.131. Given that this variable is Normally distributed, these estimates imply 58% of respondents prefer grocery stores to farmers markets and 42% prefer farmers markets to grocery stores. The distribution of preferences is tighter for local. In Phoenix, the mean and standard deviation associated with local is 0.477 and 0.166, respectively, implying 99% of respondents prefer local to non-local. Because price and time are specified log-normal, additional calculations are needed to determine the mean preference. For Model 1, for example, the mean price effect is $-1*\exp(-0.159+0.5*1.31^2) = -1.64$.

Comparing Model 1 to Model 2 and comparing Model 3 to Model 4 shows whether and to what extent preference estimates are affected by the arbitrary anchor that is the price starting point. According to a likelihood ratio test comparing Models 1 and 2, the null hypothesis that extra anchoring parameters in Model 2 in Phoenix are all equal to zero is rejected at the $p=0.01$ level (chi-square value of 17.8, with 7 degrees of freedom). Similarly, a likelihood ratio test comparing Models 3 and 4 for Detroit rejects the null the anchoring parameters are all zero at the $p<0.001$ level (chi-square value of 35.3 with 7 degrees of freedom). AIC values are smaller for the models incorporating anchoring effects, although the opposite is true of the BIC fit criteria, likely because BIC imposes a greater penalty on extra parameters, and it appears the primary anchoring effect occurs with the price parameter. Given the results of the likelihood ratio test and the AIC values, our preferred models include the starting point effects.

In both locations, the starting point significantly affected price sensitivity; higher starting points were associated with less price sensitivity. For example, in Phoenix (Model 2), individuals who saw $0.99 in their first choice for option A have a mean price effect of $-1*\exp(0.239 - 0.123*0.99 +0.5*1.296^2) = -2.150$; by contrast, individuals who saw $4.99 in their first choice for option A have a mean price effect of $-1*\exp(0.239 - 0.123*4.99 +0.5*1.296^2) = -1.314$. All else equal, this will imply that respondents who were randomly shown higher starting points will have higher WTP, in absolute value. On top of this, in Detroit (Model 4), the starting point was associated with significantly higher mean parameters for farmers market, local, and travel time.

Table 2 reports mean and median WTP estimates obtained from Models 2 and 4 in table 1 evaluated at the low- and high-anchors. Because WTP involves dividing two random coefficients, we simulate 1,000 draws from the unconditional distributions to calculate WTP, and because WTP involves dividing a normal distribution by a log-normal (or in the case of travel time, a log-normal by a log-normal), the distribution is skewed with the mean differing substantively from the median. Ninety-five percent confidence intervals are calculated using the Krinsky-Robb method and the bootstrapping approach in Poe et al. (2005) is used to calculate a p-value associated with the null that the WTP from the high-anchor is higher in absolute value than the WTP from the low anchor. Results indeed indicate mean and median WTP from the high-anchor condition are often statistically higher in absolute value than from the low-anchor condition. For example, for the value of local vs. non-local, on average consumers in Phoenix are willing to pay a \$0.69 premium for local tomatoes when confronted with a starting point of \$0.99, the WTP for local increases by a factor of 3 to \$2.08 when confronted with a starting point of \$4.99. As another example, the median WTP for an extra minute of travel time in Detroit is -\$0.06/minute (i.e., people would need to be compensated an extra \$0.06 for every minute they must travel to buy tomatoes) when shown the \$0.99 anchor; when the anchor is \$4.99, the median WTP for travel time falls by a factor of more than 2 to \$0.15/minute.

CCE Results

Having determined the baseline conventional CE results and established that the conventional analysis of CE data from 2020 suffers from starting point bias, we now explore the CCE approach. Table 3 shows the distribution of “best guess” estimates shown to respondents in 2020 immediately following their completion of the CE. Again, these "best guesses" were determined by combining, in real time, 2020 respondents' choices with the estimated prior distribution obtained from the 2017 data utilizing equation (5) for three attributes: organic vs. non-organic, local vs. non-local, and urban farm vs. grocery. In the case of organic in Phoenix, the median best-guess was $0.44, and the minimum and maximum were -$0.45 and $0.84.

Table 4 shows how respondents reacted to these "best guesses." About 60% of respondents in both locations indicated "yes" that they thought the estimated values for organic were accurate. In both locations, about 30% of respondents said the WTP for organic was too high. By contrast, table 4 shows that, in both locations, respondents thought the estimated values implied by their choices were too low for both local and urban farms. About 60% of respondents in both locations thought the estimated WTP value for local was too low and about 70% of respondents in both locations thought the estimated value for urban farms was too low.

While the data in table 4 provide some insight into how accurate were our "best guesses," table 5 provides a different way to explore the issue. In particular, we compared the "best guesses" shown to respondents (the conditional WTP using 2017 priors and 2020 choices) to the more conventional individual-specific conditional WTP (the conditional WTP using 2020 priors and 2020 choices). As shown in table 5, these values are positively but not perfectly correlated. Pearson correlations range from 0.22 for WTP for urban farms to 0.44 for local. While the "best guesses" are not perfect predictors of the conditional WTPs (based on the distribution of 2020 preferences), this need not be problematic. After all, as shown in table 2, the WTP values based on the 2020 data suffer from starting-point bias. Moreover, the CCE approach allows respondents to indicate whether the "best guesses" are too high or too low.

Table 6 reports ordered logit estimates associated with the likelihood that respondents indicated their true WTP was higher than the provided "best guess"; the dependent values are provided in table 4 and are coded 0 = no I am not willing to pay that much extra, 1 = yes, and 2 = no, I am willing to pay more. Results indicate that the likelihood of indicating whether the "best guess" was too high was *not* affected by the anchor. As expected, the higher was the "best guess," the lower the likelihood a respondent indicated their true WTP was too high. The likelihood of indicating one's WTP was higher than the "best guess" was, as expected, negatively affected by the individual-specific WTP conditioned on the 2020 distribution, but the effect was only significant for organic. There were few other consistent demographic or attitudinal effects across the three WTP values except for the variable related to worry about COVID. Consumers who were more worried about COVID were less likely to indicate their true WTP for local and urban farm vs. grocery tomatoes was too high.

The statements in table 4 are combined with the "best guess" WTP values (the conditional "individual specific" estimates based on 2020 choices and 2017 priors) and equation (7) to arrive at an updated or calibrated mean WTP. These values are shown in table 7. The results reveal no significant differences across locations. The calibrated mean for organic is \$0.393, and for local and urban farms vs. grocery are \$0.704 and -\$0.180, respectively. The latter value is substantively higher than the original un-calibrated means shown in table 2 at either anchor. This is not surprising since 73% of respondents across both locations (see table 4) indicated the "best guess" of their WTP for urban farm was too low.

Table 8 extends the results of table 7 to further explore how calibrated mean WTP varies with demographics and other variables in addition to location. Importantly, we find that the calibrated mean WTP was much less affected by the starting-point anchor relative to the uncalibrated values (see table 2). For local and urban, calibrated mean values were not significantly related to the starting point, which is in stark contrast to the uncalibrated values in WTP table 2. The mean calibrated WTP for organic was statistically affected by the starting point, but by a trivial amount. Going from the lowest to the highest starting point only increases the mean calibrated WTP by ($4.99-$0.99)*0.011=$0.033. Females were WTP more for organic and local than were males. People who had more trust in others were WTP more for organic, local, and urban farms.

Conclusions

This paper introduced a new extension to the popular choice experiment method. In particular, we introduced an approach to give people feedback about the implications of their choices. Such feedback might prompt people to provide WTP estimates that are more consistent with their true underlying preferences and behaviors. In our application, we found the majority of people thought their estimated willingness-to-pay from the choice experiment was understated for local food and for urban farms. On the other hand, the majority of respondents felt the estimated willingness-to-pay implied by their choices for organic was accurate. These findings suggest that allowing respondents to provide feedback on their answers can substantially alter the implied WTP premiums; in our case, particularly for food from urban farms relative to grocery stores. Moreover, we find that whereas WTP from the conventional CE was significantly affected by the arbitrary starting point represented by the first price respondents witnessed, the calibrated WTP was relatively unaffected by these anchors.

There are a number of avenues for future research worth considering. In the case of the calibrated choice experiment method, there are a variety of different ways that feedback could be provided. For example, upon being provided estimated willingness-to-pay values, respondents could be given the opportunity to provide more precise reactions as opposed to the simple “yes” or “no” options we provided in this study. It is even possible that the estimated WTP values from the choice experiment could be entered into a non-hypothetical auction similar to the approach used in Norwood and Lusk (2009). Another possibility is to provide feedback, not in terms of WTP values, but in terms of predictions about which options respondents would choose. Whether and to what extent these variations would lead to different valuations remains a question for future research. Our hope is that further refinements to stated preference methods, such as the calibrated choice experiment method introduced here, can lead to more refined predictions of preferences and market behavior.

References

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

Avitia, J., Costa-Font, M., Gil, J.M. and Lusk, J.L., 2015. Relative importance of price in forming individuals' decisions toward sustainable food: A calibrated auction-conjoint experiment. *Food Quality and Preference* 41:1-11.

Azevedo, C.D., Herriges, J.A. and Kling, C.L., 2003. Combining revealed and stated preferences: consistency tests and their interpretations. *American Journal of Agricultural Economics* 85(3):525-537.

Bateman, I.J., Burgess, D., Hutchinson, W.G. and Matthews, D.I., 2008. Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. *Journal of Environmental Economics and Management*, 55(2), pp.127-141.

Braga, J. and Starmer, C., 2005. Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics*, 32(1), pp.55-89.

Brooks, K. and Lusk, J.L., 2010. Stated and revealed preferences for organic and cloned milk: combining choice experiment and scanner data. *American Journal of Agricultural Economics* 92(4):1229-1241.

Cameron, T.A., 1988. A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics and Management* 15(3):355-379.

Caputo, V., Lusk, J.L. and Nayga Jr, R.M., 2018. Choice experiments are not conducted in a vacuum: The effects of external price information on choice behavior. *Journal of Economic Behavior & Organization* 145:335-351.

Caputo, V., Lusk, J.L. and Nayga, R.M., 2020. Am I Getting a Good Deal? Reference-Dependent Decision Making When the Reference Price Is Uncertain. *American Journal of Agricultural Economics* 102(1):132-153.

Carlsson, F., Mørkbak, M.R. and Olsen, S.B., 2012. The first time is the hardest: A test of ordering effects in choice experiments. *Journal of Choice Modelling*, 5(2):19-37.

Cason, Timothy N., and Charles R. Plott. "Misconceptions and game form recognition: Challenges to theories of revealed preference and framing." *Journal of Political Economy* 122, no. 6 (2014): 1235-1270.

Chang, J.B., Lusk, J.L. and Norwood, F.B., 2009. How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agricultural Economics* 91(2):518-534.

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

Crampton, L. 2020. "Coronavirus has more Americans turning directly to farms for food." *Politico*. March 31. Available online at: https://www.politico.com/news/2020/03/31/coronavirus-demand-for-local-farms-157538

Czajkowski, M., Giergiczny, M. and Greene, W.H., 2014. Learning and fatigue effects revisited: Investigating the effects of accounting for unobservable preference and scale heterogeneity. *Land Economics* 90(2):324-351.

Darby, K., Batte, M.T., Ernst, S. and Roe, B., 2008. Decomposing local: A conjoint analysis of locally produced foods. *American Journal of Agricultural Economics* 90(2):476-486.

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R. and Wang, P., 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management* 63(1):73-91.

Dekker, T., Koster, P. and Brouwer, R., 2014. Changing with the tide: Semiparametric estimation of preference dynamics. *Land Economics* 90(4):717-745.

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

Dillaway, R., Messer, K.D., Bernard, J.C. and Kaiser, H.M., 2011. Do consumer responses to media food safety information last? *Applied Economic Perspectives and Policy* 33(3):363-383.

Diamond, P. A., & Hausman, J. A., 1994. Contingent valuation: is some number better than no number?. *Journal of Economic Perspectives* 8(4):45-64.

Dyer, D. and Kagel, J.H., 1996. Bidding in common value auctions: How the commercial construction industry corrects for the winner's curse. *Management Science* 42(10):1463-1475.

Fang, D., Nayga Jr, R.M., West, G.H., Bazzani, C., Yang, W., Lok, B.C., Levy, C.E. and Snell, H.A., 2020. On the Use of Virtual Reality in Mitigating Hypothetical Bias in Choice Experiments. *American Journal of Agricultural Economics*. forthcoming. https://doi.org/10.1111/ajae.12118

Grebitus, C., Lusk, J.L. and Nayga Jr, R.M., 2013. Effect of distance of transportation on willingness to pay for food. *Ecological Economics* 88:67-75.

Greene, W.H. and Hensher, D.A., 2003. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological* 37(8):681-698.

Hanley, N., Adamowicz, W. and Wright, R.E., 2005. Price vector effects in choice experiments: an empirical test. *Resource and Energy Economics* 27(3):227-234.

Hausman, J., 2012. Contingent valuation: from dubious to hopeless. *Journal of Economic Perspectives* 26(4):43-56.

Hensher, D.A., Greene, W.H. and Rose, J.M., 2006. Deriving willingness-to-pay estimates of travel-time savings from individual-based parameters. *Environment and Planning A* 38(12):2365-2376.

Hensher, D., Louviere, J. and Swait, J., 1998. Combining sources of preference data. *Journal of Econometrics*, 89(1-2):197-221.

Herriges, J.A. and Shogren, J.F., 1996. Starting point bias in dichotomous choice valuation with follow-up questioning. *Journal of Environmental Economics and Management*, 30(1), pp.112-131.

Hess, S. and Hensher, D.A., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B: Methodological* 44(6):781-790.

Hu, W., Adamowicz, W.L. and Veeman, M.M., 2006. Labeling context and reference point effects in models of food attribute demand. *American journal of Agricultural Economics*, 88(4):1034-1049.

Huffstutter, P.J. and R. Nickel. 2020. 'How about next June?' Small meat processors backlogged as virus idles big plants. *Reuters*. May 26. Available online at: https://www.reuters.com/article/us-health-coronavirus-meatpacking/how-about-next-june-small-meat-processors-backlogged-as-virus-idles-big-plants-idUSKBN23217V

Ito, N. and Kuriyama, K., 2017. Averting behaviors of very small radiation exposure via food consumption after the Fukushima nuclear power station accident. *American Journal of Agricultural Economics* 99(1):55-72.

Izrael, D., Hoaglin, D.C. and Battaglia, M.P., 2000. A SAS macro for balancing a weighted sample. In Proceedings of the twenty-fifth annual SAS users group international conference. Cary (NC): SAS Institute. Paper 258-25. April, pp. 9-12.

Ji, Y., Keiser, D.A. and Kling, C.L., 2020. Temporal Reliability of Welfare Estimates from Revealed Preferences. *Journal of the Association of Environmental and Resource Economists* 7(4):659-686.

Johns Hopkins. New Case Trends in all 50 State. Coronavirus resource center. https://coronavirus.jhu.edu/data/new-cases-50-states

Krinsky, I. and Robb, A.L. 1986. On Approximating the Statistical Properties of Elasticities. *Review of Economics and Statistics* 68:715–19.

Kovalsky, K.L. and Lusk, J.L., 2013. Do consumers really know how much they are willing to pay?. *Journal of Consumer Affairs* 47(1):98-127.

Ladenburg, J., & Olsen, S. B. 2006. Starting point anchoring effects in choice experiments. Center for Skov, Landskab og Planlægning/Københavns Universitet. IFRO Working Paper Vol. 2006 No. 6. Available online at: https://curis.ku.dk/ws/files/134876468/6_version_4.pdf.

List, J.A., 2003. Does market experience eliminate market anomalies?. *Quarterly Journal of Economics* 118(1):41-71.

List, J.A., 2004. Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica* 72(2):615-625.

Lobasenko, V. and Llerena, D., 2017. Elicitation of willingness to pay for upgradeable products with calibrated auction-conjoint method. *Journal of Environmental Planning and Management* 60(11):2036-2055.

Lusk, J.L., 2017. Consumer research with big data: applications from the food demand survey (FooDS). *American Journal of Agricultural Economics* 99(2):303-320.

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

Lusk, J.L., Fields, D. and Prevatt, W., 2008. An incentive compatible conjoint ranking mechanism. *American Journal of Agricultural Economics* 90(2):487-498.

Meas, T., Hu, W., Batte, M.T., Woods, T.A. and Ernst, S., 2015. Substitutes or complements? Consumer preference for local and organic food attributes. *American Journal of Agricultural Economics* 97(4):1044-1071.

Mørkbak, M.R., Christensen, T. and Gyrd-Hansen, D., 2010. Choke price bias in choice experiments. *Environmental and Resource Economics* 45(4):537-551.

Norwood, F.B. and Lusk, J.L., 2011. A calibrated auction-conjoint valuation method: Valuing pork and eggs produced under differing animal welfare conditions. *Journal of Environmental Economics and Management* 62(1):80-94.

Ortega, D.L., Lusk, J.L., Lin, W. and Caputo, V., 2020. Predicting responsiveness to information: consumer acceptance of biotechnology in animal products. *European Review of Agricultural Economics*. forthcoming https://doi.org/10.1093/erae/jbaa003

Plott, C. 1996. Rational individual behavior in markets and social choice processes: the discovered preference hypothesis. K. Arrow, E. Colombatto, M. Perlemann, C. Schmidt (Eds.), *The Rational Foundations of Economic Behavior*, London: Macmillan and St. Martin Press, pp. 225-250.

Plott, C.R. and Zeiler, K., 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.

Poe, Gregory L., Kelly L. Giraud, and John B. Loomis. 2005. Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics* 87(2):353-365.

Printezis, I., Grebitus C., and Hirsch, S. 2019. The price is right!? A meta-regression analysis on willingness to pay for local food. *PLoS ONE*, 14(5), e0215847

Printezis, I. and Grebitus, C., 2018. Marketing channels for local food. *Ecological Economics* 152:161-171.

Revelt, D. and K. Train. 2000. Customer-specific taste parameters and mixed logit: households' choice of electricity supplier. Working Paper, Department of Economics, University of California at Santa Barbara

Richards, T.J., Hamilton, S.F., Gomez, M. and Rabinovich, E., 2017. Retail intermediation and local foods. *American Journal of Agricultural Economics* 99(3):637-659.

Sarrias, M., 2020. Individual-specific posterior distributions from Mixed Logit models: Properties, limitations and diagnostic checks. *Journal of Choice Modelling* 36:100224.

Sarrias, M. and Daziano, R.A., 2018. Individual-specific point and interval conditional estimates of latent class logit parameters. *Journal of Choice Modelling* 27:50-61.

Scarpa, R., Zanoli, R., Bruschi, V. and Naspetti, S., 2013. Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics* 95(1):165-180.

Scarpa R., Campbell D., Hutchinson W. G. 2007. Benefit Estimates for Landscape Improvements: Sequential Bayesian Design and Respondents' Rationality in a Choice Experiment Study *Land Economics* 83:617-34.

Schmiess, J. and J.L. Lusk. "Tradeoffs between Animal Welfare and Environmental Impacts of Beef Production: An Analysis of Presentation Effects on Consumer Choice." *Journal of Agricultural and Resource Economics*. forthcoming.

Schwarzer, R. and Jerusalem, M. 1995. Generalized self-efficacy scale. *Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs*. pp. 35-37.

Su, L., Adam, B.D., Lusk, J.L. and Arthur, F., 2017. Anchoring, information, and fragility of choice experiments: An application to consumer willingness to pay for rice with improved storage management. *Journal of Agricultural and Resource Economics*, 42:255-274.

Swait, J. and Andrews, R.L., 2003. Enriching scanner panel models with choice experiments. *Marketing Science* 22(4):442-460.

Swait, J. and Louviere, J., 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 30(3), pp.305-314.

Taylor, L.O., 1998. Incentive compatible referenda and the valuation of environmental goods. *Agricultural and Resource Economics Review*, 27:132-139.

Taylor, R. and Villas-Boas, S.B., 2016. Food store choices of poor households: A discrete choice analysis of the National Household Food Acquisition and Purchase Survey (FoodAPS). *American Journal of Agricultural Economics* 98(2):513-532.

Thilmany, D., Bond, C.A. and Bond, J.K., 2008. Going local: Exploring consumer behavior and motivations for direct food purchases. *American Journal of Agricultural Economics* 90(5):1303-1309.

Toler, S., Briggeman, B.C., Lusk, J.L. and Adams, D.C., 2009. Fairness, farmers markets, and local production. *American Journal of Agricultural Economics* 91(5):1272-1278.

Train, K.E., 2009. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Walljasper, C. and Polansek, T., 2020. Home gardening blooms around the world during coronavirus lockdowns. *Sustainable Business*, April 19, 2020.

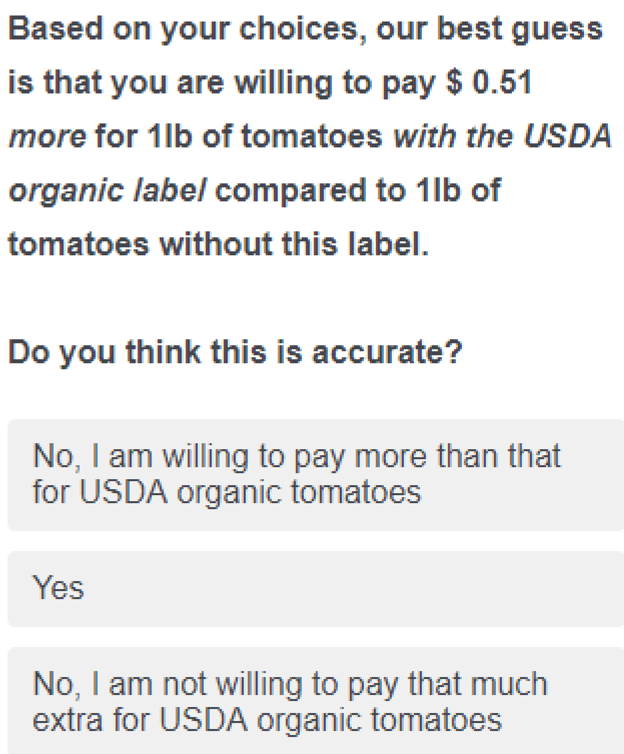
Wittink, D.R., Huber, J., Zandan, P. and Johnson, R.M., 1992, June. The number of levels effect in conjoint: Where does it come from, and can it be eliminated. In Sawtooth Software Conference Proceedings. Ketchum, ID: Sawtooth Software, pp. 355-364. Available online at: https://sawtoothsoftware.com/resources/technical-papers/the-number-of-levels-effect-in-conjoint-where-does-it-come-from-and-can-it-be-eliminated

Wüepper, D., Clemm, A. and Wree, P., 2019. The preference for sustainable coffee and a new approach for dealing with hypothetical bias. *Journal of Economic Behavior & Organization* 158:475-486.

Imagine you would like to purchase 1lb of tomatoes you usually buy.
Please choose your preferred alternative or choose "none of these."

Alternative A	Alternative B	Alternative C	Alternative D
$4.99	$2.99	$0.99	$0.99
At urban farm	At farmers market	At grocery store	At urban farm
	USDA Organic	USDA Organic	USDA Organic
Locally grown	Locally grown		
Travel time one way 15 min.	Travel time one way 15 min.	Travel time one way 5 min.	Travel time one way 15 min.

Figure 1. Example Choice Experiment Question



Based on your choices, our best guess is that you are willing to pay $ 0.51 *more* for 1lb of tomatoes *with the USDA organic label* compared to 1lb of tomatoes without this label.

Do you think this is accurate?

No, I am willing to pay more than that for USDA organic tomatoes

Yes

No, I am not willing to pay that much extra for USDA organic tomatoes

Figure 2. Example Feedback Question Following Choice Experiment

Table 1. Random Parameter Logit Models fit to 2020 data by Location

	Phoenix		**Detroit**	
	Model 1 (no anchoring)	**Model 2 (anchor[f])**	**Model 3 (no anchoring)**	**Model 4 (anchor[f])**
Mean				
-1*Price[a]	-0.159*[b] (0.077)[c]	0.239 (0.155)	-0.377* (0.068)	-0.024 (0.150)
Farmers market[d]	-0.242* (0.101)	-0.058 (0.202)	-0.184* (0.081)	-0.497* (0.150)
Urban farm[d]	-0.834* (0.104)	-0.610* (0.227)	-0.584* (0.088)	-0.928* (0.187)
Organic[e]	0.459* (0.093)	0.282 (0.190)	0.299* (0.067)	0.150 (0.151)
Local[e]	0.477* (0.072)	0.253 (0.158)	0.395* (0.071)	0.014 (0.162)
-1*Travel time[a]	-2.357* (0.084)	-2.501* (0.189)	-2.639* (0.079)	-2.971* (0.181)
None	-11.318* (0.716)	-12.534* (0.942)	-10.563* (0.575)	-11.181* (0.840)
Interactions with Starting Point				
Price	---	-0.123* (0.046)	---	-0.101* (0.045)
Farmers market	---	-0.063 (0.061)	---	0.103* (0.047)
Urban farm	---	-0.073 (0.063)	---	0.105 (0.054)
Organic	---	0.056 (0.058)	---	0.051 (0.045)
Local	---	0.071 (0.044)	---	0.134* (0.046)
Travel time	---	0.037 (0.047)	---	0.114* (0.049)
None	---	0.347* (0.153)	---	0.059 (0.193)
Standard Deviation/Scale				
Price[a]	1.310* (0.076)	1.296* (0.077)	1.300* (0.064)	1.308* (0.078)
Farmers market	1.131* (0.103)	1.160* (0.108)	0.856* (0.095)	0.848* (0.104)
Urban farm	1.272* (0.134)	1.337* (0.136)	1.218* (0.106)	1.264* (0.112)
Organic	1.240* (0.085)	1.265* (0.090)	0.767* (0.085)	0.804* (0.092)
Local	0.166* (0.246)	0.077 (0.287)	0.718* (0.096)	0.682* (0.100)
Travel[a]	1.058* (0.082)	1.283* (0.096)	1.083* (0.080)	1.144* (0.081)
None	6.465* (0.515)	6.997* (0.560)	5.598* (0.439)	6.067* (0.465)
Log Likelihood	-3673.1	-3664.2	-4273.6	-4255.9
AIC	7374.2	7370.4	8575.2	8553.9
BIC	7461.4	7501.2	8663.5	8684.6
% Correct	0.59	0.59	0.58	0.58
N choices	3735	3735	4041	4041
N people	415	415	449	449

[a]Coefficient distributed lognormal, all others distributed Normally.
[b]One asterisk represents statistical significance at the 0.05 level or lower.
[c]Numbers in parentheses are standard errors.
[d]Effect relative to shopping in grocery store.
[e]Effect relative to no label present.
[f]The anchor is the price of option A in the first choice task completed by respondents (the order of choice tasks was randomized across respondents).

Table 2. Willingness-to-Pay Estimates from Unconditional Distribution of Random Parameter Logit Models with Price Anchors fit to 2020 data

	Phoenix			Detroit		
	Anchor $0.99	**Anchor $4.99**	**p-value[a]**	**Anchor $0.99**	**Anchor $4.99**	**p-value[a]**
Mean WTP						
Farmers market[c]	-0.24	-1.23	0.10	-1.05	0.04	0.06
	[-1.01, 0.51][b]	[-2.55, 0.10]		[-1.88, -0.39]	[-1.14, 1.23]	
Urban farm[c]	-1.46	-3.37	0.03	-2.24	-1.63	0.74
	[-2.64, -0.58]	[-5.49, -1.88]		[-3.8, -1.17]	[-3.17, -0.29]	
Organic[d]	0.71	1.93	0.07	0.54	1.65	0.06
	[-0.02, 1.55]	[0.69, 3.54]		[-0.15, 1.36]	[0.60, 2.86]	
Local[d]	0.69	2.08	0.01	0.41	2.73	0.00
	[0.16, 1.38]	[1.14, 3.21]		[-0.26, 1.17]	[1.54, 4.29]	
Travel time	-0.42	-0.77	0.06	-0.30	-0.72	0.01
	[-0.24, -0.71]	[-0.44, -1.25]		[-0.18, -0.47]	[-0.43, -1.13]	
Median WTP						
Farmers market[c]	-0.04	-0.28	0.09	-0.25	0.01	0.03
	[-0.21, 0.09]	[-0.64, 0]		[-0.47, -0.09]	[-0.23, 0.26]	
Urban farm[c]	-0.36	-0.93	0.01	-0.60	-0.36	0.82
	[-0.7, -0.15]	[-1.47, -0.52]		[-0.96, -0.31]	[-0.78, -0.08]	
Organic[d]	0.15	0.46	0.05	0.11	0.41	0.04
	[0.02, 0.34]	[0.16, 0.87]		[-0.01, 0.31]	[0.14, 0.79]	
Local[d]	0.23	0.80	0.01	0.08	0.82	0.00
	[0.05, 0.50]	[0.42, 1.22]		[-0.04, 0.28]	[0.44, 1.35]	
Travel time	-0.08	-0.14	0.02	-0.06	-0.15	0.00
	[-0.05, -0.12]	[-0.10, -0.21]		[-0.04, -0.10]	[-0.10, -0.23]	

[a]p-values associated with null that WTP with a $4.99 anchor is higher in absolute value than WTP with the $0.99 according to the Poe et al. (2005) test.
[c]Numbers in brackets are 95% confidence intervals determined by the Krinsky-Robb (1986) method.
[c]Effect relative to shopping in grocery store.
[d]Effect relative to no label present.

Table 3. Distribution of Best-Guess Willingness-to-Pay values Shown to Respondents Calculated from Individual-Specific Conditional Distribution Using 2017 Priors and 2020 Choices

	Minimum	25th Percentile	Median	Mean	75th Percentile	Maximum
Phoenix (N=449)						
Organic	-\$0.45	\$0.35	\$0.44	\$0.41	\$0.50	\$0.84
Local	\$0.38	\$0.54	\$0.57	\$0.58	\$0.62	\$0.87
Urban farm	-\$1.52	-\$0.50	-\$0.46	-\$0.51	-\$0.44	-\$0.42
Detroit (N=415)						
Organic	-\$0.45	\$0.36	\$0.44	\$0.42	\$0.50	\$1.06
Local	\$0.43	\$0.54	\$0.58	\$0.58	\$0.61	\$1.06
Urban farm	-\$1.41	-\$0.50	-\$0.46	-\$0.50	-\$0.44	-\$0.41

Table 4. Responses to Post-Choice Experiment Feedback Questions

Based on your choices, our best guess is that you are willing to pay <$X> more for 1lb of tomatoes with <attribute> compared to 1lb of tomatoes without <attribute>. Do you think this is accurate?	Organic	Local	Urban Farm
Pooled			
Yes	58.6%	9.8%	11.5%
No, I am willing to pay more	10.1%	59.8%	73.0%
No, I am not willing to pay that much extra	31.4%	30.3%	15.5%
Number of observations	864	844	859
Phoenix			
Yes	58.3%	7.7%	10.9%
No, I am willing to pay more	8.7%	57.9%	73.1%
No, I am not willing to pay that much extra	33.0%	34.4%	16.0%
Number of observations	415	404	413
Detroit			
Yes	58.8%	11.8%	12.1%
No, I am willing to pay more	11.4%	61.6%	72.9%
No, I am not willing to pay that much extra	29.8%	26.6%	15.0%
Number of observations	449	440	446

Table 5. Correlations between Best Guesses and Willingness-to-Pay from Individual-Specific Conditional Distribution Using 2020 Priors and 2020 Choices, Pooled across Location

	Pearson	Spearman Rank
Organic	0.39 {<0.01}[a]	0.55 {<0.01}
Local	0.44 {<0.01}	0.48 {<0.01}
Urban farm	0.22 {<0.01}	0.29 {<0.01}

[a]Numbers in brackets { } are p-values associated with a test of the null hypothesis that the correlation is equal to zero

Table 6. Probability True WTP is Higher than Best Guess; Ordered Logit Estimates[a]

Variable	Organic	Local	Urban Farm
Threshold 1[b]	2.250*[c] (0.754)[d]	4.490* (0.987)	-0.305 (0.872)
Threshold 2[b]	5.610* (0.780)	4.975* (0.990)	0.470 (0.872)
Phoenix vs. Detroit	0.108 (0.147)	0.237 (0.151)	-0.079 (0.167)
Anchor	-0.056 (0.044)	0.039 (0.047)	-0.090 (0.050)
Best Guess WTP[e]	-1.924* (0.496)	-7.581* (1.264)	-2.858* (0.524)
2020 Conditional WTP[f]	-0.095* (0.041)	-0.042 (0.077)	-0.009 (0.028)
Female vs. Male	-0.165 (0.151)	-0.262 (0.155)	-0.348* (0.173)
Age: <30 vs. ≥ 60	-0.887* (0.288)	-0.300 (0.298)	0.451 (0.316)
Age: 30-44 vs. ≥ 60	-0.489* (0.238)	-0.380 (0.248)	0.558* (0.264)
Age: 45-60 vs. ≥ 60	-0.229 (0.193)	-0.199 (0.196)	0.322 (0.217)
Children vs. No Children	-0.316 (0.239)	0.039 (0.252)	-0.615* (0.274)
White vs. Other Race	-0.298 (0.266)	-0.099 (0.277)	-0.424 (0.286)
Black vs. Other Race	-0.489 (0.335)	0.248 (0.347)	-0.210 (0.357)
Obtained BS/BA vs. not	0.110 (0.157)	0.184 (0.162)	0.136 (0.180)
Income: >$30k vs. ≥ $90k	0.303 (0.248)	-0.052 (0.258)	0.447 (0.274)
Income: >$30-59k vs. ≥ $90k	-0.020 (0.199)	-0.201 (0.207)	-0.064 (0.230)
Income: >$60k vs. ≥ $89k	0.101 (0.194)	-0.176 (0.200)	0.037 (0.223)
Household size: 1 vs. 5 or more	-0.162 (0.337)	0.477 (0.360)	0.133 (0.378)
Household size: 2 vs. 5 or more	-0.366 (0.312)	0.208 (0.339)	0.026 (0.355)
Household size: 3 vs. 5 or more	-0.484 (0.315)	0.413 (0.341)	-0.008 (0.365)
Household size: 4 vs. 5 or more	0.003 (0.323)	0.307 (0.350)	-0.063 (0.378)
Political Party: Dem vs. Indep	-0.016 (0.189)	-0.092 (0.195)	-0.232 (0.212)
Political Party: Rep vs. Indep	0.226 (0.188)	0.091 (0.191)	-0.116 (0.210)
Political Party: Other vs. Indep	0.454 (0.306)	0.084 (0.313)	-0.300 (0.348)
Risk Aversion[g]	-0.140* (0.047)	0.006 (0.050)	-0.009 (0.053)
Worried about COVID[h]	-0.051 (0.064)	-0.135* (0.065)	-0.149* (0.071)
Worried about Food Security[i]	-0.146* (0.071)	-0.025 (0.074)	-0.069 (0.080)
Trust[j]	-0.261 (0.149)	-0.304 (0.155)	-0.101 (0.171)
Self Efficacy[k]	-0.016 (0.021)	-0.023 (0.021)	-0.068* (0.023)

[a]Dependent variable has three response categories for each attribute as shown in table 4
[b]Threshold intercept parameters associated with ordered logit categorical responses
[c]One asterisk implies coefficient is statistically different from zero at the 0.05 level or lower
[d]Numbers in parentheses are standard errors
[e]Individual-Specific Conditional WTP Using 2017 Distribution Priors and 2020 Choices
[f]Individual-Specific Conditional WTP Using 2020 Distribution Priors and 2020 Choices
[g]"Would you say that the decisions you make are rather risky or not risky?"; 9 = more than extremely risky; 1 = not at all risky
[h]"How worried are you about the following … COVID-19 in general"; 5 = a great deal; 1 = not at all
[i]"How worried are you about the following … Having enough to eat"; 5 = a great deal; 1 = not at all
[j]"Generally speaking, would you say that most people can be trusted or that you should be very careful in dealing with people?" 1 = most people can be trusted; 0 otherwise
[k]Average response to 10-item self-efficacy scale from Schwarzer and Jerusalem (1995)

Table 7. Calibrated Mean WTP Estimates; Interval Censored Estimates

Variable	Organic	Local	Urban Farm
Intercept[a]	0.393*[b] (0.011)[c]	0.704* (0.02)	-0.180* (0.038)
Phoenix vs. Detroit	-0.029 (0.015)	-0.045 (0.025)	-0.010 (0.039)
Standard Deviation	0.204* (0.007)	0.288* (0.026)	0.436* (0.035)

[a]Intercept corresponds to the mean WTP in Detroit
[b]One asterisk implies coefficient is statistically different from zero at the 0.05 level or lower
[c]Numbers in parentheses are standard errors

Table 8. Determinants of Calibrated Mean WTP; Interval Censored Estimates

Variable	Organic	Local	Urban Farm
Intercept	-0.031 (0.076)[a]	0.249*[b] (0.118)	-0.988* (0.197)
Phoenix vs. Detroit	-0.021 (0.015)	-0.036 (0.023)	0.001 (0.039)
Anchor	0.011* (0.005)	0.005 (0.007)	0.022 (0.012)
Best Guess WTP[e]	0.030 (0.016)	0.047* (0.024)	0.074 (0.041)
2020 Conditional WTP[f]	0.122* (0.028)	0.12* (0.044)	-0.071 (0.074)
Female vs. Male	0.081* (0.024)	0.113* (0.037)	-0.08 (0.063)
Age: <30 vs. ≥ 60	0.024 (0.02)	0.048 (0.03)	-0.082 (0.052)
Age: 30-44 vs. ≥ 60	0.016 (0.024)	0.016 (0.037)	0.088 (0.065)
Age: 45-60 vs. ≥ 60	0.030 (0.027)	0.013 (0.042)	0.123 (0.069)
Children vs. No Children	0.011 (0.034)	-0.02 (0.052)	0.017 (0.086)
White vs. Other Race	-0.002 (0.016)	-0.017 (0.025)	-0.021 (0.043)
Black vs. Other Race	-0.038 (0.025)	-0.034 (0.038)	-0.101 (0.066)
Obtained BS/BA vs. not	-0.017 (0.021)	0.015 (0.031)	0.009 (0.055)
Income: >$30k vs. ≥ $90k	-0.015 (0.02)	0.005 (0.03)	-0.006 (0.053)
Income: >$30-59k vs. ≥ $90k	0.008 (0.034)	-0.042 (0.054)	-0.064 (0.09)
Income: >$60k vs. ≥ $89k	0.025 (0.032)	-0.007 (0.05)	-0.007 (0.084)
Household size: 1 vs. 5 or more	0.033 (0.032)	-0.021 (0.05)	-0.007 (0.086)
Household size: 2 vs. 5 or more	0.002 (0.033)	-0.025 (0.052)	0.005 (0.088)
Household size: 3 vs. 5 or more	0.015* (0.005)	0.01 (0.007)	-0.003 (0.013)
Household size: 4 vs. 5 or more	0.019* (0.007)	0.032* (0.01)	0.047* (0.017)
Political Party: Dem vs. Indep	-0.064* (0.032)	-0.025 (0.048)	0.036 (0.082)
Political Party: Rep vs. Indep	0.003 (0.015)	0.040 (0.024)	-0.012 (0.040)
Political Party: Other vs. Indep	0.007* (0.002)	0.008* (0.003)	0.019* (0.006)
Risk Aversion[g]	0.011 (0.007)	0.012 (0.011)	0.017 (0.019)
Worried about COVID[h]	0.018 (0.019)	0.034 (0.029)	0.080 (0.050)
Worried about Food Security[i]	-0.005 (0.019)	0.002 (0.029)	0.019 (0.050)
Trust[j]	0.191* (0.006)	0.249* (0.022)	0.406* (0.033)
Self Efficacy[k]	-0.031 (0.076)	0.249* (0.118)	-0.988* (0.197)
Standard Deviation	-0.021 (0.015)	-0.036 (0.023)	0.001 (0.039)

[a]Numbers in parentheses are standard errors
[b]One asterisk implies coefficient is statistically different from zero at the 0.05 level or lower
[e]Individual-Specific Conditional WTP Using 2017 Distribution Priors and 2020 Choices
[f]Individual-Specific Conditional WTP Using 2020 Distribution Priors and 2020 Choices
[g]"Would you say that the decisions you make are rather risky or not risky?"; 9 = more than extremely risky; 1 = not at all risky
[h]"How worried are you about the following … COVID-19 in general"; 5 = a great deal; 1 = not at all
[i]"How worried are you about the following … Having enough to eat"; 5 = a great deal; 1 = not at all
[j]"Generally speaking, would you say that most people can be trusted or that you should be very careful in dealing with people?" 1 = most people can be trusted; 0 otherwise
[k]Average response to 10-item self-efficacy scale from Schwarzer and Jerusalem (1995)

Appendix

Table A1. Demographic Characteristics of Samples in 2017 and 2020 by Location

Characteristic	**Detroit**		**Phoenix**	
	2017	**2020**	**2017**	**2020**
Female	50.2%	56.3%	50.0%	49.4%
Age < 30 years	22.6%	16.0%	22.7%	10.4%
Age 30-44 years	27.2%	16.5%	27.9%	22.7%
Age 45-60 years	27.6%	22.9%	26.0%	24.6%
Age ≥ 60 years	22.6%	44.5%	23.5%	42.4%
Children under 12 in household	24.1%	17.1%	28.1%	21.5%
White	73.2%	75.9%	81.3%	84.6%
Black	18.2%	18.0%	4.8%	3.9%
Other race	8.6%	6.0%	13.9%	11.6%
Obtained BS or BA degree	39.1%	53.2%	40.1%	49.9%
Income < $30,000	29.9%	17.4%	26.5%	16.1%
Income $30,000-$59,999	33.7%	27.2%	35.9%	25.3%
Income $60,000-$89,999	19.9%	21.8%	16.8%	23.1%
Income ≥ $90,000	16.5%	33.6%	20.8%	35.4%
# Obs	522	449	524	415

Table A2. Characteristics of Detroit and Phoenix According to U.S. Census Bureau in 2019

Characteristic	**Detroit**		**Phoenix**	
	City	**MSA**[a]	**City**	**MSA**[b]
Female	52.2%	51.3%	50.1%	50.3%
Age < 30 years	23.1	18.9	22.3	20.3
Age 30-44 years	25.4	23.7	30.1	26.6
Age 45-60 years	23.8	26.7	25.5	24.3
Age ≥ 60 years	27.7	30.7	22.2	28.7
Children under 18 in household	24.7%	21.8%	24.9%	23.3%
White	14.7%	69.4%	72.9%	77.8%
Black	78.3%	22.2%	7.1%	5.5%
Other race	7.0%	8.4%	20.0%	16.7%
Obtained BS or BA degree	32.4%	16.7%	28.6%	32.2%
Median Income	$33,965	$63,474	$60,931	$67,896
Income ≥ $100,000	10.5%	29.8%	27.3%	31.8%
Population	670,052	4,319,629	1,680,988	4,948,203

[a]Metropolian Statistical Area for Detroit-Warren-Dearborn, MI Metro Area
[b]Metropolian Statistical Area for Phoenix-Mesa-Chandler, AZ Metro Area

Table A3. Fit Statistics Associated with Competing Models for 2017 Samples Where Data are Combined Across Locations

Model	# parms	Pseudo R2	AIC	BIC	% Correct Predictions	Size of smallest class
MNL constants only	4	---	29257.3	29285.9	58.9%	---
MNL	7	0.24	22124.8	22174.9	58.9%	---
LCM-2 classes	15	0.30	20628.4	20735.7	64.9%	0.15
LCM-3 classes	23	0.35	19096.6	19261.1	66.4%	0.14
LCM-4 classes	31	0.36	18758.5	18980.2	71.5%	0.11

Table A4. Latent Class Model Estimates Fit to Pooled 2017 data (N=1,046 individuals; 9,494 choices) Used as Priors to Project Willingness-to-Pay in Feedback Questions in 2020

Variable	**Class 1**	**Class 2**	**Class 3**	**Class 4**
Price	-0.816	-1.348	-0.736	-0.088
	(0.066)	(0.051)	(0.043)	(0.021)
Farmers market	-0.975	0.052	0.488	-0.242
	(0.174)	(0.106)	(0.124)	(0.077)
Urban farm	-1.329	-0.522	-0.095	-0.498
	(0.178)	(0.093)	(0.126)	(0.078)
Organic	-0.476	0.017	1.507	-0.169
	(0.132)	(0.076)	(0.122)	(0.062)
Local	0.362	0.304	1.157	0.14
	(0.124)	(0.087)	(0.109)	(0.061)
Travel time	-0.116	-0.123	-0.174	-0.024
	(0.01)	(0.006)	(0.009)	(0.004)
None	-2.352	-6.686	-5.857	-2.593
	(0.216)	(0.228)	(0.375)	(0.164)
Class probability	0.112	0.405	0.271	0.211
	(0.012)	(0.025)	(0.024)	(0.015)

Table A5. Comparison of Parameter Means and Standard Deviations from Unconditional and Conditional Posterior Distributions from 2017 Data

Parameters	**Unconditional Parameters Implied by Table A4**	**Conditional Posterior "Individual Specific" Parameters**
Means		
Price	-0.856	-0.856
Farmers market	-0.007	-0.007
Urban farm	-0.491	-0.491
Organic	0.327	0.327
Local	0.508	0.508
Travel time	-0.115	-0.115
None	-5.111	-5.111
Standard Deviations		
Price	0.476	0.428
Farmers market	0.430	0.399
Urban farm	0.349	0.322
Organic	0.735	0.627
Local	0.403	0.336
Travel time	0.052	0.047
None	1.830	1.700

Table A6. Fit Statistics Associated with Competing Models Estimated with 2020 Data

Model	# parms	Phoenix Pseudo R2	Phoenix AIC	Phoenix BIC	Phoenix % Correct Predictions	Detroit Pseudo R2	Detroit AIC	Detroit BIC	Detroit % Correct Predictions
MNL constants only	4	---	11704.3	11729.2	59.7%	---	12443.3	12443.3	57.7%
MNL	7	0.22	9109.1	9152.7	59.7%	0.19	10039.6	10083.7	57.7%
RPL (price & time constant)	12	0.31	8120.5	8195.3	59.7%	0.25	9412.0	9487.6	57.6%
RPL (price & time lognormal)	14	**0.37**[a]	**7374.2**[a]	**7461.4**[a]	59.7%	0.31	8575.2	**8663.5**[a]	57.8%
RPL (price & time triangular)	12	0.35	7657.6	7732.3	59.7%	0.29	8841.1	8916.8	58.3%
LCM-2 classes	15	0.29	8317.6	8411.0	66.9%	0.26	9196.9	9291.5	58.3%
LCM-3 classes	23	0.34	7743.6	7886.8	70.5%	0.30	8712.5	8857.5	63.9%
LCM-4 classes	31	0.36	7533.9	7726.8	**75.0%**[a]	0.32	8545.9	8741.3	71.3%
LCM-5 classes[c]	39	n/a[b]	n/a	n/a	n/a	**0.33**[a]	**8459.0**[a]	8704.8	**75.6%**[a]
LCM-6 classes	47	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

[a]Best fitting model according to criteria in the respective column.
[b]n/a implies the model didn't converge or had parameters or standard errors that were not identified.
[c]The 5-class LCM model in Detroit resulted in one class with a positive price parameter.

Table A7. Multinomial Logit and Willingness-to-Pay Estimates by Treatment

	Pooled	Pooled Phoenix	Pooled Detroit	2017			2020[e]		
				Pooled	Phoenix	Detroit	Pooled	Phoenix	Detroit
Price	-0.550	-0.569	-0.532	-0.580	-0.595	-0.564	-0.516	-0.539	-0.498
	(0.007)[a]	(0.011)	(0.010)	(0.010)	(0.014)	(0.014)	(0.010)	(0.016)	(0.014)
Farmers	-0.146	-0.159	-0.135	-0.100	-0.106	-0.090	-0.198	-0.218	-0.180
market[b]	(0.027)	(0.038)	(0.037)	(0.037)	(0.052)	(0.051)	(0.039)	(0.056)	(0.054)
Urban	-0.468	-0.518	-0.421	-0.445	-0.445	-0.441	-0.491	-0.608	-0.390
farm[b]	(0.025)	(0.037)	(0.035)	(0.035)	(0.050)	(0.049)	(0.037)	(0.055)	(0.051)
Organic[c]	0.219	0.268	0.172	0.190	0.261	0.122	0.253	0.276	0.231
	(0.021)	(0.030)	(0.028)	(0.028)	(0.040)	(0.039)	(0.030)	(0.044)	(0.041)
Local[c]	0.329	0.360	0.301	0.344	0.373	0.317	0.314	0.344	0.288
	(0.023)	(0.033)	(0.032)	(0.032)	(0.045)	(0.044)	(0.034)	(0.050)	(0.046)
Travel	-0.076	-0.078	-0.075	-0.082	-0.082	-0.082	-0.069	-0.074	-0.066
time	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
None	-3.306	-3.199	-3.430	-3.428	-3.260	-3.611	-3.165	-3.126	-3.226
	(0.044)	(0.062)	(0.062)	(0.060)	(0.084)	(0.085)	(0.064)	(0.091)	(0.091)
Willingness-to-Pay									
Farmers	-$0.27	-$0.28	-$0.25	-$0.17	-$0.18	-$0.16	-$0.38	-$0.41	-$0.36
market[b]	[-0.36, -0.18][d]	[-0.41, -0.15]	[-0.39, -0.12]	[-0.29, -0.06]	[-0.34, -0.02]	[-0.33, 0.01]	[-0.52, -0.25]	[-0.61, -0.20]	[-0.57, -0.15]
Urban	-$0.85	-$0.91	-$0.79	-$0.77	-$0.75	-$0.78	-$0.95	-$1.13	-$0.78
farm[b]	[-0.93, -0.77]	[-1.03, -0.79]	[-0.92, -0.66]	[-0.88, -0.65]	[-0.90, -0.59]	[-0.96, -0.61]	[-1.08, -0.82]	[-1.32, -0.94]	[-0.98, -0.59]
Organic[c]	$0.40	$0.47	$0.32	$0.33	$0.44	$0.22	$0.49	$0.51	$0.46
	[0.32, 0.47]	[0.37, 0.57]	[0.22, 0.42]	[0.24, 0.42]	[0.31, 0.57]	[0.08, 0.35]	[0.37, 0.61]	[0.35, 0.67]	[0.30, 0.63]
Local[c]	$0.60	$0.63	$0.57	$0.59	$0.63	$0.56	$0.61	$0.64	$0.58
	[0.52, 0.68]	[0.52, 0.74]	[0.45, 0.68]	[0.49, 0.7]	[0.48, 0.78]	[0.41, 0.71]	[0.47, 0.75]	[0.44, 0.84]	[0.39, 0.76]
Travel	-$0.14	-$0.14	-$0.14	-$0.14	-$0.14	-$0.15	-$0.13	-$0.14	-$0.13
time	[-0.14, -0.13]	[-0.15, -0.13]	[-0.15, -0.13]	[-0.15, -0.13]	[-0.15, -0.13]	[-0.16, -0.13]	[-0.14, -0.13]	[-0.15, -0.12]	[-0.15, -0.12]
N Choices	17190	8451	8739	9414	4716	4698	7776	3735	4041
N people	1910	939	971	1046	524	522	864	415	449
Log LF	-20653.4	-10096.4	-10530.6	-11055.4	-5540.0	-5501.1	-9576.4	-4547.5	-5012.8

[a]Numbers in parentheses are standard errors; [b]Effect relative to shopping in grocery store; [c]Effect relative to no label present; [d]Numbers in brackets are 95% confidence intervals determined by the Krinsky-Robb method. [e]Data from 2020 are weighted to match demographics of Pre-COVID samples

Table A8. Results of Likelihood Ratio Tests from Multinomial Logit Models

Hypothesis	χ^2	**p-value**
Parameters equal for both locations and both time periods	103.88	<0.001
Parameters equal for both locations (time periods pooled)	52.80	<0.001
Parameters equal for both locations in 2017	28.52	<0.001
Parameters equal for both locations in 2020	32.16	<0.001
Parameters equal in 2017 and 2020 (locations pooled)	43.20	<0.001
Parameters equal in 2017 and 2020 in Phoenix	17.64	0.014
Parameters equal in 2017 and 2020 in Detroit	33.44	<0.001