Information Videos Mitigate Hypothetical Bias in Discrete Choice Experiments

Wen Lin, China Academy for Rural Development, School of Public Affairs, Zhejiang University, Hangzhou, China, 310058. Email: linwen2@zju.edu.cn.

Rodolfo M. Nayga, Jr, [1]Department of Agricultural Economics, Texas A&M University, College Station, TX, USA, 77843. [2]Department of Food and Resource Economics, Korea University, Seoul, South Korea, 02841. Email: rudy.nayga@ag.tamu.edu.

David L. Ortega, Department of Agricultural, Food, and Resource Economics, Michigan State University, East Lansing, Michigan, USA, 48910. Email: dlortega@msu.edu.

Weibin Hu*, China Academy for Rural Development, School of Public Affairs, Zhejiang University, Hangzhou, China, 310058. Email: winsbon@zju.edu.cn.

* Correspondening author.

Abstract

Efforts to reduce hypothetical bias in stated preference methods often focus on the content of ex-ante information, while this study examines the impacts of presentation formats on hypothetical

doi:10.3368/le.101.4.072424-0065R1

bias mitigation, focusing on cheap talk scripts and product explanations. Using beef alternatives characterized by attributes such as carbon label, plant-based label, favor and price, we find that video presentations result in hypothetical willingness-to-pay values closer to non-hypothetical scenarios, particularly among new and unfamiliar attributes. Subjects exposed to video information demonstrate an improved understanding of hypothetical bias and product attributes, and those with lower cognitive skills might more responsive to video displays.

Keywords: Hypothetical bias, video, choice experiment, willingness to pay

JEL classifications: D12, Q18, C25

1. Introduction

Hypothetical bias, the difference between stated and real values, is an important issue that affects the legitimacy of stated preference methods and the usefulness of those methods in policy-making and cost-benefit analysis across several economic fields (e.g. environment and natural resources, food, health and transport). The challenge of creating incentive-compatible valuation tasks, alongside the appeal and/or necessity of employing stated preference questions in certain scenarios, has motivated researchers to explore which *ex-ante* information, delivered beforehand, could reduce hypothetical bias in stated preference methods. Much emphasis has been placed on examining the content of information, including cheap talk scripts (Cummings and Taylor, 1999; Carlsson et al., 2005), consequentiality scripts (Vossler and Watson, 2013; Wei and Khachatryan, 2023), honesty priming (De-Magistris et al., 2013; Howard et al., 2017), and opt-out option reminders (Ladenburg and Olsen, 2014). Relatively less investigated is whether information presentation formats matter for hypothetical bias mitigation.

Information can be viewed as a heterogeneous good whose value depends on presentation formats, considering the behavioral insight that decision-making is context-dependent. Visual information is more salient than text (Li and Camerer, 2022). Meanwhile, dynamically displaying a concept compensates for individuals' limitations in imagination (Salomon, 1979) and reduces cognitive load by allowing people to process information using fewer elements in working memory (Carlson et al., 2003; Höffler and Leutner, 2007). Relative to text formats, presenting information in video format enhances attentiveness and engagement, leading to effective processing and greater value of the information (Shepard, 1967; Mayer and Moreno, 2003; Castro-Alonso et al., 2019).

Given that stated preference methods are usually questioned by hypothetical bias, a strand of literature has emerged to assess the effectiveness of various techniques designed to mitigate such bias, such as cheap talk scripts (Cummings and Taylor, 1999), alternative measures include oath (Jacquemet et al., 2013), honesty priming (De-Magistris et al., 2013; Howard et al., 2017), Bayesian truth serum (Menapace and Raffaelli, 2020), inferred valuation (Lusk and Norwood, 2009; Entem et al., 2022), and consequentiality scripts (Vossler and Watson, 2013; Carson et al., 2014). These measures are centered on crafting the information content provided to individuals before DCE tasks, aiming to mitigate hypothetical bias. Unlike their approach, our research keeps information content and choice experimental design consistent while experimenting with different presentation formats across hypothetical DCE groups. We are novel in examining whether video displays of the cheap talk script and production explanations in hypothetical DCE would yield similar preferences under non-hypothetical scenarios.

Specifically, this study tests which format scheme when used to present ex-ante information in stated preference methods, results in mitigated hypothetical bias. Utilizing discrete choice experiments (DCEs), the study focuses on two pieces of information commonly presented to individuals before engaging in hypothetical choice tasks: a cheap talk script and the introduction/explanation of products or services of interest. The DCE focuses on a well-researched product, beef, that is characterized by alternatives that include attributes such as carbon label, plant-based label, favor and price. The attributes under consideration provide both public and private benefits. Our design adopts a between-subject design, including one non-hypothetical DCE group and four hypothetical DCE groups. In the non-hypothetical group, individuals receive product attribute explanations and then make incentive-compatible choices. The second group differs only in the use of hypothetical DCE choices. The remaining three groups involve hypothetical DCEs. The third group presents the cheap talk script in video format and retains the product attribute explanation in text. In the fourth group, the product attribute explanation is conveyed via video while the cheap talk script remains in text. In the fifth group, the two pieces of information are presented in video format. Table 1 provides an overview of the five groups, with all other aspects, including choice tasks and information content, identical across groups. The difference in willingness to pay (WTP) between the first and second groups establishes a benchmark for assessing the effects of video information on hypothetical bias. A comparison of this benchmark and the difference in WTP between the first and third groups determines if a video display of the cheap talk script is effective in mitigating hypothetical bias. Likewise, if the difference in WTP between the first and fourth groups is smaller than the benchmark, it suggests a larger reduction of hypothetical bias with video product explanation. Comparing the benchmark and the difference in WTP between the first and fifth groups inform the effects of jointly presenting the cheap talk script and the product explanation in video.

Our findings reveal varying degrees of hypothetical bias across all hypothetical DCE groups, confirming the persistent presence of hypothetical bias (Penn and Hu, 2018) and highlighting the need for hypothetical bias mitigation. When compared to hypothetical DCEs with text displays of information, DCEs incorporating video displays of cheap talk script or/and product explanation elicit most WTPs that are closer to those in the non-hypothetical DCE. This implies that, despite identical content, the use of video displays to present ex-ante information reduces hypothetical bias. That is, the presentation format of ex-ante information indeed matters in mitigating hypothetical bias. Additionally, hypothetical bias is higher for the carbon label compared to the plant-based label across all hypothetical treatments. The plant-based label, offering a private benefit, shows a more notable reduction in hypothetical bias, while the carbon label, which provides a public benefit, exhibits less fluctuation in hypothetical bias, indicating that our video treatment effects are more pronounced for private valuations than for public benefits. We conduct two additional analyses. First, responses to follow-up questions after the choice tasks indicate that subjects in the hypothetical DCE groups exposed to information videos demonstrate a more accurate understanding of hypothetical bias and product attributes, potentially contributing to alleviating hypothetical bias. Second, heterogeneity analysis finds that individuals with high cognitive skills, as measured using the Raven Advanced Progressive Matrices test (Raven, 1938), tend to be more responsive to video displays of information, showing lower hypothetical bias compared to those with lower cognitive skills.

Several studies have highlighted differing preferences or even a reduction in hypothetical bias through visualizing choice tasks in hypothetical DCEs (Bateman et al., 2009; Matthews et al., 2017; Fang et al., 2020; Guilfoos et al., 2023). However, we emphasize video displays presenting information to individuals before DCE tasks rather than altering the tasks themselves. Two closely related studies used video or audio to deliver the cheap talk script and compared respondents' attention and WTP with those in the text-based cheap talk group (Penn and Hu, 2021; Agossadou and Nayga, 2023). Our study differs by incorporating a non-hypothetical choice experiment and directly comparing WTPs between non-hypothetical and hypothetical valuations.

We also relate to some meta-analyses showing that hypothetical bias is greater when valuing public goods, all else being equal (List and Gallet, 2001; Murphy et al., 2005; Penn and Hu, 2018). In our study, the carbon label represents a public good attribute, while the plant-based label aligns more with a private good attribute. Our results confirm these findings, showing greater hypothetical bias for the carbon label across all treatments. Several factors may explain this discrepancy in hypothetical bias between public and private goods. First, familiarity and experience play a role, as individuals are more comfortable valuing goods they commonly purchase, leading to fewer errors in valuation (List and Gallet, 2001). This is supported by evidence that people often misjudge the costs of providing public goods in stated preference surveys (Lang et al., 2024). Second, social desirability bias could inflate willingness-to-pay (WTP) for public goods, as respondents may feel pressure to respond in a way that aligns with socially or environmentally responsible behaviors (Entem et al., 2021). This bias is less likely to affect private goods, where decisions are more personal.

This study contributes to the broader literature that investigates the effects of informational nudges on individual behaviors. A particular focus is on identifying the effective information content for behavioral change, such as gain and loss framing of messages (Hallsworth et al., 2017; Balew et al., 2023), scientific and narrative framing (Dahlstrom, 2014; Yang and Hobbs, 2020), positive and negative framing (Andreoni, 1995). We relate to some studies that have emerged to investigate how should messages be diffused to this end. Bahety et al. (2021) discovered that an SMS-based information nudge has minimal impact on knowledge or adaptation of preventive health behavior in India during the COVID-19 pandemic. Under the same study context, in contrast, Banerjee et al. (2020) propose that the video mode of information communication is effective.

There is also a growing literature that evaluates consumer preferences and demand for beef alternatives, considering options that exclude animal products or address some of the environmental concerns occurring throughout the supply chain. Focused primarily on developed markets and hypothetical choice scenarios, studies yield mixed findings on individual valuations for plant-based beef and lower carbon footprint certified beef products (Li et al., 2016; Slade, 2018; Van Loo et al., 2020; Kilders and Caputo, 2024). We add to this literature by including an incentive-compatible DCE to elicit consumer price premiums for plant-based and low-carbon beef products in the emerging Chinese market. Given that the food system is responsible for one-third of the global anthropogenic greenhouse gas emissions (Crippa et al., 2021), promoting sustainable meat alternatives is important for the world's largest consumer market (Ortega et al., 2022).

2. Design

Discrete choice experiment design

We design a beef jerky product choice experiment to compare WTP estimates and hypothetical bias across treatments. We focus on beef products for several reasons. Beef is a commonly investigated food item in studies utilizing stated preference methods (Lin et al., 2023), and recent DCE studies investigating methodological issues have selected beef as their experimental products (Lim and Hu, 2023; Kliders and Caputo, 2024), so we selected beef to to inform this body of literature. In addition to its significance in research, beef consumption in China has experienced a substantial increase, reaching 7.7 million tons in 2019, making a 55% growth from 2000. This surge accounted for 11% of the global beef consumption. Projections indicate that China's beef consumption is expected to grow by 12% to 39% in the next 10 years, and by 19 to 71% in 15 years, depending on income levels (Zhu et al., 2021). Among different beef products, we opt to use beef jerky given that it is a popular and convenient snack for our study sample. Unlike fresh beef, beef jerky is typically consumed as a snack by college students, as confirmed by our pre-test.

The experimental design is developed based on our research objectives and previous DCE studies. One primary objective of this study is to examine whether an exogenous manipulation that affects information processing could mitigate hypothetical bias. Compared to familiar product attributes, respondents may need to allocate more cognitive resources for understanding when dealing with novel ones. We thus include novel product attributes in the design, especially whether the beef product is plant-based and whether the beef has a lower carbon label. Several demand-side studies indicate that individuals prefer traditional beef to plant-based beef but value environmentally-friendly labels on meat products, albeit results exhibit significant heterogeneity across consumers and certification types (Grebitus et al., 2016; Meyerding et al., 2019; Van Loo et al., 2020; Ortega et al., 2022; Lin et al., 2022). Given this variability in individual preferences and limited focus on the Chinese population, the market success of plant-based beef products or beef products certified under a low-carbon verification scheme in China is worth investigating. The carbon label is related to the natural environment because it informs consumers about the carbon emissions associated with a product, and it generates positive externalities by reducing carbon emissions and provides a public benefit. Two other product attributes, familiar to consumers, have also been selected to describe beef jerky products: flavor and price. Indicating the flavor of a food snack is typically observed on the packaging as it provides a private benefit to consumers, and we set two levels for the taste attribute: spicy and original flavor. The price levels in this experiment are representative of the actual market price range for beef jerky. Actual prices of real and planted-based beef jerky were sourced from large online retailers like JD and Alibaba, as well as from offline retailers such as supermarkets and convenience stores, ensuring a realistic representation of prices. In line with the majority of DCE studies that were designed with four price levels (Caputo and Scarpa, 2022), we implement four different prices ranging from 9 RMB to 22 RMB per 50 grams. Appendix Table A1 presents the attributes used in the DCE.

Given the selected attributes and attribute levels, a full-factorial design with two product alternatives in a single choice set would require 1,024 ($4^{2*1}*2^{2*3}$) choice questions. Following this, an orthogonal optimal design is applied to substantially reduce the number of choice questions, resulting in a total of 16 choice questions (Appendix Table A2 shows the 16 choice tasks and corresponding product alternatives) and a D-optimality of 97.21%. The DCE design, and thus the choice tasks, remain constant across all treatments. Therefore, the difference in WTPs between two treatments are not attributed to variations in our experimental design. Each choice task includes two product alternatives and one opt-out option (see a sample choice set in Figure 1).

Treatment design

Our experiment employs a between-subject design with five treatments. In each treatment, the subject completes a survey in exchange for money.

The treatments vary in the binding nature of DCE questions and the form in which cheap talk scripts and product attribute explanations are displayed. In a *Real Treatment* (*Real*), individuals engage with non-hypothetical DCE questions where they are explicitly informed that each choice task had an equal probability of becoming their actual purchase, in line with incentive-compatible value elicitation designs (La Nauze and Myers, 2023). They are obligated to purchase the product they selected in the binding choice task, at the specified price. This group is referred to as the "Real Treatment" because when respondents believe that their answers will be consequential with any positive probability, their dominant strategy is to answer truthfully (Herriges et al., 2010; Vossler et al., 2012).

The four remaining treatments all involve hypothetical DCE with non-binding questions, where we provide subjects with a cheap talk script and product attribute explanation before choice sets. We exogenously manipulate cognitive burden across the four treatments by introducing either text displays or video displays of cheap talk scripts and product attribute explanations. Note that the wording of the cheap talk script and product attribute explanation is identical across treatments. In *a Both Text Treatment* (*Both_text*), individuals are presented with text displays of a cheap talk script as well as product attribute explanation. In *a Video Cheap Talk Treatment* (*V_cheap*), we present the cheap talk script in video format and retain the product attribute explanation in text format. In *a Video Instruction Treatment* (*V_instruc*), the product attribute explanation is conveyed through video while the cheap talk script remains in text. In *a Both Video Treatment* (*Both_video*), both the cheap talk script and product attribute explanation are presented via video format. In video treatments, respondents are required to view the video(s) before proceeding; in text treatments, they are given one minute to read the text beforehand. Appendix B details the text and video for the cheap talk script and product attribute explanation.

The difference in WTP between Real and Both_text establishes a benchmark for assessing hypothetical bias, which we use as a control in our experiment. A comparison of this benchmark and the difference in WTP between Real and V_cheap determines if a video display of the cheap talk script is more effective in mitigating hypothetical bias, relative to the text display. Likewise, a comparison of the benchmark and the difference in WTP between Real and V_intruc determines whether hypothetical bias would be reduced by the Video Instruction Treatment. Comparing the benchmark and the difference in WTP between Real and Both_video assesses the joint treatment effects of using video displays for both the cheap talk script and the product attribute explanation.

Questionnaire design

The questionnaire used in this study consists of three parts: (1) Cognitive ability test; (2) beef jerky DCE (see section 2.2), followed by the collection of subjective and objective understanding related to hypothetical bias and DCE attributes; and (3) questions on socio-demographics and beef purchasing habits, whose variables are presented in Table 1. The DCE choice tasks and DCE follow-up questions are presented in Appendix A. As we aim to investigate whether individuals with differing cognitive abilities respond differently to our treatments, we also assess the cognitive ability of each respondent using the Raven Advanced Progressive Matrices test (Raven, 1938), hereafter, the Raven test. The Raven test is a leading non-verbal measure of cognitive skills comprised entirely of pictorial questions related to spatial reasoning and pattern matching, which has been extensively employed in cross-cultural contexts (Boissiere et al., 1985; Bartling and

Netzer, 2016; Zhao et al., 2019). The Raven test presents 12 problems that do not rely on knowledge or verbal skills. Each item features a 3×3 matrix of abstract figures with one figure missing. Subjects must identify the missing figure from a selection of eight possible solutions. This task demands pattern recognition across rows and columns. The problems become more difficult throughout the 12 items. Before beginning the test, participants are asked to solve two practice items to confirm their understanding of the test.

3. Data

In June and July 2023, we recruited students to a lab on the campus of a university in China[1]. Using a student sample in the field of experimental economics is a common practice. Studies have found that the behavior of students closely mirrors that of non-students in non-hypothetical DCE questions (Depositario et al., 2009) and in a battery of behavioral attributes (Snowberg and Yariv, 2021). Eligibility criteria included being at least 18 years old and having either consumed beef jerky in the past three months or being willing to do so. In Real, each respondent is asked for and consents to make actual purchases. After participants completed the discrete choice experiments and questionnaires offline in a laboratory setting, they received a payment of 25 RMB (1 USD=7.09 RMB at the time of the data collection), regardless of group. We collected a total of 488 completed responses, which were randomly assigned into one of the five treatments (Real, Both_text, V_cheap, V_instruc, and Both_video). The distribution of participants across treatment is reported in Appendix Table C1. In determining our sample size, we followed sample size requirements for DCE from Lancsar and Louviere (2008) and de Bekker-Grob et al. (2015), as well as insights from recent studies using a between-subject design like Xie et al. (2022). While

63 participants per group would have sufficed, we chose approximately 90 to ensure robustness in our analysis.

Table 1 provides summary statistics of socio-demographic, beef jerky consumption and cognitive ability variables of the participants. Across all groups, the female proportion averaged between 67.3% and 72.3%. Participants' ages averaged around 22 years, primarily consisting of undergraduate students pursuing diverse majors. An average respondent purchased meat jerky less than weekly, spending between 50-100 RMB per purchase. Flavor preferences were evenly split, with half favoring original flavor and half preferring spicy jerky. In terms of cognitive ability, the average individual solved 9 cognitive exercises correctly, with the mode being 10. The mean accuracy rate of the Raven test is 75%, consistent with findings from prior studies on Chinese college students (Zhu et al., 2017). Besides, the randomization of blocks within each group works well. On average, approximately 51% of the sample belongs to Block 1. Importantly, chi-squared tests suggest that all the variables in Table 1 are balanced across treatments, and joint orthogonality tests (McKenzie, 2015) show that all variables are jointly orthogonal concerning treatments.

4. WTP between treatment

We design a choice experiment survey to assess preferences regarding beef jerky products and estimate marginal utilities. In each choice scenario, respondents have the opportunity to select one of three options: the status quo and two beef jerky products. Following Random Utility Theory, Train and Weeks (2005) and Scarpa et al. (2008), the utility of individual i choosing an alternative j in choice scenario t in the willingness to pay space:

$$U_{ijt} = \sigma_i'(-P_{ijt} + \boldsymbol{\beta}_i' \boldsymbol{X}_{ijt}) + \varepsilon_{ijt} \qquad (1)$$

whereP_{ijt} is the price presented to individual i for alternative j in choice scenario t, $\boldsymbol{X_{ijt}}$ denotes the non-price product attribute vector (carbon label, plant-based beef, spicy flavor), ε_{ijt} is an *i.i.d.* extreme value type one error term with individual-specific scale σ_i. $\boldsymbol{\beta_i'} = \boldsymbol{b_i'}/\alpha$ is the vector of marginal WTPs for beef jerky attributes, for example, the marginal WTP for plant-based beef jerky relative to the conventional beef jerky (for simplicity, we refer marginal WTP as WTP)). $\sigma_i' = \sigma_i\alpha$ is the scale parameter which also represents the price coefficient (Scarpa et al., 2008). We consider random parameter specifications that account for preference heterogeneity and specify normal distributions for the non-price attribute parameters, $\boldsymbol{\beta_i'}$, and a log-normal distribution for σ_i'. In particular, $\sigma_i' = exp[\bar{\sigma} + \tau w_i]$, where w_i follows a standard normal distribution. τ, $\bar{\sigma}$, $\boldsymbol{\beta_i'}$ are coefficients to be estimated.

Table 2 presents the mixed logit model estimation results in WTP space for each of the five treatments. Looking at the first two columns of Table 2 which provides the WTP outcomes of the Real treatment, not surprisingly, we find that the WTPs in Real produce the smallest values in all treatments, implying that subjects in the four hypothetical experiments tend to overstate their valuations. In particular, in Real, the average valuation of plant-based beef jerky is 5 RMB per 50 grams lower than that of traditional beef jerky products. The mean WTP of low-carbon labeled beef jerky is approximately 1.3 RMB per 50 grams, meaning that participants are willing to pay a modest price premium for low-carbon products. Neither the preference for spicy flavor nor original flavor is pronounced in Real, with 48% of the participants favoring original beef. The remaining columns in this table report WTP estimates from treatments involving hypothetical DCEs. Relative to traditional beef jerky, the mean valuation of plant-based beef jerky is reduced by 7.6 RMB, 7 RMB, 6.8 RMB, and 5.7 RMB per 50grams in Both_text, V_cheap, V_instruc and Both_video, respectively. Conversely, the mean WTP for low-carbon labeled beef jerky increases by around 2 RMB per 50 grams in these four treatments, compared to unlabeled beef products. Similar to Real, the average WTP for spicy flavor is statistically indifferent from zero in the four treatments.

5. WTP in the pooled sample

Despite some suggestive evidence in the previous section that video treatment reduces hypothetical bias, in this analysis we generate dummy variables to indicate the treatment group and adopt a pooled sample approach to account for more precise differences in WTP between treatments. The utility in WTP space:

$$U_{ijt} = -\sigma_i' P_{ijt} + \sigma_i' \boldsymbol{X_{ijt}}' (\boldsymbol{\beta_{i0}} + \boldsymbol{\beta_{i1}} Both_{text} + \boldsymbol{\beta_{i2}} V_{cheap} + \boldsymbol{\beta_{i3}} V_{instruc} + \boldsymbol{\beta_{i4}} Both\_video) + \varepsilon_{ijt} \qquad (2)$$

where Both_text, V_cheap, V_instruc, and Both_video are dummy variables and equal to 1 if individual i belongs to the corresponding group, and zero otherwise. $\boldsymbol{\beta_{i0}}$ are the willingness to pay value in Real, which serves as the reference group in Equation (2). $\boldsymbol{\beta_{i1}}, \ldots \boldsymbol{\beta_{i4}}$ capture the differences in willingness to pay between Real and a hypothetical treatment (Both_text, V_cheap, V_instruc, and Both_video, respectively), namely hypothetical bias.

Table 3 reports the mixed logit estimation result of Equation (2), where Model 1 lets the difference in WTP between Real and other groups be fixed. First, the WTP values of Real in the model are close to those in Table 2, suggesting the robustness of our results. Second, looking at Model 1, we find that in general, the extent of hypothetical bias for plant-based and low-carbon labeled beef is most evident, while there is little difference in the WTP for spicy beef between Real and any hypothetical treatment. On average, individuals in Both_text tend to exaggerate their valuation, overstating by approximately 2.7 RMB for plant-based beef jerky, 1 RMB for low carbon beef jerky. Moreover, the extent of hypothetical bias varies across treatment and product attributes. In terms of plant-based beef, Both_text exhibits strongest hypothetical bias, followed by V_cheap and V_instruc, while the value of Both_video is indifferent from that of Real. Regarding the WTP for carbon labeled beef, Both_text, V_cheap and V_instruc show similar degrees of hypothetical bias. Relative to Real, participants of Both_text are willing to pay an additional amount of 1 RMB per 50 grams for low-carbon beef products (0.75 RMB in V_cheap, 1 RMB in V_instruc). Last, for all the three WTP values, there exists insignificant difference between Real and Both_video.

To explicitly assess the treatment effects of video displays, we compare the WTPs across the four hypothetical treatments following Poe et al. (2005) and Lim and Hu (2022). While WTPs for plant-based and carbon labeled beef jerkey are overstated in the Both_text, V_cheap, and V_intruc, as indicated by the significant interaction terms in Table 3, the overstatement is less pronounced in treatments involving video displays compared to text-based information (Both_text) (Hypotheses 1-3 in Appendix D). Besides, our analysis does not yield statistical evidence to reject the null hypotheses of equal WTP between V_cheap and V_instruc (Hypothesis 4). But hypothesis testing indicates significantly lower WTPs for plant-based and spicy beef products in Both_video compared to V_cheap or V_instruc. Thus, the treatment effect of Both_video in mitigating hypothetical bias is the most pronounced, followed by V_cheap and V_instruc with a weaker effect.

Besides, we calculate the degree of hypothetical bias for plant-based label and carbon label across treatments (Appendix Table E1). The video treatments reduce hypothetical bias the two product attributes more effectively, compared to the Both_text treatment, consistent with the above results. Hypothetical bias is greater for the carbon label than for the plant-based label across all hypothetical treatments. The plant-based label, which provides a private benefit, shows a more significant reduction in hypothetical bias across treatments, with the decrease from 1.505 in Both_text to 1.133 in Both_video. In contrast, the carbon label, which provides a public benefit , exhibits less variation in hypothetical bias, with values ranging from 1.570 to 1.707 across treatments. This suggests that our video treatment effects are more pronounced for private valuations than for public benefits.

Last, we conduct two robustness checks to guard our results. Model 2 of Table 3 enables the WTP difference between Real and hypothetical treatments to be random and normally distributed. Results from Model 2 are largely consistent with those obtained from Model 1, where WTP differences are assumed to be fixed, with some exceptions. For example, the coefficient of Carbon × V_instruc in Model 1 is 1.007. In Model 2, this coefficient has a mean of 1.056 and a standard deviation of 1.267, and the probability of it being smaller than 0 is 0.2. This indicates that while V_instruc generally exhibits upward hypothetical bias, there is also a chance of encountering downward hypothetical bias, reflecting some variability in the treatment's impact.

Additionally, we estimate a generalized multinomial logit model by Fiebig et al. (2010) which accommodates scale heterogeneity. Capturing the treatment's impacts on the variance of the scale parameter serves a dual purpose, it helps disentangle the effects of the scale parameter and the treatment (Lim and Hu, 2021). On the other hand, since a larger scale parameter implies that choices are less stochastic or more deterministic, the specification enables us to investigate whether video displays affect individual choice errors. Arentze et al. (2003) found no impact of image presentations on the scale factor while Matthews et al. (2017) observed reduced choice variability within the virtual reality treatment. Our study adds to the limited literature on the effects of display formats on scale factors. Table 4 provides the GMNL model estimation results using the pooled sample. All the treatment coefficients, θ', are statistically insignificant, indicating that any differences between hypothetical treatments and Real are attributed to the difference in individual preferences, instead of the scale[2].

6. Heterogeneity across subject's cognitive ability

In this section, we explore the mechanism through which video information mitigates hypothetical bias in two ways. First, we examine both the subjective and objective understanding of hypothetical bias and DCE product attributes across treatments, as outlined in Table 5. We anticipate that a diminished cognitive burden in treatments incorporating video displays would enhance information processing (Khaw et al., 202l; Ball et al., 2023). Results from Table 5 suggest that there exists little difference in self-reported understanding regarding hypothetical bias or DCE product attributes between treatments. However, in the assessment of hypothetical bias definition, individuals in V_cheap and Both_video demonstrate higher correct responses compared to Both_text. Specifically, the correct understanding of hypothetical bias is observed in nearly 95% of participants in Both_text, while the correct rate is 99% in V_cheap and Both_video. Shifting the focus to the definition of low carbon labeling, 90% of participants in Both_text exhibit a correct understanding of the label, a proportion significantly lower than that in V_instruc (97%) and Both_video (100%). These findings highlight that, in comparison to Both_text, hypothetical treatments with video displays enhance individual understanding of cheap talk scripts and product attributes, which would contribute to the mitigation of hypothetical bias.

Additionally, we investigate heterogeneity in results across individuals with low versus high cognitive ability in Table 6. Individuals are categorized into two groups based on their Raven test scores, with those scoring above the sample median being the high cognitive ability subsample. Those scoring above the sample median constitute the high cognitive ability subsample (Above_Raven=1), while others are in the low cognitive ability subsample (Above_Raven=0). We introduce interaction terms by multiplying the variable, Above_Raven, with each of the four hypothetical treatment dummies. Subsequently, we conduct a regression analysis where the four interaction terms and four hypothetical treatment dummies are regressed on conditional mean WTPs (or individual-specific mean WTP conditional on observed choices (Hess and Rose, 2007), from Table 2) for low-carbon beef jerky and plant-based beef jerky, whose results are in columns (1) and (2) of Table 6, respectively. First, the coefficients of the hypothetical treatment dummies in Table 6 indicate that the difference between Real and each of the hypothetical treatments. Importantly, the coefficients associated with the four interaction terms, with statistical significance of at least 10%, suggest low cognitive ability subsample is more responsive to video displays. For the WTP for low-carbon beef jerky, the interaction term coefficients between Above_Raven and each of Both_text, V_cheap, V_instruc, and Both_video are -0.858, -0.249, -0.321, and -0.232, respectively. Similarly, for the valuation of plant-based beef jerky, the coefficients for Both_text×Above_Raven, V_cheap×Above_Raven, V_instruc×Above_Raven, and Both_video×Above_Raven are 1.093, 0.660, 0.841, and 0.324, respectively. Across the four treatments incorporating hypothetical DCEs, the treatment effects are smaller among high-cognitive individuals, implying that text information is sufficient for high-cognitive individuals. This implies that limited cognitive ability may be a source of hypothetical bias in DCEs.

7. Discussions and Conclusions

In light of the challenge of creating incentive-compatible valuation tasks and the necessity of employing stated preference methods in certain scenarios, an expanding body of research has explored measures to reduce hypothetical bias. The prior literature emphasizes *ex-ante* information, delivered before stated preference questions, which could reduce hypothetical bias. This study explores whether diffusing messages through video or text effectively mitigates hypothetical bias.

This study examines the impacts of presentation formats on mitigating hypothetical bias in stated preference methods. Using discrete choice experiments on beef jerky products, we focus on two pre-choice task information components: a cheap talk script and the explanation of products. Cheap talk is a mainstay technique among stated preference practitioners to reduce hypothetical bias, while the explanation of products is typically presented to improve individuals' understanding, resembling real-life scenarios. Our between-subject design includes one non-hypothetical DCE group and four hypothetical DCE groups varying in the presentation format of the cheap talk script and the explanation, using text or video format. Findings reveal the presence of hypothetical bias across products and hypothetical DCE groups, with varying degrees of bias. Illustrating with Both_text, the conventional practice of hypothetical DCE in the literature, the mean WTPs of plant-based beef jerky are -7.573 RMB per 50 grams, respectively. This represents 1.5 times the values observed in Real, consistent with the results of Penn and Hu (2018). They conducted an extensive meta-analysis that included 131 studies and showed that the mean (median) ratio of hypothetical WTP to real WTP in the dataset is 2.29 (1.39).

The value of information hinges on its communication mode, be it through video or text. While the information content remains unchanged across hypothetical DCE groups, those incorporating video displays of ex-ante information designed to mitigate hypothetical bias achieve the intended goal more successfully. Compared to Both_text, the remaining three hypothetical

DCE groups, each incorporating at least one video display, produce WTP values more closely aligned with those in Real. Notably, Both_video exhibits the smallest extent of hypothetical bias, whereas V_cheap and V_instruc show comparable levels. This parallels findings that relative to text-based choice tasks, framing the tasks with visual reality or video affects WTP estimates (Bateman et al., 2009; Matthews et al., 2017; Mokas et al., 2021) and reduces hypothetical bias (Fang et al., 2019). Some have compared the impacts of pictorial versus text choice tasks on choice consistency and found mixed results (Townsend and Kahn, 2014; Uggeldahl et al., 2016; Shr et al., 2019). There has been limited discussion in the literature of how hypothetical bias is influenced by communication mode, although hypothetical bias presents a measure that directly challenges the validity of stated preference methods. We contribute to the literature by being the first to examine the effective communication mode of pre-choice task information for hypothetical bias mitigation.

In an attempt to understand possible reasons underlying the effect of information videos, we investigate individuals' objective and subjective understanding of hypothetical bias and product attributes after the presentation of corresponding information. There is evidence that video displays enhance subjects' objective understanding, attributable to the argument that visualization techniques can reduce respondents' cognitive load and improve the evaluation and interpretation of complex information (Mokas et al., 2021; Penn and Hu, 2021). Second, we find that the treatment effect of information videos is attribute-specific. Information video reduces hypothetical bias in WTP for plant-based beef jerky most, followed by the WTP for and carbon labeled beef jerky. However, the valuation of spicy beef is similar in all groups. Relative to plant-based and carbon labels, taste (original or spicy) is a familiar attribute to respondents which may suffer less from hypothetical bias. Besides, in examining whether individuals with high and low cognitive abilities respond differently to information videos, we find that the high cognitive ability subsample exhibits lower hypothetical bias in Both_text. This aligns with the study by DeShazo and Fermo (2002) and Dohmen et al. (2010) that individuals with lower cognitive abilities tend to be more impatient and prone to choice randomness. However, the disparity in the degree of hypothetical bias between the high and low cognitive ability subsamples tends to diminish in the remaining three hypothetical DCE groups, notably in Both_video. In other words, the impact of information videos on reducing hypothetical bias is more consistent across individuals with different cognitive abilities. This implies that videos mitigate the biases induced by individual differences in cognitive ability, leading to the narrowing gap in hypothetical bias between those with high and low cognitive capacities. Our results highlight the role of cognitive capabilities in designing and implementing measures to mitigate hypothetical bias.

Meta-analyses on hypothetical bias consistently indicate greater hypothetical bias when public goods are valued, with others being equal (List and Gallet, 2001; Murphy et al., 2005; Penn and Hu, 2018). In our DCE design, although individuals are evaluating a private good (beef jerky), the attributes being evaluated differ in their alignment with public versus private benefits. Specifically, the carbon label reflects characteristics typically associated with public goods, given its implications for collective environmental benefits, while the plant-based label is more akin to a private good. Following the approach outlined in previous meta-analyses, we calculate the degree of hypothetical bias. Table R1 indicates that the hypothetical bias is greater for the carbon label than for the plant-based label across all hypothetical treatments, consistent with the findings of these meta-analyses. Several factors could contribute to the difference in hypothetical bias when valuing public versus private goods. First, familiarity and experience. Most subjects are more comfortable valuing goods they commonly purchase, so they may make less errors in valuing these types of goods than valuing public goods, which they may have little valuation experience (List and Gallet, 2001). This is evidenced by a study showing that people misperceive the costs of providing public goods in stated preference questions (Lang et al., 2022). Second, social desirability bias. When valuing public goods, respondents might feel societal pressure to provide responses that align with what is perceived as "socially responsible" or "environmentally friendly," which can inflate their WTP for public goods (Entem et al., 2022). This is less likely to occur for private goods, where the decision is more individualistic. These may explain why a higher degree of hypothetical bias is observed when valuing public goods compared to private goods. Furthermore, we show that video treatments reduce hypothetical bias for the two product attributes more effectively, compared to the Both_text treatment. Table 5 of this submission explains these treatment effects. Table 5 suggests that the video-based cheap talk script improves respondent's objective understanding about hypothetical bias relative to Both_text, making them more aware of the potential discrepancy between hypothetical and real choices. Similarly, the video-based product explanation enhances their knowledge about the carbon label by providing a clearer, more vivid presentation of the carbon label, increasing respondent's familiarity and comprehension.

Finally, it is noted that the plant-based label, providing a private benefit, shows a more significant reduction in hypothetical bias across treatments, suggesting that our treatment effects are more pronounced for private valuations, different from some meta-analyses indicating that cheap talk script is more effective for public goods (Penn and Hu, 2019). Considering the context of our study, further research is needed to explore this issue more comprehensively. We note that information videos are not limited to laboratory environments. Since we focus on pre-choice task information, i.e. the cheap talk script and product explanation, our measures would open up the opportunity to mainstream the application of information videos in stated preference methods for hypothetical bias reduction. While our study keeps the DCE design constant across groups, future research could test our findings by exploring different products, choice experimental designs, and sample populations.

Ackowledgments

Many thanks to Wenxin Hou and Tingyu Wang for their assistance with the lab experiment and survey on this project. This work was supported by the National Natural Science Foundation of China (No. 72103180), the National Social Science Fund of China (No.22BGL304) and the Fundamental Research Funds for the Central Universities. All errors remain ours.

References

Andreoni, J. (1995). Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics*, *110*(1), 1-21.

Bahety, G., Bauhoff, S., Patel, D., & Potter, J. (2021). Texts don't nudge: An adaptive trial to prevent the spread of COVID-19 in India. *Journal of Development Economics*, *153*, 102747.

Balew, S., Bulte, E., Abro, Z., & Kassie, M. (2023). Incentivizing and nudging farmers to spread information: Experimental evidence from Ethiopia. *American Journal of Agricultural Economics*, *105*(3), 994-1010.

Banerjee, A., Alsan, M., Breza, E., Chandrasekhar, A. G., Chowdhury, A., Duflo, E., ... & Olken, B. A. (2020). Messages on COVID-19 prevention in India increased symptoms reporting and adherence to preventive behaviors among 25 million recipients with similar effects on non-recipient members of their communities. *National Bureau of Economic Research,* No. w27496).

Bartling, B., & Netzer, N. (2016). An externality-robust auction: Theory and experimental evidence. *Games and Economic Behavior*, *97*, 186-204.

Bateman, I. J., Day, B. H., Jones, A. P., & Jude, S. (2009). Reducing gain–loss asymmetry: a virtual reality choice experiment valuing land use change. *Journal of Environmental Economics and Management*, *58*(1), 106-118.

Boissiere, M., Knight, J. B., & Sabot, R. H. (1985). Earnings, schooling, ability, and cognitive skills. *The American Economic Review*, *75*(5), 1016-1030.

Caputo, V., & Scarpa, R. (2022). Methodological advances in food choice experiments and modeling: current practices, challenges, and future research directions. *Annual Review of Resource Economics*, *14*, 63-90.

Carlson, R., Chandler, P., & Sweller, J. (2003). Learning and understanding science instructional material. *Journal of Educational Psychology*, *95*(3), 629.

Carlsson, F., Frykblom, P., & Lagerkvist, C. J. (2005). Using cheap talk as a test of validity in choice experiments. *Economics Letters*, *89*(2), 147-152.

Carson, R. T., Groves, T., & List, J. A. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists*, *1*(1/2), 171-207.

Castro-Alonso, J. C., Ayres, P., & Sweller, J. (2019). Instructional visualizations, cognitive load theory, and visuospatial processing. *Visuospatial Processing for Education in Health and Natural Sciences*, 111-143.

Crippa, M., Solazzo, E., Guizzardi, D., Monforti-Ferrario, F., Tubiello, F. N., & Leip, A. J. N. F. (2021). Food systems are responsible for a third of global anthropogenic GHG emissions. *Nature Food*, *2*(3), 198-209.

Cummings, R. G., & Taylor, L. O. (1999). Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American Economic Review*, *89*(3), 649-665.

Dahlstrom, M. F. (2014). Using narratives and storytelling to communicate science with nonexpert audiences. *Proceedings of the National Academy of Sciences*, *111*(supplement_4), 13614-13620.

de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *The Patient-Patient-Centered Outcomes Research*, *8*, 373-384.

De-Magistris, T., Gracia, A., & Nayga Jr, R. M. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American Journal of Agricultural Economics*, *95*(5), 1136-1154.

Depositario, D. P. T., Nayga Jr, R. M., Wu, X., & Laude, T. P. (2009). Should students be used as subjects in experimental auctions?. *Economics Letters*, *102*(2), 122-124.

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

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability?. *American Economic Review*, *100*(3), 1238-1260.

Entem, A., Lloyd-Smith, P., Adamowicz, W. V. L., & Boxall, P. C. (2022). Using inferred valuation to quantify survey and social desirability bias in stated preference research. *American Journal of Agricultural Economics*, *104*(4), 1224-1242.

Fang, D., Nayga Jr, R. M., West, G. H., Bazzani, C., Yang, W., Lok, B. C., ... & Snell, H. A. (2021). On the use of virtual reality in mitigating hypothetical bias in choice experiments. *American Journal of Agricultural Economics*, *103*(1), 142-161.

Grebitus, C., Steiner, B., & Veeman, M. M. (2016). Paying for sustainability: A cross-cultural analysis of consumers' valuations of food and non-food products labeled for carbon and water footprints. *Journal of Behavioral and Experimental Economics*, *63*, 50-58.

Hallsworth, M., List, J. A., Metcalfe, R. D., & Vlaev, I. (2017). The behavioralist as tax collector: Using natural field experiments to enhance tax compliance. *Journal of Public Economics*, *148*, 14-31.

Herriges, J., Kling, C., Liu, C. C., & Tobias, J. (2010). What are the consequences of consequentiality?. *Journal of Environmental Economics and Management*, *59*(1), 67-81.

Hess, S., & Rose, J. M. (2007). Effects of distributional assumptions on conditional estimates from Mixed Logit models. *Arbeitsberichte Verkehrs-und Raumplanung*, 423.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, *17*(6), 722-738.

Howard, G., Roe, B. E., Nisbet, E. C., & Martin, J. F. (2017). Hypothetical bias mitigation techniques in choice experiments: do cheap talk and honesty priming effects fade with repeated choices?. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 543-573.

Jacquemet, N., Joule, R. V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management*, *65*(1), 110-132.

Khaw, M. W., Li, Z., & Woodford, M. (2021). Cognitive imprecision and small-stakes risk aversion. *The Review of Economic Studies*, *88*(4), 1979-2013.

Kilders, V., & Caputo, V. (2024). A reference-price-informed experiment to assess consumer demand for beef with a reduced carbon footprint. *American Journal of Agricultural Economics*, forthcoming.

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

La Nauze, A., & Myers, E. (2023). Do Consumers Acquire Information Optimally? Experimental Evidence from Energy Efficiency (No. w31742). *National Bureau of Economic Research.*

Ladenburg, J., & Olsen, S. B. (2014). Augmenting short cheap talk scripts with a repeated opt-out reminder in choice experiment surveys. *Resource and Energy Economics*, *37*, 39-63.

Lancsar, E., & Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics*, *26*, 661-677.

Li, X., & Camerer, C. F. (2022). Predictable effects of visual salience in experimental decisions and games. *The Quarterly Journal of Economics*, *137*(3), 1849-1900.

Li, X., Jensen, K. L., Clark, C. D., & Lambert, D. M. (2016). Consumer willingness to pay for beef grown using climate-friendly production practices. *Food Policy*, *64*, 93-106.

Lim, K. H., & Hu, W. (2023). Contextual reference price in choice experiments. *American Journal of Agricultural Economics*, *105*(4), 1288-1306.

Lin, W. (2023). The effect of product quantity on willingness to pay: A meta-regression analysis of beef valuation studies. *Agribusiness*, forthcoming.

Lusk, J. L., & Norwood, F. B. (2009). An inferred valuation method. *Land Economics*, *85*(3), 500-514.

Matthews, Y., Scarpa, R., & Marsh, D. (2017). Using virtual environments to improve the realism of choice experiments: A case study about coastal erosion management. *Journal of Environmental Economics and Management*, *81*, 193-208.

Mayer, R. E., & Moreno, R. (2016). Nine ways to reduce cognitive load in multimedia learning. In *Cognitive Load Theory* (pp. 43-52). Routledge.

McKenzie, D. (2015). Tools of the Trade: a joint test of orthogonality when testing for balance. *World Bank Blog.* https://blogs.worldbank.org/en/impactevaluations/tools-trade-joint-test-orthogonality-when-testing-balance#comments.

Menapace, L., & Raffaelli, R. (2020). Unraveling hypothetical bias in discrete choice experiments. *Journal of Economic Behavior & Organization*, *176*, 416-430.

Meyerding, S. G., Bauchrowitz, A., & Lehberger, M. (2019). Consumer preferences for beer attributes in Germany: A conjoint and latent class approach. *Journal of Retailing and Consumer Services*, *47*, 229-240.

Mokas, I., Lizin, S., Brijs, T., Witters, N., & Malina, R. (2021). Can immersive virtual reality increase respondents' certainty in discrete choice experiments? A comparison with traditional presentation formats. *Journal of Environmental Economics and Management*, *109*, 102509.

Ortega, D. L., Sun, J., & Lin, W. (2022). Identity labels as an instrument to reduce meat demand and encourage consumption of plant based and cultured meat alternatives in China. *Food Policy*, *111*, 102307.

Penn, J. M., & Hu, W. (2018). Understanding hypothetical bias: An enhanced meta-analysis. *American Journal of Agricultural Economics*, *100*(4), 1186-1206.

Penn, J., & Hu, W. (2021). Videos Improve Attention and Cheap Talk in Online Surveys. *Frontiers of Economics in China*, *16*(2).

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

Raven, J. C. (1938). Raven standard progressive matrices. *Journal of Cognition and Development*.

Salomon, G. (1979). Media and symbol systems as related to cognition and learning. *Journal of Educational Psychology*, *71*(2), 131.

Scarpa, R., Thiene, M., & Train, K. (2008). Appendix to Utility in WTP space: a tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics*, *90*(314-2016-6062).

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156-163.

Shr, Y. H. J., Ready, R., Orland, B., & Echols, S. (2019). How do visual representations influence survey responses? Evidence from a choice experiment on landscape attributes of green infrastructure. *Ecological Economics*, *156*, 375-386.

Snowberg, E., & Yariv, L. (2021). Testing the waters: Behavior across participant pools. *American Economic Review*, *111*(2), 687-719.

Townsend, C., & Kahn, B. E. (2014). The “visual preference heuristic”: The influence of visual versus verbal depiction on assortment processing, perceived variety, and choice overload. *Journal of Consumer Research*, *40*(5), 993-1015.

Train, K., & Weeks, M. (2005). *Discrete choice models in preference space and willingness-to-pay space* (pp. 1-16). Springer Netherlands.

Uggeldahl, K., Jacobsen, C., Lundhede, T. H., & Olsen, S. B. (2016). Choice certainty in discrete choice experiments: Will eye tracking provide useful measures?. *Journal of Choice Modelling*, *20*, 35-48.

Van Loo, E. J., Caputo, V., & Lusk, J. L. (2020). Consumer preferences for farm-raised meat, lab-grown meat, and plant-based meat alternatives: Does information or brand matter?. *Food Policy*, *95*, 101931.

Vossler, C. A., & Watson, S. B. (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior & Organization*, *86*, 137-147.

Vossler, C. A., Doyon, M., & Rondeau, D. (2012). Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, *4*(4), 145-171.

Wei, X., & Khachatryan, H. (2023). How consequential is policy consequentiality? Evidence from online discrete choice experiment with ornamental plants. *Journal of Behavioral and Experimental Economics*, *107*, 102122.

Xie, L., Adamowicz, W., Kecinski, M., & Fooks, J. R. (2022). Using economic experiments to assess the validity of stated preference contingent behavior responses. *Journal of Environmental Economics and Management*, *114*, 102659.

Yang, Y., & Hobbs, J. E. (2020). The power of stories: Narratives and information framing effects in science communication. *American Journal of Agricultural Economics*, *102*(4), 1271-1296.

Zhao, Q., Wang, X., & Rozelle, S. (2019). Better cognition, better school performance? Evidence from primary schools in China. *China Economic Review*, *55*, 199-217.

Zhu, W. B., Chen, Y. F., Jing, Z. H. A. O., & Wu, B. B. (2021). Impacts of household income on beef at-home consumption: Evidence from urban China. *Journal of Integrative Agriculture*, *20*(6), 1701-1715.

Tables

Table 1. Socio-demographics, beef jerky consumption and cognitive ability across treatment

Variable	Definition	Real	Both_text	V_cheap	V_instruc	Both_video	p-value
Socio-demographics							
Female	=1, female; =0, otherwise	0.703	0.673	0.723	0.703	0.702	0.243
		(0.491)	(0.471)	(0.449)	(0.451)	(0.454)	
Age	Year	22.068	21.832	21.610	21.593	22.074	0.524
		(2.142)	(2.191)	(2.055)	(1.997)	(2.774)	
Grade	=1, freshman/sophomore; =2, junior/senior; =3, graduate	2.017	1.981	1.895	1.855	1.923	0.549
		(0.791)	(0.801)	(0.784)	(0.776)	(0.834)	
Major	=1, humannities; =2, social sciences; =3, natural sciences; =4, engineering; =5, medicine; =6, others	3.336	3.664	3.333	3.441	3.463	0.361

		(1.497)	(1.473)	(1.645)	(1.692)	(1.649)	
Monthly expenditure (RMB)	=1, below 1000; =2, 1000-1999; =3, 2000-2999; =4, 3000-3999; =5, 4000-4999; =6, above 5000	2.552	2.561	2.457	2.602	2.620	0.442
		(0.837)	(0.803)	(0.679)	(0.775)	(0.924)	
Monthly food expenditure (RMB)	=1, below 500; =2, 500-999; =3, 1000-1499; =4, 1400-1999; =5, 2000-2499; =6, 2500-2999; =7 above 3000	2.965	2.878	2.819	2.983	3.064	0.837
		(0.833)	(0.908)	(0.841)	(0.877)	(1.025)	
Residence	=1, urban; =0, rural	0.741	0.766	0.790	0.762	0.722	0.410
		(0.440)	(0.425)	(0.409)	(0.384)	(0.445)	
Hunger status	=1, extremely hungry;… =5, not hungry at all	3.267	3.364	3.305	3.254	3.324	0.826

		(0.838)	(0.732)	(0.785)	(0.753)	(0.747)	
Meat jerky purchasing habit							
Average purchasing frequency	=1, everryday; =2, 3-4 times per week; =3,1-2 times per week; =4, less than weekly	3.327	3.187	3.314	3.389	3.389	0.689
		(0.778)	(0.848)	(0.788)	(0.806)	(0.771)	
Average expenditure per purchase (RMB)	=1, below 50; =2, 50-100; =3, 101-150; =4, above 150	1.155	1.290	1.333	1.263	1.204	0.179
		(0.386)	(0.583)	(0.531)	(0.530)	(0.427)	
Preferred taste	=1, original; =2, spicy; =3, others	1.466	1.607	1.571	1.508	1.537	0.308
		(0.566)	(0.527)	(0.586)	(0.535)	(0.537)	
Raven test score							

0 to 12	8.931	8.813	9.019	9.211	9.056	0.926
	(2.166)	(2.295)	(2.472)	(2.164)	(2.531)	

Notes: standard deviations are in parentheses. The Raven test is borrowed from the measurement by Bartling and Netzer (2016), and higher Raven test scores indicates higher cognitive ability. P-values are generated by chi-squared tests examining equality across treatments. We also conducted joint orthogonality tests which suggests all variables are jointly orthogonal with the respect to the treatments (F-statistics=0.608, p-value=0.652).

Table 2. Mixed logit models in WTP space

	Real		Both_text		V_cheap		V_instruc		Both_video	
	Coefficient	Z-value	Coefficient	Z-value	Coefficient	Z-value	Coefficient	Z-value	Coefficient	Z-value
Means										
Opt-out	-18.355***	-37.57	-19.239***	-26.60	-18.457***	-28.78	-18.023***	-31.71	-15.425***	-25.34
Plant-based	-5.031***	-8.42	-7.573***	-9.45	-6.944***	-8.55	-6.775***	-10.94	-5.699***	-7.24
Carbon	1.280***	2.81	2.185***	3.91	2.044***	4.57	2.079***	5.09	2.009***	3.78
Spicy	0.174	0.30	-0.474	-0.58	-1.197	-1.54	-1.486	-1.07	-0.183	-0.18
$\bar{\sigma}$	-0.949***	-11.62	-1.312***	-12.07	-1.135***	-11.58	-1.147***	-11.56	-1.121***	-9.17
Standard Deviations										
Opt-out	5.107***	91.5	5.461***	9.37	5.399***	13.05	5.265***	9.28	5.968***	8.76
Plant-based	4.029***	7.23	4.225***	4.91	4.595***	5.34	3.247***	4.81	3.963***	5.47
Carbon	0.580	1.41	0.955**	2.43	1.569***	6.12	1.361	1.57	2.393***	3.33
Spicy	3.819***	7.62	4.965***	5.59	4.943***	11.31	4.290***	4.03	6.983***	5.61
τ	0.528***	3.93	0.498*	1.82	0.181	0.77	0.205	0.71	0.417**	2.32

Model statistics					
Observations	1488	1648	1488	1616	1568
Log-likelihood	-635.774	-660.462	-573.212	-603.791	-610.586

*Notes: Observation denotes the number of choices scenarios for all respondents, with standard errors clustered at the respondent level. ***, **, *indicate significance levels at 0.01, 0.05, and 0.10. $\bar{\sigma}$ and τ are the mean and standard deviation of the lognormally distributed scale parameter, $\sigma_i' = exp[\bar{\sigma} + \tau w_i]$, in equation (2).*

Table 3. Mixed logit model in WTP space, a pooled sample

	Model 1		Model 2	
	Coefficient	Z-value	Coefficient	Z-value
Means				
Opt-out	-16.732***	-37.57	-17.152***	-43.39
Opt-out × Both_text	-1.434***	-2.65	-1.180	-1.45
Opt-out × V_cheap	1.311**	1.96	0.509	0.68
Opt-out × V_instruc	-0.182	-0.27	-0.172	-0.39
Opt-out × Both_video	1.666**	2.44	1.004	1.31
Plant-based	-5.102***	-8.38	-5.241***	-8.58
Plant-based × Both_text	-2.711***	-3.19	-2.815***	-3.70
Plant-based × V_cheap	-2.040**	-2.31	-2.016***	-2.82
Plant-based × V_instruc	-2.076***	-2.79	-2.010***	-3.44
Plant-based × Both_video	-1.064	-1.18	-1.097	-1.38
Carbon	1.275***	2.83	1.521***	3.70
Carbon × Both_text	0.992***	2.64	0.668**	2.21
Carbon × V_cheap	0.754***	2.82	0.712***	3.15
Carbon × V_instruc	1.007**	2.17	1.056***	3.12
Carbon × Both_video	0.873	1.24	0.788	1.37
Spicy	0.242	0.30	-0.562	-1.10
Spicy × Both_text	0.219	0.822	0.522	0.62
Spicy × V_cheap	-1.039	-1.04	-1.546	-1.77

Spicy × V_instruc	-0.864	-0.90	-0.534	-0.62
Spicy × Both_video	0.426	0.42	-0.009	-0.21
$\bar{\sigma}$	-0.966***	-11.74	-0.919***	-14.76
Standard Deviations				
Opt-out	5.410***	91.50	5.216***	89.42
Opt-out × Both_text			4.625***	6.51
Opt-out × V_cheap			3.904***	6.00
Opt-out × V_instruc			4.622***	8.49
Opt-out × Both_video			5.248***	7.62
Plant-based	4.123***	7.23	4.216***	17.14
Plant-based × Both_text			0.550	0.691
Plant-based × V_cheap			2.045	1.52
Plant-based × V_instruc			1.968**	2.04
Plant-based × Both_video			0.922***	2.02
Carbon	0.370	1.53	1.656***	5.15
Carbon × Both_text			0.288	0.27
Carbon × V_cheap			1.127**	2.23
Carbon × V_instruc			1.267***	2.99
Carbon × Both_video			0.196	0.24
Spicy	3.219***	7.26	4.15***	8.40
Spicy × Both_text			1.233***	2.82
Spicy × V_cheap			1.148***	3.59
Spicy × V_instruc			1.201**	2.49

Spicy × Both_video			0.704***	7.38
τ	0.503***	3.91	0.656***	8.11
Model statistics				
Number of Parameters	26		42	
Observations	7808		7808	
Log-likelihood	-3069.745		-2971.733	

Notes: Observation denotes the number of choices scenarios for all respondents, with standard errors clustered at the respondent level. ***, **, ** indicate significance levels at 0.01, 0.05, and 0.10. $\bar{\sigma}$ and τ are the mean and standard deviation of the lognormally distributed scale parameter, $\sigma_i' = exp[\bar{\sigma} + \tau w_i]$, in equation (2).*

Table 4. Conditional logit model in preference space, a pooled sample

	Model 1		Model 2	
	Coefficient	Z-value	Coefficient	Z-value
Opt-out	-4.273***	-17.63	-4.289***	-17.26
Opt-out × Both_text	0.990***	2.88	0.981***	2.76
Opt-out × V_cheap	0.655*	1.83	0.671*	1.81
Opt-out × V_instruc	0.492	1.40	0.524	1.45
Opt-out × Both_video	0.821**	2.28	0.855**	2.31
Price	-0.214***	-17.05	-0.195***	-16.75
Price × Both_text	0.066***	3.32	0.069***	3.31
Price × V_cheap	0.025**	2.57	0.024**	2.48
Price × V_instruc	0.032*	1.66	0.033*	1.70
Price × Both_video	0.021	1.47	0.022	1.01
Plant-based	-0.938***	-9.30	-0.941***	-9.26
Plant-based × Both_text	-0.443**	-2.31	-0.451**	-2.34
Plant-based × V_cheap	-0.412***	-2.63	-0.418***	-2.75
Plant-based × V_instruc	-0.427***	-2.78	-0.426***	-2.98
Plant-based × Both_video	-0.232	-1.45	-0.238	-0.99
Carbon	0.295***	3.07	0.296***	3.07
Carbon × Both_text	0.212***	3.03	0.230***	3.00
Carbon × V_cheap	0.185**	2.42	0.200**	2.39

Carbon × V_instruc	0.221**	2.23	0.229**	2.24
Carbon × Both_video	0.117	0.79	0.122	0.81
Spicy	-0.033	-0.34	-0.033	-0.35
Spicy × Both_text	0.092	0.67	0.091	0.92
Spicy × V_cheap	-0.233	-1.57	0.236	0.93
Spicy × V_instruc	-0.194	-1.37	-0.195	-0.96
Spicy × Both_video	0.112	0.79	0.121	0.91
Scale Heterogeneity				
Both_text			-0.116	0.92
V_cheap			-0.110	1.21
V_instruc			-0.097	1.43
Both_video			-0.098	0.96
τ			1.628***	4.19
γ			1.171**	2.45
Model statistics				
Number of Parameters	25		31	
Observations	7808		7808	
Log-likelihood	-3327.338		-3326.060	

*Notes: Observation denotes the number of choices scenarios for all respondents, with standard errors clustered at the respondent level. ***, **, *indicate significance levels at 0.01, 0.05, and 0.10. In the estimation, we permit γ to take any value, in line with Keane and Wasi (2013) who point out that there is no* *reason to restrict* γ *between 0 and 1, and* $\gamma < 0$ *or* $\gamma > 1$ *still permits sensible behavioral interpretations.*

Table 5. Understanding about cheap talk and DCE attributes across treatment

Variable	Definition	Real	Both_text	V_cheap	V_instruc	Both_video
Subjective understanding of hypothetical bias	=1, useless;... =5, extremely useful	N.A.	4.000	4.129*	4.059	4.157*
		N.A.	(0.728)	(0.679)	(0.822)	(0.740)
Objective understanding of hypothetical bias	=1, correct; =0, otherwise	N.A.	0.951	0.989**	0.970	0.988*
		N.A.	(0.216)	(0.104)	(0.171)	(0.110)
Subjective understanding of DCE product attributes	=1, useless;... =5, extremely useful	4.198	4.258	4.265	4.314	4.330
		(0.800)	(0.706)	(0.717)	(0.467)	(0.549)
Objective understanding of DCE product attributes	=1, correct; =0, otherwise	0.925	0.901	0.952	0.971**	1.000***
		(0.265)	(0.300)	(0.215)	(0.169)	(0.000)

Notes: the analysis is conducted at the individual level. Standard deviations are in parentheses. N.A. stands for not applicable. Subjective understanding of hypothetical bias is evaluated using *the question "To what extent do you believe the explanation of hypothetical bias improves your understanding of hypothetical bias?". We use a similar approach to obtain a subjective understanding of DCE attributes. We assess an objective understanding of hypothetical bias with the question: "In a hypothetical purchasing environment, respondents' willingness to pay is often ___ their willingness to pay in actual purchases",' with the correct response being 'lower than'. We evaluate an objective understanding of DCE attributes through the question: "A product with the following carbon label has ___ carbon emissions than those without", with the correct response being 'fewer'. All the understanding questions were presented to respondents immediately after providing explanations on hypothetical bias and DCE attributes, presented in either text or video formats. $^{***, **, *}$ indicate significance levels at 0.01, 0.05, and 0.10. For all the variables, we refer to Both_text as the reference group and compare it with each of the remaining treatments.*

Table 6. Heterogenous treatment effects by cognitive ability

	Carbon	Plant-based
Both_text	1.062***	-2.644***
	(0.297)	(0.576)
V_cheap	0.898***	-2.002***
	(0.237)	(0.575)
V_instruc	0.793***	-1.965***
	(0.263)	(0.422)
Both_video	0.804**	-0.670
	(0.359)	(0.617)
Both_text×Above_Raven	-0.858**	1.093*
	(0.356)	(0.625)
V_cheap×Above_ Raven	-0.249**	0.660***
	(0.117)	(0.222)
V_instruc×Above_ Raven	-0.321*	0.841**
	(0.189)	(0.397)
Both_video×Above_ Raven	-0.232**	0.324*
	(0.103)	(0.187)
Constant	1.267***	-5.176***
	(0.165)	(0.301)
R^2	0.21	0.15
Observations	473	473

Notes: Observation denotes the number of respondents. The dependent variable in the first column is individual conditional WTP for beef jerky with a low carbon certification, and the variable in the second column is individual conditional WTP for plant-based beef jerky (RMB per 50 grams). Conditional WTP refers to individual-specific mean WTP conditional on observed choices. Standard errors are in parentheses. ***, **, *and* **indicate significance levels at 0.01, 0.05, and 0.10, respectively. The coefficient of the constant term refers to the average individual conditional WTP of Real. Above_Raven is a dummy variable, equal one if a person's Raven test score is above the median of the full sample, and zero otherwise.*

Figure 1. Sample choice task

Endnotes

[1] This study obtained ethical approval from the institutional review board and was pre-registered before data collection began (https://aspredicted.org/MLW_TLM).

[2] One reviewer kindly refers us to the paper by Train and Hess (2017) which discussed that the GMNL model could not disentangles preference heterogeneity from scale heterogeneity. We caution that this result needs to be interpreted with caution.

售价:9元

选项A

售价:13元

选项B

我不购买选项A
也不购买选项B

选项C