

RESEARCH ARTICLE

# Do labeled versus unlabeled treatments of alternatives' names influence stated choice outputs? Results from a mode choice study

Wen Jin<sup>1</sup>, Hai Jiang<sup>1\*</sup>, Yimin Liu<sup>2</sup>, Erica Klampf<sup>2</sup>

**1** Department of Industrial Engineering, Tsinghua University, Beijing 100084, China, **2** Research & Advanced Engineering, Ford Motor Company, 2101 Village Road MD-2149, Dearborn, MI 48121, United States of America

\* [haijiang@tsinghua.edu.cn](mailto:haijiang@tsinghua.edu.cn)



## Abstract

Discrete choice experiments have been widely applied to elicit behavioral preferences in the literature. In many of these experiments, the alternatives are *named alternatives*, meaning that they are naturally associated with specific names. For example, in a mode choice study, the alternatives can be associated with names such as car, taxi, bus, and subway. A fundamental issue that arises in stated choice experiments is whether to treat the alternatives' names as labels (that is, labeled treatment), or as attributes (that is, unlabeled treatment) in the design as well as the presentation phases of the choice sets. In this research, we investigate the impact of labeled versus unlabeled treatments of alternatives' names on the outcome of stated choice experiments, a question that has not been thoroughly investigated in the literature. Using results from a mode choice study, we find that the labeled or the unlabeled treatment of alternatives' names in either the design or the presentation phase of the choice experiment does not statistically affect the estimates of the coefficient parameters. We then proceed to measure the influence toward the willingness-to-pay (WTP) estimates. By using a random-effects model to relate the conditional WTP estimates to the socioeconomic characteristics of the individuals and the labeled versus unlabeled treatments of alternatives' names, we find that: a) Given the treatment of alternatives' names in the presentation phase, the treatment of alternatives' names in the design phase does not statistically affect the estimates of the WTP measures; and b) Given the treatment of alternatives' names in the design phase, the labeled treatment of alternatives' names in the presentation phase causes the corresponding WTP estimates to be slightly higher.

## OPEN ACCESS

**Citation:** Jin W, Jiang H, Liu Y, Klampf E (2017) Do labeled versus unlabeled treatments of alternatives' names influence stated choice outputs? Results from a mode choice study. PLoS ONE 12(8): e0178826. <https://doi.org/10.1371/journal.pone.0178826>

**Editor:** Xiaolei Ma, Beihang University, CHINA

**Received:** April 26, 2017

**Accepted:** May 19, 2017

**Published:** August 14, 2017

**Copyright:** © 2017 Jin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** YL and EK are supported in the form of salary by Ford Motor Company. The specific roles of these authors are articulated in the 'author contributions' section. This research is supported by the Ford University Research Programs. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Introduction

Discrete choice experiments have been widely applied to elicit behavioral preferences in many fields such as transportation [1–3], marketing [4], environmental research [5, 6], health economics [7, 8], and housing studies [9, 10]. In these experiments, respondents are asked to complete one or more choice tasks, each of which contains several alternatives with hypothetical

**Competing interests:** YL and EK are supported in the form of salary by Ford Motor Company. This study was supported by the Ford University Research Programs. There are no patents, products in development, or marketed products to declare. This commercial affiliation does not alter our adherence to PLOS ONE policies on sharing data and materials.

attribute values. Within each choice task, the respondent needs to indicate his or her most preferred alternative, which can later be used to estimate appropriate discrete choice models and quantify respondents’ preferences toward the attributes. In many of these experiments, the alternatives are *named alternatives*, meaning that they are naturally associated with specific names, which convey additional information regarding the tangible or intangible qualities of the alternatives [11]. For example, in mode choice studies, the alternatives’ names can be car, taxi, bus, and subway [12, 13]; in recreational site choice, the names can be river walk, hill walk, and field walk [14]; in air travel choice, the names can be Air France, KLM, and Iberia [15, 16]; and in the choice of colorectal cancer screening methods, the names can be fecal occult blood tests (FOBTs), sigmoidoscopy, and colonoscopy [17].

A fundamental issue that arises in stated choice experiments involving named alternatives is whether to treat the alternatives’ names as labels (that is, labeled treatment), or as attributes (that is, unlabeled treatment). These decisions have important ramifications toward the design and the presentation of the choice sets:

- In the design of the choice sets, that is, the process that generates the underlying variations of the attribute values to maximize identification capabilities of discrete choice models, the labeled treatment of alternatives’ names views the names as titles of the alternatives, while the unlabeled treatment view the names as attributes of the alternatives. The labeled or unlabeled treatment of alternatives’ names necessitates either the labeled or the unlabeled design method [18]. The labeled design method can easily generate realistic choice tasks because it accommodates alternative specific attributes and attribute levels. Since we must allocate one alternative in the choice set for each distinct alternative name, the number of alternatives in the choice sets can become rather large when we have four or more distinct alternatives’ names, which may place considerable cognitive burden on the respondents [19]. The unlabeled design methods, however, treat the alternatives’ names as attributes. Therefore, it offers more flexibility with respect to the number of alternatives to include in a choice set. In addition, by restricting certain combinations of attribute levels, it can also avoid generating unrealistic combinations of attribute levels;
- In the presentation of the choice sets, that is, the process when we show the choice sets to the respondents and solicit their responses, the labeled treatment displays the alternatives’ names as their labels in the header row of the choice task as is illustrated by the example shown in Table 1. We refer to this as the labeled presentation style and since the alternatives’ names are displayed rather prominently, it more or less affects how respondents reach their choice outcomes [18]. The unlabeled treatment presents the alternatives’ names as attributes while referencing the alternatives with generic names, such as “Option 1”, “Option 2”, and “Option 3”, in the header row as is shown by the example in Table 2. Since the unlabeled treatment displays the alternatives’ names along with the other attributes, it could potentially force the respondents to compare the attribute values before reaching their decisions.

**Table 1. An example for the labeled treatment of alternatives’ names in the presentation phase.** The names (that is, car, bus, and subway) are displayed in the header row of the choice task.

	Car	Bus	Subway
Travel time	20 min.	30 min.	15 min.
Travel cost	20 Yuan	0.4 Yuan	2 Yuan
Your choice	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

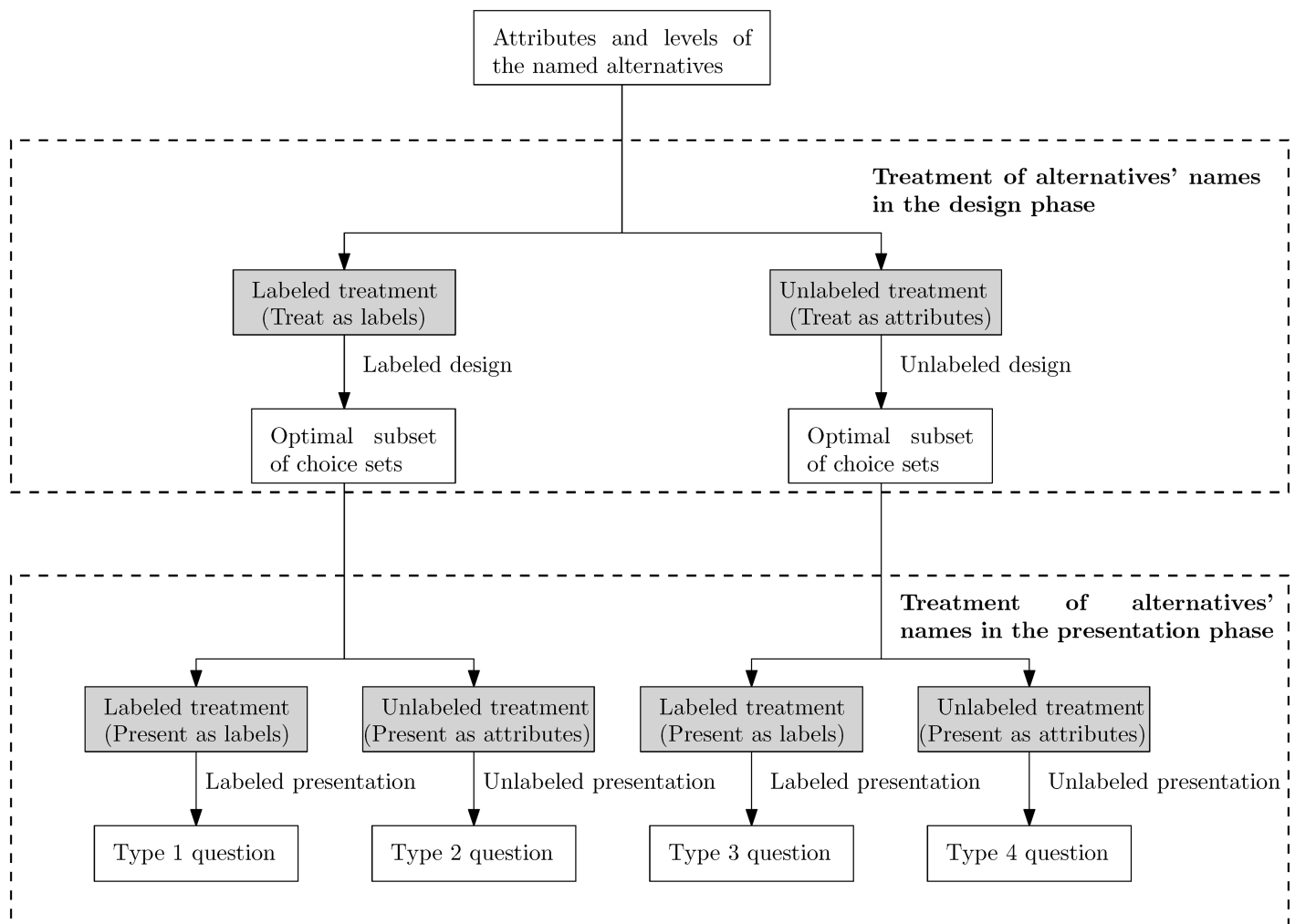
<https://doi.org/10.1371/journal.pone.0178826.t001>

**Table 2. An example for the unlabeled treatment of alternatives' names in the presentation phase.** The names (that is, car, bus, and subway) are displayed in the row that corresponds to "travel mode".

	Option 1	Option 2	Option 3
Travel time	20 min.	30 min.	15 min.
Travel cost	20 Yuan	0.4 Yuan	2 Yuan
Travel mode	Car	Bus	Subway
Your choice	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<https://doi.org/10.1371/journal.pone.0178826.t002>

The major contribution of this research is that we investigate the impact of labeled versus unlabeled treatments of alternatives' names on the outcome of stated choice experiments involving named alternatives, a problem that has not been thoroughly investigated in the literature. Since the decision to use either the labeled or unlabeled treatment of alternatives' names in the design phase is independent of that in the presentation phase, for a stated choice experiment involving named alternatives we can produce four types of questions as are shown in Fig 1. In



**Fig 1. The labeled versus unlabeled treatments of alternatives' names in the design and the presentation phases produce four types of questions for a choice experiment.**

<https://doi.org/10.1371/journal.pone.0178826.g001>

the very beginning of a choice experiment, we need to determine the set of alternatives, their attributes, and the associated attribute levels. Then, depending on the treatment of alternatives' names in the design phase, we adopt either the labeled or the unlabeled design method to obtain the corresponding optimal subset of choice sets. Regardless of which treatment we take in the design phase, when presenting the choice sets in the optimal subset to the respondent, we can choose the labeled or unlabeled treatment of alternatives' names in the presentation phase. Note that in the literature, there is also a fifth type of question, where the alternatives' names are completely removed and the cognitive significance of the alternatives' names are examined. The fifth type of questions essentially turns the named alternatives into *non-named alternatives*, and the unlabeled design method and the unlabeled presentation style have to be adopted because the alternatives' names are no longer there. Table 3 shows the relationship among the five types of questions. The rows and the columns indicate the three treatments of alternatives' names in the design and presentation phases, respectively.

Existing literature investigating the impact of labeled versus unlabeled treatments of alternatives' names in stated choice experiments is scarce. Table 3 summarizes the types of questions compared in previous research and we notice that most, if not all, research focuses on comparing Type 1 and Type 5 questions. In an effort to understand the influence of policy labels in environmental choice modeling, [20] study the effects of employing labeled rather than generic choice set configurations. In their work, Types 1 and 5 questions are generated and given to a split sample. Results show that while the mean unobserved utility component differs, the vectors of attribute taste parameters do not differ significantly when scale differences are accommodated. Later, [17] compare the feasibility, dominant effects, and convergent validity between Types 1 and 5 questions in the context of colorectal cancer screening programs in the Netherlands. They find that the inclusion of labels plays a significant role in individual choices and reduces the attention to the actual attributes. [6] assess the effect of energy technology labels on preferences. They find that the use of labeled alternatives leads to significantly different estimated parameters, while most implicit prices and welfare measures are found to be statistically equivalent. [14] investigate whether and to what extent respondents reach their choices based on the labels only. They employ a discrete mixture modeling approach on data collected using Type 1 choice tasks. They find that a portion of respondents reach their choices based on the alternatives' names only and show that this has a large impact toward welfare measures.

**Table 3. If we allow the alternatives' names to be removed, altogether five types of questions involving named alternatives can be produced.**

		Treatment of names in the presentation phase		
		Labeled treatment (Present as labels)	Unlabeled treatment (Present as attributes)	Remove the names
Treatment of names in the design phase	Labeled treatment (Treat as labels)	<b>Type 1 question</b> [20] [17] [6] [14] <u>This paper</u>	<b>Type 2 question</b> <u>This paper</u>	—
	Unlabeled treatment (Treat as attributes)	<b>Type 3 question</b> <u>This paper</u>	<b>Type 4 question</b> <u>This paper</u>	—
	Remove the names	—	—	<b>Type 5 question</b> [20] [17] [6]

<https://doi.org/10.1371/journal.pone.0178826.t003>

In this research, we compare the first four types of questions in [Table 3](#) that include the alternatives' names in both the design and the presentation of the choice sets. We quantify the influence associated with the labeled versus unlabeled treatments toward stated choice outputs in the context of a mode choice study, where residents' preferences to public transportation, private car, taxi, and vehicle sharing are studied. We generate choice tasks for each type of questions in our survey and ask respondents to indicate their preferred travel mode for each choice task. We then estimate mixed logit models using responses to each question type to assess the influence of the different treatments toward parameter estimates. Finally, we analyze the conditional willingness-to-pay (WTP) measures estimated from these data sets using a random-effects model.

The remainder of this paper is organized as follows. Section 2 gives the background of named alternatives and discuss in detail their implications in stated choice experiments. Section 3 briefly reviews the mixed logit model, the likelihood ratio test for parameter comparison, the conditional WTP estimates, and the random-effects model used in our analysis. Section 4 introduces the mode choice study, the design and the administration of the survey. Section 5 presents the results of the case study and our findings. Finally, we conclude our discussion in Section 6.

## Named alternatives and stated choice experiments

In Section 2.1 we discuss the implications associated with the treatment of alternatives' names in the design of the choice sets, and in Section 2.2 we discuss that associated with the treatment of alternatives' names in the presentation of the choice sets.

### Treatment of alternative's name in the design of the choice sets

In the design of the choice sets, we generate the underlying variations of the attribute values to maximize our model's identification capabilities. The labeled treatment of alternatives' names views the alternatives' names as labels (or titles), while the unlabeled treatment views the names as attributes. Depending on the treatment, we need to employ either the labeled or the unlabeled design method [18]. In the labeled design method, the alternatives' names are treated as labels and alternatives with different labels come from different candidate alternative sets. In the unlabeled design method, the names are treated as attributes and the alternatives all come from the same candidate alternative set.

The two design methods have their respective benefits and drawbacks. First of all, in the labeled design method, the number of alternatives in a choice set typically equals the number of distinct alternatives' names. That is, one for each distinct alternative's name. For example, in a mode choice study involving five different transportation modes, the choice sets generated by the labeled design method typically need to have exactly five alternatives. When the number of distinct alternatives' names grows, the number of alternatives in the choice set also increases, which may create information overload to the respondents and negatively impact their decision process. The unlabeled design method, however, gives us the flexibility with respect to the number of alternatives in the choice set, which can help reduce the choice complexity when the alternatives have a large number of distinct names.

Secondly, the universal choice set, that is, the set of all possible choice sets, in the labeled design is typically smaller than that in the unlabeled design, which makes the labeled design method computationally amenable when we search for the optimal design, that is, the optimal subset of choice sets. Suppose that the alternatives we consider have  $M$  distinct names and we have decided that the numbers of alternatives in the choice sets are  $M$  for both design methods. Let us further assume that in addition to its name, each of the alternatives has  $A$  attributes with

$L$  levels. Then, there are  $L^{MA}$  possible choice sets under the labeled design. In the unlabeled design, however, because the names are treated as attributes, we now have  $ML^A$  possible alternatives that can be included in each of the  $M$  positions of a choice set. Since identical alternatives are not allowed to appear in the same choice set, there are “ $ML^A$  choose  $M$ ” possible choice sets under the unlabeled design. For a typical choice experiment with four alternatives, five attributes (excluding the names of the alternatives), and three levels, the numbers of possible choice sets in the labeled design is  $3^{(4 \times 5)} = 3,486,784,401$ . However, in the unlabeled design, the number of possible alternatives is  $4 \times 3^5 = 972$  and the number of possible choice sets is  $C_{972}^4 = 36,963,217,215$ , which is substantially greater than that in the labeled design. Note that in the unlabeled design method, we can reduce the size of the universal choice set by decreasing the number of alternatives in the choice set. In the previous example, if we display three alternatives in each choice set, the number of possible choice sets can then be reduced to  $C_{972}^3 = 152,582,940$ .

Thirdly, the choice set produced by the labeled design method typically has one alternative for each distinct name, which may not be desirable when respondents make decisions solely based on the alternatives' names. For example, in a mode choice study, some respondent may always choose to drive a car regardless of the actual attributes associated with the alternatives. This attribute non-attendance phenomenon has been widely recognized in the literature [14]. The unlabeled design method, however, treats the alternatives' names as attributes, which means that in the choice set generated there can be more than one alternative having the same name. For example, there can be two car alternatives with different travel times and costs. This is advantageous because having two car alternatives can force the respondents to pay attention to the actual attribute values even if he wants the car alternative.

A review of the literature shows that there is no consensus with regard to which design method is more appropriate for stated choice experiments involving named alternatives. We find that the labeled design method is often used in transportation related studies [21–25], while the unlabeled design method is commonly adopted by health and marketing related studies [11, 17].

## Treatment of alternatives' names in the presentation of the choice set

It has been recognized that the presentation of the choice tasks affects choice outcomes, for example, the order of the alternatives, the order of the attributes, the appearance of the choice tasks, and so on [26, 27]. For choice experiments involving named alternatives, regardless of the type of experimental design method used to generate the choice set, we have two ways to present them depending on where we display the alternatives' names: a) the labeled presentation style, where the names are presented in the header row of the table as is illustrated by the example shown in Table 1. In this example, the alternatives' names, that is, car, bus, and subway, are displayed in the header row as if they are the labels of the alternatives; and b) the unlabeled presentation style, where the names are presented together with the other attributes as is illustrated by the example shown in Table 2. In this example, the alternatives' names are displayed along with travel times and travel costs.

A potential issue with the labeled presentation style is that respondents may base their choices wholly or largely on the prominently displayed name and pay less attention to the actual attributes associated with the alternatives. In the literature, there has been widespread evidence that respondents often use heuristics when making their choices [17, 28, 29]. [30], [31], and [32] investigate the extent to which respondents employ information processing strategies in their decision process. [14] find evidence that a proportion of respondents reach their choices solely on the basis of the alternative's label.

For ease of exposition in the following text, we sometimes use the phrase “labeled (or unlabeled) design” as a shorthand for the labeled (or unlabeled) treatment of alternatives’ names in the design phase, and the phrase “labeled (or unlabeled) presentation” as a shorthand for the labeled (or unlabeled) treatment of alternatives’ names in the presentation phase.

## Methods

In this research, we are interested in quantifying the impact of labeled versus unlabeled treatments of alternatives’ names on both the parameter estimates and the WTP measures in stated choice experiments. Since there are two kinds of treatments available for the alternatives’ names in both the design and the presentation phases of the choice experiment, we can produce four types of questions shown in Fig 1, and use them to solicit responses from the respondents. These responses form four data sets based on the types of questions used. The parameter estimates and WTP measures obtained from these data sets are then compared and analyzed using methods detailed in this section. We first present in Section 3.1 a brief review of the mixed logit model, which is widely used in the field of transportation studies [33, 34]. We then discuss the statistical tests to compare parameter estimates across data sets in Section 3.2. Finally, in Sections 3.3 and 3.4 we show how conditional WTP estimates can be obtained from the mixed logit model and how the random-effects model can be used to explain the variation in WTP by the socioeconomic characteristics of the individuals as well as the different treatments of alternatives’ names.

### The mixed logit model

Let us assume that we have  $N$  individuals indexed by  $n$  and each of them faces  $T$  types of questions indexed by  $t$ . Given question type  $t$ , there are  $S$  choice tasks indexed by  $s$ . In the following text, we briefly review the mixed logit model and its estimation based on responses collected for Type  $t$  questions. Let  $C_{stn}$  be the set of alternatives in the  $s$ -th choice task of question type  $t$  faced by individual  $n$ . Let  $y_{istn}$  take value 1 if respondent  $n$  chooses alternative  $i$  in  $C_{stn}$ ; and 0, otherwise. Define  $\mathbf{y}_{stn} = (y_{1stn}, y_{2stn}, \dots, y_{|C_{stn}|stn})$ , where  $|C_{stn}|$  is the cardinality of set  $C_{stn}$ . Let  $\mathbf{y}_{tn} = (\mathbf{y}_{1tn}, \mathbf{y}_{2tn}, \dots, \mathbf{y}_{Stn})$  be the observed choices from individual  $n$  for all choice tasks of type  $t$  and  $\mathbf{Y}_t = (\mathbf{y}_{t1}, \mathbf{y}_{t2}, \dots, \mathbf{y}_{tN})$  be the observed choices from all individuals for all choice tasks of type  $t$ . Let  $\boldsymbol{\beta}_{tn}$  be the coefficient vector for individual  $n$  when facing question type  $t$ . The utility function for alternative  $i \in C_{stn}$  can be written as

$$U_{istn} = \boldsymbol{\beta}'_{tn} \mathbf{x}_{istn} + \epsilon_{istn},$$

where  $\mathbf{x}_{istn}$  is a vector of length  $\kappa$ , which corresponds to the observed attributes related to alternative  $i$  in choice task  $s$  for type  $t$  questions faced by individual  $n$ ; and  $\epsilon_{istn}$  is the random term representing the unobserved utility component. In the mixed logit model,  $\boldsymbol{\beta}_{tn}$  is assumed to vary over individuals to capture preference heterogeneity in the population. It is common to assume that  $\boldsymbol{\beta}_{tn}$  conforms to a multivariate normal distribution with mean  $\mathbf{b}_t$  and variance-covariance matrix  $\mathbf{W}_t$ , and its density function is written as  $\phi(\boldsymbol{\beta}_{tn} | \mathbf{b}_t, \mathbf{W}_t)$ . The probability for individual  $n$  to choose alternative  $i$  in choice set  $C_{stn}$  conditional on  $\boldsymbol{\beta}_{tn}$  is defined as

$$P_{stn}(i | \boldsymbol{\beta}_{tn}, C_{stn}) = \frac{\exp(\boldsymbol{\beta}'_{tn} \mathbf{x}_{istn})}{\sum_{j \in C_{stn}} \exp(\boldsymbol{\beta}'_{tn} \mathbf{x}_{jstn})}.$$

The probability of individual  $n$ 's choices for all  $S$  choice tasks conditional on  $\beta_{tn}$  is

$$L(y_{tn}|\beta_{tn}) = \prod_{s=1}^S \prod_{i \in C_{stn}} \left( \frac{\exp(\beta'_{tn} x_{istn})}{\sum_{j \in C_{stn}} \exp(\beta'_{tn} x_{jstn})} \right)^{y_{istn}}$$

The unconditional probability is therefore the integral of  $L(y_{tn}|\beta_{tn})$  over  $\beta_{tn}$

$$L(y_{tn}|\mathbf{b}_t, \mathbf{W}_t) = \int L(y_{tn}|\beta_{tn}) \phi(\beta_{tn}|\mathbf{b}_t, \mathbf{W}_t) d\beta_{tn},$$

and the final likelihood function is

$$L(\mathbf{b}_t, \mathbf{W}_t) = \prod_{n=1}^N L(y_{tn}|\mathbf{b}_t, \mathbf{W}_t).$$

By maximizing the above likelihood function, we can obtain the estimates for  $\mathbf{b}_t$  and  $\mathbf{W}_t$ . In this research, we estimate parameters  $\mathbf{b}_t$  and  $\mathbf{W}_t$  using the hierarchical Bayes approach based on the Markov Chain Monte Carlo method, which offers two advantages over the maximum simulated likelihood approach: a) It does not attempt to maximize the likelihood function, which is very challenging numerically and convergence can be hard to achieve; and b) Desirable estimation properties, such as consistency and efficiency, can be obtained under more relaxed conditions [33]. Let  $k(\mathbf{b}_t, \mathbf{W}_t)$  be the prior density function for  $\mathbf{b}_t$  and  $\mathbf{W}_t$ , then the posterior distribution of  $\mathbf{b}_t$  and  $\mathbf{W}_t$  is

$$K(\mathbf{b}_t, \mathbf{W}_t|Y_t) \propto \prod_{n=1}^N L(y_{tn}|\mathbf{b}_t, \mathbf{W}_t) k(\mathbf{b}_t, \mathbf{W}_t).$$

It is customary to assume that  $\mathbf{b}_t$  and  $\mathbf{W}_t$  are independent, the prior on  $\mathbf{b}_t$  is normal with an unboundedly large variance, and the prior on  $\mathbf{W}_t$  is inverted Wishart with  $\kappa$  degrees of freedom and scale matrix  $\mathbf{I}$ , which is an identity matrix with rank  $\kappa$ . To improve computational efficiency, the hierarchical Bayes approach treats each  $\beta_{tn}$  as a parameter along with  $\mathbf{b}_t$  and  $\mathbf{W}_t$ , and the corresponding joint posterior for  $\mathbf{b}_t$ ,  $\mathbf{W}_t$ , and  $\beta_{tn} \forall n$  is then given by

$$K(\mathbf{b}_t, \mathbf{W}_t, \beta_{tn} \forall n|Y_t) \propto \prod_{n=1}^N L(y_{tn}|\beta_{tn}) \phi(\beta_{tn}|\mathbf{b}_t, \mathbf{W}_t) k(\mathbf{b}_t, \mathbf{W}_t), \tag{1}$$

which can be efficiently drawn using the Gibbs sampler [33].

### Likelihood ratio test for parameter comparison

Note that there are two types of parameters in the mixed logit model: a) the coefficient parameters, that is,  $\mathbf{b}_t$  and  $\mathbf{W}_t$ , which describe  $\beta_{tn}$ , the coefficients for the attributes; and b) the scale parameter, which measures the variability of the random term  $\epsilon_{istn}$ . When estimating the mixed logit model based on a single data set, the scale parameter is often normalized to one [33]. However, when several mixed logit models with the same specification are estimated on different data sets, it is then necessary to control for the differences in the scale parameters. This is typically done by normalizing the scale parameter associated with one of the data sets to one and estimating the scale parameters associated with the others.

Let  $D_1, D_2, \dots, D_T$  be the  $T$  data sets we collect (that is, one data set for each of the  $T$  types of questions), we can estimate  $T$  mixed logit models and get the corresponding parameter estimates. We want to see whether the parameters, that is, the coefficient parameters and the scale parameters, are the same across these data sets. This can be accomplished by the likelihood



ratio test, whose steps for the mixed logit model are similar to those for the multinomial logit model and we just need to account for the potential difference in the scale parameters of the data sets [35].

Suppose that we want to compare the parameter estimates obtained from Type 1 and Type 2 questions. For data set  $D_t$ , where  $t \in \{1, 2\}$ , let  $\mathbf{b}_t$  and  $\mathbf{W}_t$  be the coefficient parameters and  $\mu_t$  be corresponding scale parameter of the mixed logit model. The likelihood ratio test proceeds in two steps. In the first step, we test the hypothesis that the coefficient parameters are identical while allowing the scale parameters to be different. That is,

$$H_a : \quad \mathbf{b}_1 = \mathbf{b}_2 \text{ and } \mathbf{W}_1 = \mathbf{W}_2. \tag{2}$$

The likelihood ratio test involves estimating an unrestricted model and a restricted one. The unrestricted model consists of two separate mixed logit models, where the first one is fit over data set  $D_1$  and the second one is fit over data set  $D_2$ , independently of each other. The unrestricted model essentially returns the estimates for  $\mathbf{b}_1$ ,  $\mathbf{b}_2$ ,  $\mathbf{W}_1$ , and  $\mathbf{W}_2$ . In addition, the construction of the unrestricted model implies that we allow the scale parameters in  $D_1$  and  $D_2$  to be different. Let the sum of the loglikelihood function values for the two mixed logit models be  $L_1$ , which is also referred to as the loglikelihood for the unrestricted model. The restricted model is obtained by enforcing  $H_a$  in the unrestricted model. This can be conveniently achieved by estimating a mixed logit model over the following data matrix

$$\Omega = \begin{bmatrix} D_1 \\ \mu_2 D_2 \end{bmatrix},$$

which is created by multiplying all independent variables in  $D_2$  by the scale parameter  $\mu_2$ , whose value let us for now assume is known, and then concatenating with  $D_1$ . This implies that we normalize  $\mu_1$ , the scale parameter for  $D_1$ , to 1, and that  $\mu_2$  now reflects the relative scale of  $D_2$  with respect to  $D_1$ . The above construction imposes the restriction that  $\mathbf{b}_1 = \mathbf{b}_2$  and  $\mathbf{W}_1 = \mathbf{W}_2$  while allowing the scale parameters to be different. The estimation of the mixed logit model for  $\Omega$  is conducted by systematically varying the value of  $\mu_2$  and finding the one that maximizes the value of the maximum loglikelihood function. Let the maximum value be  $L_2$ . The test statistic for  $H_a$  is  $-2(L_2 - L_1)$ , which is  $\chi^2$  distributed with  $2\kappa - 1$  degrees of freedom.

If  $H_a$  cannot be rejected, we proceed to the second step and test the hypothesis that the scale parameters are identical between  $D_1$  and  $D_2$ , that is,

$$H_b : \quad \mu_1 = \mu_2. \tag{3}$$

This can be achieved by conducting a second likelihood ratio test, where the unrestricted model is estimated based on data matrix  $\Omega$  and the restricted model is estimated based on data matrix  $\Omega'$  defined below

$$\Omega' = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}. \tag{4}$$

This implies that we equalize  $\mu_1$  and  $\mu_2$ , and that both are normalized to one. Let  $L_3$  be the loglikelihood for the model based on  $\Omega'$  and the test statistics is  $-2(L_3 - L_2)$ , which is  $\chi^2$  distributed with 1 degree of freedom. The above procedure can be easily extended to situations where we need to compare parameter estimates across more than two data sets as is shown in Section 5.1.

### Conditional WTP estimates

A key output of a discrete choice model is the WTP measurement for various attributes of the alternatives. In addition to producing the WTP estimates for the entire sample, the mixed logit model can also estimate each individual's WTP conditional upon his or her observed choices. We can then conduct further analysis, for example, using a random-effects model, to relate the variation in the individual specific WTP to factors such as the characteristics of the individuals and the treatment of alternatives' names. This type of two-step analysis, that is, a mixed logit model followed by a random-effects model, is first proposed by [5] to estimate the economic benefits associated with rural landscape improvement. Recently, [9] and [36] apply this method to analyze preference heterogeneity in residential choice problems.

Recall that when estimating the mixed logit model based on  $Y_t$  in Section 3.1, that is, the responses to type  $t$  questions, through the hierarchical Bayes method, each individual's  $\beta_{tn}$  is treated as a random variable along with  $b_t$  and  $W_t$ . Based on the joint posterior for  $b_t$ ,  $W_t$ , and  $\beta_{tn} \forall n$  defined in Eq (1), we can obtain the posterior distribution of  $\beta_{tn}$  by integrating out variables  $b_t, W_t, \beta_{t1}, \beta_{t2}, \dots, \beta_{t, n-1}, \beta_{t, n+1}, \dots, \beta_{tN}$  in the joint posterior:

$$h(\beta_{tn}|Y_t) = \int \int \dots \int K(b_t, W_t, \beta_{tn} \forall n | Y_t) db_t dW_t d\beta_{t1} d\beta_{t2} \dots d\beta_{t, n-1} d\beta_{t, n+1} \dots d\beta_{tN}. \tag{5}$$

As a result, the mean  $\beta_{tn}$  for individual  $n$  in the sampled population can be written as

$$\bar{\beta}_{tn} = E[\beta_{tn}|Y_t] = \int \beta_{tn} h(\beta_{tn}|Y_t) d\beta_{tn} \tag{6}$$

$$= \int \int \dots \int \beta_{tn} K(b_t, W_t, \beta_{tn} \forall n | Y_t) db_t dW_t d\beta_{t1} d\beta_{t2} \dots d\beta_{tN} \tag{7}$$

The last equality in Eq (7) is obtained by plugging in the result for  $h(\beta_{tn}|Y_t)$  from Eq (5). The above multi-dimensional integral does not have a closed form and is calculated by simulation. Note that in the hierarchical Bayes approach, we have already obtained the draws for  $b_t, W_t$ , and  $\beta_{tn} \forall n$  from their joint posterior defined in Eq (1). Let  $R$  be the number of draws after the burn-in period, and  $\beta_{tn}^r$  be the sampled value in the  $r$ -th draw, we have

$$\bar{\beta}_{tn} = \frac{1}{R} \sum_{r=1}^R \beta_{tn}^r. \tag{8}$$

Let  $\bar{\beta}_{tnj}$  be the coefficient for attribute  $j$  in vector  $\bar{\beta}_{tn}$ , then individual  $n$ 's WTP for attribute  $j$  is given by

$$WTP_{tnj} = - \frac{\bar{\beta}_{tnj}}{\beta_{tp}}, \tag{9}$$

where  $\beta_{tp}$  is the price coefficient. This individual level WTP estimate can then be related to the socioeconomic characteristics of each individual as well as the treatments of alternatives' names.

### The random-effects model

Once we obtain the conditional WTP estimates for each individual and each question type, we can pool them together to create a panel. According to [37], panel data can be analyzed by three types of models: the pooled regression model, the fixed-effects model, and the random-effects model. We adopt the random-effects model because it is capable of controlling for

unobserved heterogeneity among the individuals. Besides, it also allows us to include panel-invariant variables, for example, the socio-demographic characteristics of the individuals, among the regressors. The econometric specification of the random-effects model is as follows:

$$y_{tn} = \beta' x_{tn} + (\alpha + v_n) + \omega_{tn},$$

where  $n$  represents a given individual,  $t$  refers to a specific question type,  $y_{tn}$  is the WTP estimate of individual  $n$  for question type  $t$ ,  $x_{tn}$  is the vector of explanatory variables,  $\alpha$  is the intercept of the model,  $v_n$  is the unobserved individual-specific error term, and  $\omega_{tn}$  is the random disturbance. Note that the random-effects model divide the unobserved effect into an error term that is individual specific ( $v_n$ ) and a component that is common to all individuals and question types ( $\omega_{tn}$ ). The individual specific disturbance is constant across all WTP estimates observed for this individual regardless of the question type. Due to this treatment, our model is also referred to as the one-way random-effects model.

In our random-effects model, the dependent variable is the conditional WTP estimate specific to each individual and each question type. The independent regressors include the socio-economic characteristics and life styles of the individuals, dummy variables indicating the labeled or unlabeled treatment of alternatives' names in the design and the presentation phases of the choice experiment, and variables that describe the choice consistency of the individuals.

## Case study

The main objective of this research is to investigate the influence of labeled versus unlabeled treatments of alternatives' names in stated choice experiments. We examine the problem in the context of a mode choice study involving car, taxi, public transportation, and vehicle sharing in the city of Beijing. Section 4.1 gives the background of the study. Sections 4.2 and 4.3 present the design and administration of the questionnaire, respectively.

## Background to the mode choice study

Beijing has been undergoing a rapid motorization process in the past two decades. This rapid motorization process has created significant economic benefits as well as severe congestion in the city. To alleviate congestion, Beijing has been exploring a variety of means [38, 39], among which vehicle sharing has received great attention [40]. Vehicle sharing, also known as car sharing, is a program through which individuals can rent vehicles for a short period of time, often by the hour. Unlike classical car rental, vehicles used in vehicle sharing are positioned in unstaffed neighbourhood locations. Since vehicle sharing turns the fixed costs of ownership into variable costs, it provides individuals and families with periodic access to automobiles with relatively low cost. The major social benefit of vehicle sharing is reduced vehicle ownership and use intensity [41, 42].

The municipal government in Beijing is very interested in knowing residents' choice behavior between existing travel modes and vehicle sharing should it be introduced. This would help the government estimate its impact to vehicle ownership and/or usage. Moreover, the government also wants to identify the socioeconomic characteristics of the subpopulation that would favor vehicle sharing. This information can be used to better target potential vehicle sharing customers. However, little research has been conducted on Chinese residents' preferences toward vehicle sharing. [43] carry out a survey within Beijing in spring 2006 to explore vehicle sharing familiarity and response. They report that over 25% of respondents expressed a high level of interest in vehicle sharing and those interested in vehicle sharing were skewed towards younger age categories, higher education levels, and slightly higher income. [44] study vehicle

sharing in Shanghai and they find that those interested in vehicle sharing were more likely to be educated and had longer commuting times.

### Survey design

There are three major urban transportation modes in Beijing, that is, private car, taxi, and public transportation [45]. These three modes together with vehicle sharing form the four travel alternatives available to the respondents. These alternatives are named alternatives and the travel modes can be viewed as their names. Table 4 summarizes the attributes and their levels associated with these alternatives in our survey. The first row indicates the treatment of alternatives' names: either as attributes, or as labels. The first column shows the attributes for the alternatives, that is, mode, travel cost, parking cost, in-vehicle time, out-of-vehicle time, and number of transfers. Note that we break down the overall cost into two components: 1) travel cost, which refers to the gasoline expense for private car, the fare for public transportation and taxi, or the rental cost for vehicle sharing; and 2) parking cost, which refers to the cost paid to park the private car or rental car at the destination. The second column shows the attribute levels in the unlabeled design when alternative's names are treated as attributes. To prevent generating unrealistic alternatives, restrictions are applied to prohibit certain combinations of attribute levels in the unlabeled design. For example, when the mode is public transportation, the parking cost is restricted to zero. The third column shows the attribute levels in the labeled design when alternatives' names are treated as labels. In this design, each alternative (or travel mode) has its own attribute levels.

**Table 4. The attributes and their levels in our mode choice study.**

Treatment of alternatives' names in the design phase	Unlabeled treatment (Treat as attributes)	Labeled treatment (Treat as labels)	
Mode	Public Transportation (PT), Private car (PC) Taxi (TX) Vehicle sharing (VS)		
Travel cost (Yuan)	0.4, 2, 5, 15, 30, 45	Public transportation:	0.4, 2, 5
		Private car:	15, 30, 45
		Taxi:	15, 30, 45
		Vehicle sharing:	15, 30, 45
Parking cost (Yuan)	0, 10, 20	Public transportation:	0
		Private car:	0, 10, 20
		Taxi:	0
		Vehicle sharing:	0, 10, 20
In-vehicle time (min)	20, 40, 60	Public transportation:	40, 50, 60
		Private car:	20, 30, 40
		Taxi:	20, 30, 40
		Vehicle sharing:	20, 30, 40
Out-of-vehicle time (min)	5, 10, 15, 20	Public transportation:	10, 15, 20
		Private car:	5, 10, 15
		Taxi:	10, 15, 20
		Vehicle sharing:	5, 10, 15
Number of transfers	0, 1, 2	Public transportation:	0, 1, 2
		Private car:	0
		Taxi:	0
		Vehicle sharing:	0

<https://doi.org/10.1371/journal.pone.0178826.t004>

Given the attributes and their levels in Table 4, we use the D-optimal design SAS macros provided by [46] to come up with the optimal subset of choice sets under both designs. For a given choice set, regardless of whether it comes from the labeled or the unlabeled design, we can present it using either the labeled presentation style, where the mode names are displayed in the header row, or the unlabeled presentation style, where the mode names are displayed along with the other attributes. In this way, we obtain four choice task pools, each of which contains one type of questions to be answered by the respondents. For ease of exposition, let  $\mathcal{P}_i$  be the pool of choice tasks for question type  $i \in \{1, 2, \dots, 4\}$ .

## Survey administration

Our goal is to assess the impact of the different question types to respondents' choice behavior. According to [17], there are two approaches to administer the survey: 1) the split sample approach, where statistically equivalent samples are presented with different versions of the questionnaires; and 2) administer versions of the questionnaire in the same group of respondents where the order is completely randomized, as long as the workload is appropriate. In our research, we adopt the second approach where the sample is presented with choice tasks from all four types of questions.

The survey is conducted on computers. When the respondent starts the survey, our computer program randomly selects three choice tasks from each of the four pools of choice tasks, that is, each of  $\mathcal{P}_1$ ,  $\mathcal{P}_2$ ,  $\mathcal{P}_3$ , and  $\mathcal{P}_4$ . These twelve choice tasks are then combined with a specially designed screening choice task which contains a dominant alternative (that is, an alternative that is logically preferable than the other alternatives in the choice set). This screening choice task is included in all questionnaires and is used to identify respondents with possible violations of rational preference. All thirteen questions are mixed in a random order to eliminate possible ordering effects among the choice tasks.

The pilot survey was administered in December 2012 to detect potential issues in the questionnaire. For example, we got inquiries from the respondents for Types 3 and 4 choice tasks. These two types of questions are generated by treating mode names as attributes in the design phase. As a result, the same mode may appear more than once in the choice tasks. Respondents asked us if we had made errors in the questionnaire. To dispel such confusion in the main survey, we explicitly told the respondents that the same mode might appear more than once and they needed to compare the detailed attribute levels to make choices. The main survey was conducted in May 2013. Both the pilot and the main surveys were administered in the city of Beijing by intercepting residents on the street. Since the concept of vehicle sharing is new to most Chinese residents, an introduction to vehicle sharing was given to the respondents before the survey. The introduction explained what vehicle sharing is, how to use shared vehicles, and the differences between vehicle sharing and traditional car rental. After they responded to all choice tasks, we also collected respondents' socio-demographic information, such as their ages, whether they own private cars, and so on. We also solicited their responses to know more about their life styles. For example, we asked about their frequently used travel modes and their environmental attitude. This information can be used to identify the subpopulation that would be interested in vehicle sharing. In the main survey, we collected responses from a total of 230 individuals, among which 198 correctly answered the screening choice task.

## Results

As is said earlier, in our survey each respondent faces three choice tasks from each of the four question types. We group all responses for type  $i$  questions into data set  $D_i$ , where  $i \in \{1, 2, 3, 4\}$ . We first investigate if the parameter estimates from the four data sets are the same in

Section 5.1. We then estimate the conditional WTP for each individual and each question type, and use a random-effects model to relate its variation to the socioeconomic characteristics of the individuals as well as the treatment of alternatives' names in Section 5.2.

### Impact toward parameter estimates

We first investigate if there are significant differences in the parameter estimates obtained from these four data sets. In the specification of the mixed logit models, we consider the following independent variables: the negative of in-vehicle time, the negative of out-of-vehicle time, the negative of the number of transfers, travel cost, parking cost, and three alternative-specific constants for private car (denoted as ASC\_PC), vehicle sharing (denoted as ASC\_VS), and taxi (denoted as ASC\_TX). For the variables that we use the negatives, their respective coefficients are modeled using the log-normal distribution, because we want the utility to decline when the the values of those variables increase. The coefficients for the three alternative-specific constants, that is, ASC\_PC, ASC\_VS and ASC\_TX, are assumed to be normally distributed, while those for travel cost and parking cost are assumed to be fixed numbers as is suggested by [33].

Let  $\mathbf{b}_i$  and  $\mathbf{W}_i$  be the coefficient parameters and  $\mu_i$  be the scale parameter estimated using data set  $D_i$ , where  $i = 1, 2, 3, 4$ . The estimation results are reported in Table 5. Columns 1 and 2 identify the coefficient parameters and the scale parameters. Columns 3 and 4 show the parameter estimates obtained from data set  $D_1$ . The coefficient for transportation cost and parking cost are both negative, which is consistent with our expectation. In addition, the coefficient for parking cost is more negative in magnitude than that for travel cost, which implies that individuals are more sensitive to parking expenses. The means for the three alternative-specific constants are all negative, which indicates that travelers would prefer public transportation if all else are held equal. We also notice that the standard deviations of the coefficients are highly significant, which suggests that there is significant behavioral heterogeneity among the respondents. The value of the scale parameter, that is,  $\mu_1$ , is normalized to 1 for convenience. The results for data sets 2, 3, and 4 are shown in Columns 5 through 10, and similar observations can be made. Note that we shall not directly compare the parameter estimates reported in Columns 3 through 10, because their respective scale parameters are normalized to one.

Following the steps in Section 3.2, we first test if the coefficient parameters are the same across all these four data sets, that is, we test the following hypothesis:

$$H_{1a}: \mathbf{b}_1 = \mathbf{b}_2 = \mathbf{b}_3 = \mathbf{b}_4 \text{ and } \mathbf{W}_1 = \mathbf{W}_2 = \mathbf{W}_3 = \mathbf{W}_4.$$

To test  $H_{1a}$ , we estimate an unrestricted model which contains four mixed logit models for  $D_1, D_2, D_3$ , and  $D_4$ , respectively. The estimation results have been reported in Columns 3 through 10 in Table 5 and the corresponding loglikelihood for the unrestricted model is  $(-801.3) + (-790.5) + (-714.3) + (-707.5) = 3013.6$ . The restricted model is obtained by estimating a mixed logit model over the data matrix formed by pooling  $D_1, D_2, D_3$ , and  $D_4$  together while allowing the scale parameters to be different. This is achieved by normalizing  $\mu_1$  to 1 and systematically varying the values of  $\mu_2, \mu_3$ , and  $\mu_4$  independently between 0 and 10.0 with a step size of 0.02. For each combination of  $\mu_2, \mu_3$ , and  $\mu_4$ , we can obtain the maximum value for the loglikelihood function. We find that the loglikelihood function reaches its maximum of -3028.4 when  $\mu_2 = 1.06, \mu_3 = 0.76$ , and  $\mu_4 = 0.72$ . The corresponding parameter estimates are reported in Columns 11 and 12 of Table 5. The statistic for the likelihood ratio test is  $-2[-3028.4 - (-3013.6)] = 29.6$ , which is less than the critical value of the Chi-squared distribution with 39 degrees of freedom at the 5% level of significance ( $\chi_{39,0.05}^2 = 54.57$ ). Therefore, we

**Table 5. Estimation results for the mixed logit models.**

Variables (1)	Parameters (2)	Data set $D_1$		Data set $D_2$		Data set $D_3$		Data set $D_4$		Pooled Data set 1 (Different scale parms.)		Pooled Data set 2 (Identical scale parms.)	
		Value (3)	S.E. (4)	Value (5)	S.E. (6)	Value (7)	S.E. (8)	Value (9)	S.E. (10)	Value (11)	S.E. (12)	Value (13)	S.E. (14)
In-vehicle time (neg.)	Mean of ln(coef.)	-3.0021 <sup>‡</sup>	0.1781	-2.5938 <sup>‡</sup>	0.1550	-3.2074 <sup>‡</sup>	0.1450	-3.0022 <sup>‡</sup>	0.1285	-2.9642 <sup>‡</sup>	0.0715	-2.9350 <sup>‡</sup>	0.0742
	S.D. of ln(coef.)	1.4942 <sup>‡</sup>	0.4097	0.7348 <sup>‡</sup>	0.2310	0.9461 <sup>‡</sup>	0.2210	1.3528 <sup>‡</sup>	0.3521	0.8904 <sup>‡</sup>	0.1671	0.9605 <sup>‡</sup>	0.1873
Out-of-vehicle time (neg.)	Mean of ln(coef.)	-3.2177 <sup>‡</sup>	0.1887	-3.5634 <sup>‡</sup>	0.4751	-3.5729 <sup>‡</sup>	0.4381	-4.9587 <sup>‡</sup>	0.6091	-3.4616 <sup>‡</sup>	0.2083	-3.4650 <sup>‡</sup>	0.2194
	S.D. of ln(coef.)	1.1561 <sup>‡</sup>	0.3023	2.5835 <sup>†</sup>	1.1595	0.9282 <sup>†</sup>	0.4448	2.8959 <sup>†</sup>	1.1505	1.1742 <sup>‡</sup>	0.3731	1.2834 <sup>‡</sup>	0.4087
No. of transfers (neg.) (Specific to public transport.)	Mean of ln(coef.)	-0.4098*	0.2456	-0.5850 <sup>†</sup>	0.2904	-1.2246 <sup>†</sup>	0.2606	-0.9408 <sup>†</sup>	0.2109	-1.7102 <sup>†</sup>	0.4376	-1.8903 <sup>†</sup>	0.5240
	S.D. of coef.	2.4049 <sup>†</sup>	1.0679	2.7559 <sup>†</sup>	1.1154	1.4520 <sup>†</sup>	0.6075	1.5409 <sup>†</sup>	0.6177	4.3402*	2.2466	5.0992 <sup>†</sup>	2.3802
ASC_PC (Specific to private car)	Mean of coef.	-0.2109	0.2685	-0.7477*	0.4102	-0.4149	0.3709	-0.9621 <sup>‡</sup>	0.3402	-0.6128 <sup>‡</sup>	0.1831	-0.6292 <sup>‡</sup>	0.1913
	S.D. of coef.	2.2965 <sup>‡</sup>	0.7152	8.1717 <sup>‡</sup>	2.9035	4.1294 <sup>†</sup>	1.8717	2.2261 <sup>‡</sup>	0.6829	4.8320 <sup>‡</sup>	1.0753	5.7049 <sup>‡</sup>	1.1437
ASC_VS (Specific to vehicle sharing)	Mean of coef.	-0.1156	0.2444	-0.3718	0.3793	-0.0366	0.2477	-0.9626 <sup>‡</sup>	0.2650	-0.3267*	0.1687	-0.2983*	0.1729
	S.D. of coef.	2.0229 <sup>†</sup>	0.9167	6.9754 <sup>‡</sup>	1.7912	2.0107 <sup>†</sup>	0.7921	2.4146 <sup>‡</sup>	0.8403	3.4518 <sup>‡</sup>	0.8295	3.8671 <sup>‡</sup>	0.7787
ASC_TX (Specific to taxi)	Mean of coef.	-0.0913	0.2045	-0.2747	0.2997	-0.1560	0.2350	-0.6545 <sup>†</sup>	0.2820	-0.2908 <sup>†</sup>	0.1447	-0.2809 <sup>†</sup>	0.1481
	S.D. of coef.	1.4496 <sup>†</sup>	0.5936	2.4592 <sup>‡</sup>	-0.6545	1.2929*	0.6739	4.5363 <sup>‡</sup>	1.4034	1.9860 <sup>‡</sup>	0.5556	2.1792 <sup>‡</sup>	0.5592
Travel cost	Coef.	-0.0630 <sup>‡</sup>	0.0059	-0.0658 <sup>‡</sup>	0.0066	-0.0530 <sup>‡</sup>	0.0065	-0.0315 <sup>‡</sup>	0.0073	-0.0477 <sup>‡</sup>	0.0034	-0.0510 <sup>‡</sup>	0.0036
Parking cost	Coef.	-0.0843 <sup>‡</sup>	0.0127	-0.1009 <sup>‡</sup>	0.0161	-0.0968 <sup>‡</sup>	0.0133	-0.0741 <sup>‡</sup>	0.0128	-0.0774 <sup>‡</sup>	0.0058	-0.0827 <sup>‡</sup>	0.0061
$\mu_1$		1.00								1.00		1.00	
$\mu_2$				1.00						1.04		1.00	
$\mu_3$						1.00				0.76		1.00	
$\mu_4$								1.00		0.74		1.00	
No. of observations		690		690		690		690		2760		2760	
Loglikelihood		-801.3		-790.5		-714.3		-707.5		-3028.4		-3034.6	
Adjusted $\rho^2$		0.1476		0.1590		0.2386		0.2458		0.2041		0.2032	

‡ significant at 1% level  
 † significant at 5% level  
 \* significant at 10% level

<https://doi.org/10.1371/journal.pone.0178826.t005>

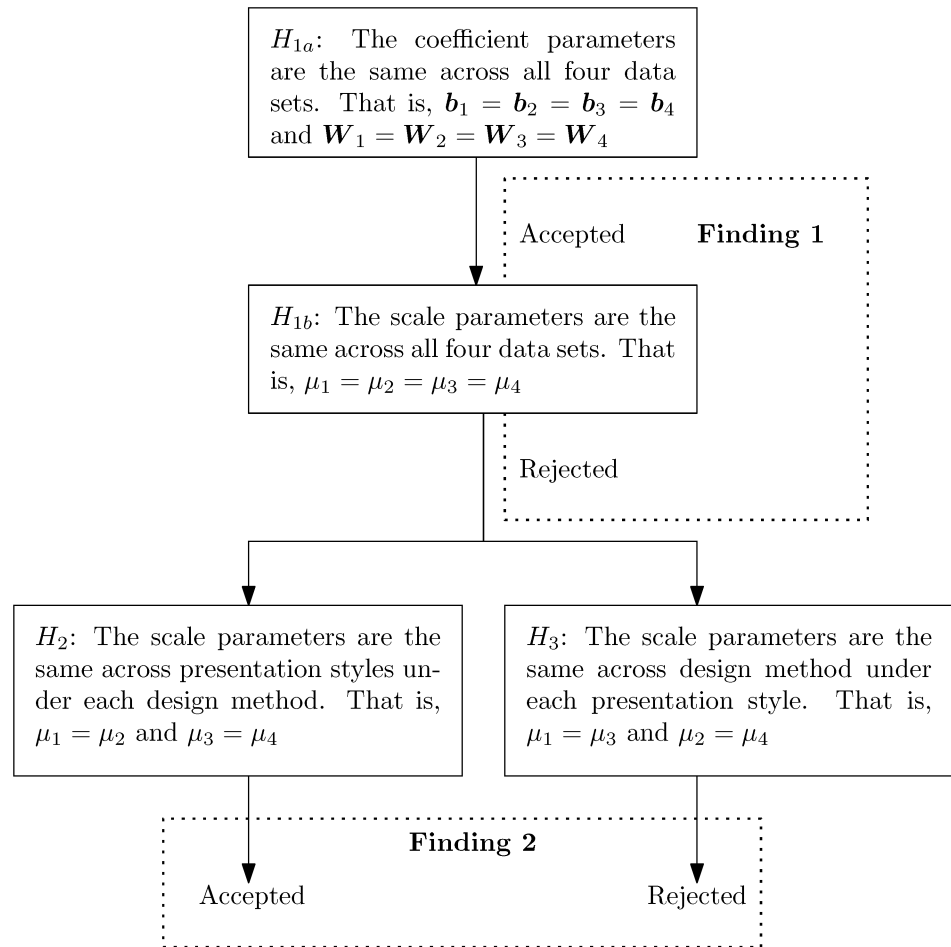
cannot reject  $H_{1a}$ , which means that the coefficient parameters are identical across these four data sets.

Since  $H_{1a}$  stands, we proceed to check if the scale parameters are the same among these four data sets:

$$H_{1b}: \mu_1 = \mu_2 = \mu_3 = \mu_4.$$

This is accomplished by estimating the unrestricted model reported in Columns 11 and 12 in Table 5 and a restricted model over the data matrix formed by pooling all four data matrices together while restricting their scale parameters to be identical. The parameter estimates for the restricted model are reported in Columns 13 and 14 in Table 5. Note that the values of the scale parameters are all normalized to 1.0 and the loglikelihood is -3034.6. The test statistic for  $H_{1b}$  is  $-2[-3034.6 - (-3028.4)] = 12.8$ , which exceeds the critical value of the Chi-square distribution with 3 degrees of freedom at the 5% level of significance ( $\chi^2_{3,0.05} = 7.82$ ). Hence, the hypothesis is rejected and there exists significant differences in the scale parameters of the four data sets. The findings related to  $H_{1a}$  and  $H_{1b}$  can be summarised as follows:

**Finding 1** The labeled or the unlabeled treatment of alternatives' names in either the design or the presentation phases of the choice experiment does not statistically affect the estimates of the coefficient parameters. However, these different treatments do cause the scale parameters to differ.



**Fig 2. The hypothesis tests we conduct on the four data sets.** The results are summarized into Findings 1 and 2 in the text.

<https://doi.org/10.1371/journal.pone.0178826.g002>

The sequence of the above hypothesis tests is illustrated by the flow chart shown in Fig 2. At the top of the figure, we test  $H_{1a}$  and accept this hypothesis. Afterwards, we test  $H_{1b}$  and it gets rejected. These two results are summarized into Finding 1. Since  $H_{1b}$  is rejected, which means that the scale parameters are not identical across the four data sets, we conduct additional hypothesis tests to further analyze the cause: Is it due to the treatment of alternatives' names in the design or in the presentation phase? This is accomplished by testing  $H_2$  and  $H_3$ . Hypothesis  $H_2$  tests the following hypothesis: Given the treatment of alternatives' names in the design phase, the scale parameters are the same between the labeled and the unlabeled treatments of alternatives' names during the presentation phase. Mathematically, this means

$$H_2 : \quad \mu_1 = \mu_2 \text{ and } \mu_3 = \mu_4.$$

Hypothesis  $H_3$  tests the following hypothesis: Given the treatment of alternatives' names in the presentation phase, the scale parameters are the same between the labeled and the unlabeled treatments of alternatives' names during the design phase. That is,

$$H_3 : \quad \mu_1 = \mu_3 \text{ and } \mu_2 = \mu_4.$$



Table 6. Detailed steps for the likelihood ratio tests.

Hypothesis		Unrestricted model		Restricted model		Result
		Data matrix	Log(L)	Data matrix	Log(L)	
$H_{1a}$	$b_1 = b_2 = b_3 = b_4;$ $W_1 = W_2 = W_3 = W_4$	$D_1$	-801.3	$\begin{bmatrix} D_1 \\ \mu_2 D_2 \\ \mu_3 D_3 \\ \mu_4 D_4 \end{bmatrix}$	-3028.4	Test statistic = 29.6 Degree of freedom = 39 Critical value = 54.57 Hypothesis is accepted
		$D_2$	-790.5		$\mu_2 = 1.04$	
		$D_3$	-714.3		$\mu_3 = 0.76$	
		$D_4$	-707.5		$\mu_4 = 0.74$	
$H_{1b}$	$\mu_1 = \mu_2 = \mu_3 = \mu_4$	$\begin{bmatrix} D_1 \\ \mu_2 D_2 \\ \mu_3 D_3 \\ \mu_4 D_4 \end{bmatrix}$	-3028.4 $\mu_2 = 1.04$ $\mu_3 = 0.76$ $\mu_4 = 0.74$	$\begin{bmatrix} D_1 \\ D_2 \\ D_3 \\ D_4 \end{bmatrix}$	-3034.6	Test statistic = 12.4 Degree of freedom = 3 Critical value = 7.82 Hypothesis is rejected
$H_2$	$\mu_1 = \mu_2$ and $\mu_3 = \mu_4$	$\begin{bmatrix} D_1 \\ \mu_2 D_2 \\ \mu_3 D_3 \\ \mu_4 D_4 \end{bmatrix}$	-3028.4 $\mu_2 = 1.04$ $\mu_3 = 0.76$ $\mu_4 = 0.74$	$\begin{bmatrix} D_1 \\ D_2 \\ \mu_3 D_3 \\ \mu_3 D_4 \end{bmatrix}$	-3029.6 $\mu_3 = 0.74$	Test statistic = 2.4 Degree of freedom = 2 Critical value = 5.99 Hypothesis is accepted
$H_3$	$\mu_1 = \mu_3$ and $\mu_2 = \mu_4$	$\begin{bmatrix} D_1 \\ \mu_2 D_2 \\ \mu_3 D_3 \\ \mu_4 D_4 \end{bmatrix}$	-3028.4 $\mu_2 = 1.04$ $\mu_3 = 0.76$ $\mu_4 = 0.74$	$\begin{bmatrix} D_1 \\ \mu_2 D_2 \\ D_3 \\ \mu_2 D_4 \end{bmatrix}$	-3034.5 $\mu_2 = 0.92$	Test statistic = 12.2 Degree of freedom = 2 Critical value = 5.99 Hypothesis is rejected

<https://doi.org/10.1371/journal.pone.0178826.t006>

The details for the likelihood ratio tests associated with  $H_{1a}$ ,  $H_{1b}$ ,  $H_2$ , and  $H_3$  are conveniently summarized in Table 6. Column 1 shows the hypothesis in question. Columns 2 and 3 show the data matrices and the loglikelihood function value for the unrestricted model. For example, while testing  $H_{1a}$  in Row 1, the log likelihood based on data matrices  $D_1$ ,  $D_2$ ,  $D_3$ , and  $D_4$  are -801.3, -790.5, -714.3, and -707.5, respectively. Columns 4 and 5 show the data matrix for the restricted model and the log likelihood function value. The values of the scale parameters are also reported if applicable. In the test for  $H_{1a}$ , the loglikelihood function of the restricted model reaches its maximum value of -3028.4 when  $\mu_1 = 1$ ,  $\mu_2 = 1.06$ ,  $\mu_3 = 0.76$ , and  $\mu_4 = 0.72$ . Finally, Column 6 shows the details of the likelihood ratio test, including the test statistic, the degree of freedom and critical value for the  $\chi^2$  test, and whether the hypothesis is accepted or rejected. Rows 2 and 3 show the results for  $H_{1a}$  and  $H_{1b}$ , respectively. In Row 4, we show that  $H_2$  is accepted, meaning that the treatment of alternatives' names in the presentation phase does not affect the magnitude of the scale parameter. In Row 5, we show that  $H_3$  is rejected, meaning that the treatment of alternatives' names in the design phase does affect the magnitude of the scale parameter.

Based on the above results,  $\beta_{tm}$  can be assumed to come from the same distribution with parameters  $\mathbf{b}$  and  $\mathbf{W}$ , for all  $t$ . As a result, the posterior distribution of  $\mathbf{b}$  and  $\mathbf{W}$  based on all four data sets is given by

$$K(\mathbf{b}, \mathbf{W}, \beta_{tm} \forall n \forall t | \mathbf{Y}) \propto \prod_{t=1}^T \prod_{n=1}^N L(\mathbf{y}_n | \beta_{tm}) \phi(\beta_{tm} | \mathbf{b}, \mathbf{W}) k(\mathbf{b}, \mathbf{W}), \tag{10}$$

where  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \mathbf{Y}_4)$ . This posterior is used in the conditional WTP analysis because it utilizes all data sets in a single model.

## Impact toward willingness-to-pay measures

An important output of behavioral models is the WTP measures. In our mode choice study, we are particularly interested in the WTP estimate for vehicle sharing. On one hand, we want to know whether the WTP estimate is affected by the design types and presentation styles; on the other hand, we are also keen to identify the sub-population that would be interested in vehicle sharing, that is, the individuals that have higher WTP for vehicle sharing, which shall help vehicle sharing companies better target their customers.

The panel is created by pooling the WTP estimates for the four types of questions for each individual. We then relate these individual level WTP estimates to the socioeconomic characteristics of each individual as well as the treatment of alternatives' names in the design and the presentation phases. For a choice task, let indicator variable `LBD_DSGN` take value 1 if it is produced by treating the alternatives' names as labels in the design phase; and zero, otherwise. Similarly, let indicator variable `LBD_PRSN` take value 1 if this question is produced by treating the alternatives' names as labels in the presentation phase; and zero, otherwise. For example, for a choice task that belongs to Type 1 question in [Table 3](#), both `LBD_DSGN` and `LBD_PRSN` equal 1. To explain the variability in  $WTP_{mj}$ , we specify the following random-effects model:

$$WTP_{mj} = \gamma'z_n + \delta LBD\_DSGN + \pi LBD\_PRSN + (\alpha + v_n) + \omega_{nt},$$

where  $z_n$  is a vector of individual  $n$ 's socioeconomic characteristics, and  $\gamma$  is the corresponding coefficient vector;  $\delta$  and  $\pi$  are the coefficients for the two dummy variables;  $\alpha$  is the intercept term and is constant for all individuals;  $v_n$  is the random heterogeneity specific to individual  $n$  and is constant across question types faced by individual  $n$ ; and  $\omega_{nt}$  is the error term. The estimate for the coefficient vector  $\gamma$  can uncover the socioeconomic causes for the preference heterogeneity toward attribute  $j$ , while those for  $\delta$  and  $\pi$  can quantify the impact of labeled versus unlabeled treatments of alternatives' names. Note that since we have a separate individual error term, the above model allows us to include individual specific socioeconomic variables in the regressors.

[Table 7](#) summarizes the dependent and independent variables used in the model. The dependent variable is the WTP estimates for vehicle sharing. The independent variables are classified into four groups, where the first group is related to the socio-demographic characteristics of the respondents: the age and gender of the respondent, and whether he/she owns a vehicle. The age of the respondent is classified into three dummy variables: `AGE1` takes value 1 if the age of the respondent is no more than 22 years old, and zero otherwise; `AGE2` takes value 1 if the age of the respondent is greater than 22 years old but no more than 27 years old; and `AGE3` takes value 1 if the age of the respondent is greater than 27 years old, and zero otherwise. This type of demographic information can be easily obtained from public sources and the insights from the random-effects model can be used to identify the sub-population that would be interested in vehicle sharing. The second group is about the life styles of the respondents, that is, whether he or she is a frequent user of public transportation, private car, or taxi; and whether he/she agrees with the statement that "I consider the environmental impact of the transportation mode I use". The third group of variables indicate the design type and presentation style of the questions: Dummy variables `LBD_DSGN` and `LBD_PRSN` distinguish the treatment of alternatives' names in the design and the presentation phase, respectively. The last group of variables detects possible violations of rational preference. Dummy variable `NON_MON` takes value 1 if the respondent fails to select the dominant alternative in the specially designed screening choice set. Note that the violation of non-monotonic preference does not necessarily imply theoretical inconsistency or irrationality of the respondent [5]. It may be due to the fact that the respondents have perceptions of missing attributes. As a result, these

**Table 7. Variables used in the random-effects model to explain the heterogeneity in the WTP estimate for vehicle sharing.**

Variable	Description
<b>Dependent variable</b>	
WTP_VS	Conditional WTP for vehicle sharing (specific to each individual and question type)
<b>Socio-demographic characteristics</b>	
AGE1	Dummy variable that equals 1 when the respondent is no more than 22 years old; and zero, otherwise
AGE2	Dummy variable that equals 1 when the respondent is greater than 22 years old but no more than 27 years old; and zero, otherwise
AGE3	Dummy variable that equals 1 when the respondent is greater than 27 years old; and zero, otherwise
OWN_CAR	Dummy variable that equals 1 when the respondent owns a private car; and zero, otherwise
<b>Life Style</b>	
FRQT_PT	Dummy variable that equals 1 when the respondent is a frequent user of public transportation; and zero, otherwise
FRQT_TX	Dummy variable that equals 1 when the respondent is a frequent user of taxi; and zero, otherwise
ENV_FRNDLY	Dummy variable that equals 1 when the respondent reports that he/she considers the environmental impact of the travel mode he/she uses; and zero, otherwise
<b>Treatment of alternatives' names</b>	
LBD_DSGN	Dummy variable that equals 1 when mode names are treated as labels in the design phase; and zero, otherwise
LBD_PRSN	Dummy variable that equals 1 when mode names are presented as labels in the presentation phase; and zero, otherwise
<b>Possible violation of rational preference</b>	
NON_MON	Dummy variable for non-monotonic preferences. It equals 1 when the respondent failed to choose the dominant alternative in the specially designed screening choice task

<https://doi.org/10.1371/journal.pone.0178826.t007>

responses are not excluded from the WTP estimation. However, it is expected that these anomalous individuals shall have greater WTP estimates.

The estimation results for the random-effects model are reported in Table 8. We find that the coefficients for the age dummy variables, that is, AGE1 and AGE3, are not significant, which suggests that the interest in vehicle sharing does not vary significantly across different age groups. This finding is similar to that reported in [44], where the authors find that the coefficient for age is not significant. The coefficient for OWN\_CAR is positive and significant, which means that vehicle owners are highly interested in vehicle sharing. This is in line with existing findings that vehicle sharing programs are fairly attractive to vehicle owners and contribute to the reduction in vehicle ownership [41]. For the three life style related variables, the coefficient for FRQT\_PT is negative and significant, implying that public transportation users are not quite interested in vehicle sharing. This is understandable because the cost of vehicle sharing is much higher than that for public transportation. The coefficient for FRQT\_TX is positive and significant, implying that frequent users of taxis have a higher WTP for vehicle sharing. This means that vehicle sharing service providers can launch targeted marketing campaigns toward taxi passengers at taxi hailing stations or through smart-phone-based taxi hailing applications. The coefficient for ENV\_FRNDLY is positive but insignificant, which suggests that the interest in vehicle sharing is only weakly correlated with the environment attitude of the individuals. The above results can be summarized into the following finding:

**Table 8. Parameter estimates for the random-effects model.**

	Coefficient	S.E.
Intercept	-10.2720 <sup>‡</sup>	3.2507
AGE1	-1.1083	2.1138
AGE3	3.2682	2.3419
OWN_CAR	4.7624 <sup>†</sup>	1.8839
FRQT_PT	-6.0769 <sup>‡</sup>	2.2323
FRQT_TX	1.5060 <sup>†</sup>	1.8702
ENV_FRNDLY	1.2559	1.9240
LBD_DSGN	0.9204	1.0859
LBD_PRSN	2.3303 <sup>†</sup>	1.0859
NON_MON	13.5620 <sup>‡</sup>	2.5832
S.D. of $v_n$	10.0510	
S.D. of $\omega_{nt}$	16.4690	
Lagrange multiplier test	76.7200	
Adjusted $R^2$	0.3045	
Num. of observations	920	

<sup>‡</sup> significant at 1% level

<sup>†</sup> significant at 5% level

<https://doi.org/10.1371/journal.pone.0178826.t008>

**Finding 2** We find that individuals who are vehicle owners and who frequently use taxis have higher WTP for vehicle sharing. However, it appears that age and environmental attitude do not affect the WTP estimate.

The coefficient for LBD\_DSGN is not significant, which suggests that there is not much difference in the WTP estimates between the labeled and unlabeled treatments of alternatives' names in the design phase. However, the coefficient for LBD\_PRSN is positive and significant at the 5% level of significance, implying that the labeled treatment of alternatives' names leads to higher WTP estimate than the unlabeled treatment in the presentation phase. This is consistent with our expectation, because when the names of the modes are displayed prominently in the header row, they are likely to become the target of choice and the actual attributes of the alternatives may be overlooked by the respondents, which would cause the WTP estimate to be greater. The dummy variable NON\_MON is used to test the sensitivity of WTP for respondents who failed to detect dominant alternatives in the screening choice task. In total, 32, or 14%, respondents did not recognize the dominant alternative. The coefficient for this variable is positive and significant at the 1% level of significance, which is consistent with existing studies that respondents who fail to detect the dominant alternative tend to have higher WTP estimates than those who do [5]. The above results are summarized into the following finding:

**Finding 3** Given the treatment of alternatives' names in the presentation phase, the treatment of alternatives' names in the design phase does not statistically affect the estimates of the WTP measures. However, given the treatment of alternatives' names in the design phase, the labeled treatment of alternatives' names in the presentation phase causes the corresponding WTP estimates to be slightly higher.

Now let us look at the standard deviations of the two error terms. We notice that the standard deviation of  $v_n$  is comparable to that of  $\omega_{nt}$ , which suggests that there is indeed considerable individual specific variability in the responses. To formally evaluate the validity of the random-effects model, the Lagrange multiplier test, developed by [47], is conducted to test the null hypothesis if the variance of  $v_n$  is zero. The test statistic is 76.7200, which exceeds the

critical value of the Chi-square distribution at the 5% level of significance ( $\chi^2_{1,0.05} = 3.84$ ). Hence, we can reject the null hypothesis and conclude that the random-effects model is more appropriate than ordinary least squares, implying there are significant differences across individuals.

## Conclusions

Discrete choice experiments have been widely applied to elicit behavioral preference in the literature. In many of these experiments, the alternatives are *named alternatives*, meaning that they are naturally associated with specific names. A fundamental issue that arises in stated choice studies is whether to adopt the labeled or unlabeled treatment of alternatives' names in the design as well as the presentation phases of the choice sets. In the design phase, the labeled treatment of alternatives' names views the alternatives' names as labels and use the labeled design methods to generate the choice tasks, while the unlabeled treatment views the alternatives' names as attributes and use the unlabeled design methods instead. In the presentation phase, the labeled treatment of alternatives' names displays the alternatives' names in the header row of the choice task, while the unlabeled treatment displays the alternatives' names together with its attributes and use generic names in the header row to reference the alternatives.

We investigate the impact of labeled versus unlabeled treatments of alternatives' names on the outcome of stated choice experiments. In a study that aims at eliciting Chinese residents' preferences toward vehicle sharing, we generate four types of survey questions that differ by the treatment of alternatives' names in the design and presentation phases. We estimate mixed logit models based on the collected data and find that parameter estimates are not affected by the treatment of alternative's names at the design as well as the presentation phases. However, we do observe considerable differences in the scale parameters, which is attributable to the treatment of alternatives' names in the design phase. We proceed to quantify the differences in WTP estimates. By building a random-effects model after we obtain the conditional WTP estimates for each individual and each question type, we find that the treatment of alternatives' names during the design phase does not affect the WTP estimates, while that during the presentation phase does. Specifically, we find that the labeled treatment of alternatives' names in the presentation phases produces slightly greater WTP estimates. We also find that vehicle owners and frequent users of taxis have higher WTP for vehicle sharing.

## Supporting information

**S1 File. The data set used in this study.**  
(ZIP)

## Acknowledgments

This research is supported by the Ford University Research Programs. While this article is believed to contain correct information, Ford Motor Company (Ford) does not expressly or impliedly warrant, nor assume any responsibility, for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, nor represent that its use would not infringe the rights of third parties. Reference to any commercial product or process does not constitute its endorsement. This article does not provide financial, safety, medical, consumer product, or public policy advice or recommendation. Readers should independently replicate all experiments, calculations, and results. The views and opinions expressed are of the

authors and do not necessarily reflect those of Ford. This disclaimer may not be removed, altered, superseded or modified without prior Ford permission.

## Author Contributions

**Conceptualization:** Hai Jiang, Yimin Liu, Erica Klampfl.

**Data curation:** Wen Jin, Hai Jiang.

**Formal analysis:** Wen Jin, Hai Jiang, Yimin Liu.

**Funding acquisition:** Hai Jiang, Yimin Liu, Erica Klampfl.

**Investigation:** Wen Jin, Hai Jiang, Yimin Liu.

**Methodology:** Wen Jin, Hai Jiang, Yimin Liu.

**Project administration:** Hai Jiang, Yimin Liu.

**Resources:** Hai Jiang.

**Software:** Wen Jin.

**Supervision:** Hai Jiang.

**Validation:** Wen Jin, Hai Jiang.

**Visualization:** Wen Jin, Hai Jiang.

**Writing – original draft:** Wen Jin.

**Writing – review & editing:** Hai Jiang.

## References

1. Saleh W, Farrell S. Implications of congestion charging for departure time choice: work and non-work schedule flexibility. *Transportation Research Part A*. 2005; 39:773–791.
2. Patil S, Burris M, Shaw D. Travel using managed lanes: an application of a stated choice model for Houston, Texas. *Transport Policy*. 2011; 18(4):595–603. <https://doi.org/10.1016/j.tranpol.2011.02.004>
3. Devarasetty PC, Burris M, Shaw D. The value of travel time and reliability-evidence from a stated preference survey and actual usage. *Transportation Research Part A*. 2012; 46(8):1227–1240.
4. Dellaert B, Donkers B, Soest AV. Complexity effects in choice experiment based models. *Journal of Marketing Research*. 2012; 49(3):424–434. <https://doi.org/10.1509/jmr.09.0315>
5. Campbell D. Willingness to pay for rural landscape improvements: combining mixed logit and random-effects models. *Journal of Agricultural Economics*. 2007; 58(3):467–483. <https://doi.org/10.1111/j.1477-9552.2007.00117.x>
6. Fimereli E, Mourato S. Assessing the effect of energy technology labels on preferences. *Journal of Environmental Economics and Policy*. 2013; 2(3):245–265. <https://doi.org/10.1080/21606544.2013.801183>
7. Viney R, Savage E, Louviere J. Empirical investigation of experimental design properties of discrete choice experiments in health care. *Health Economics*. 2005; 14(4):349–62. <https://doi.org/10.1002/hec.981> PMID: 15712274
8. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. *Journal of Health Economics*. 2012; 31(1):306–318. <https://doi.org/10.1016/j.jhealeco.2011.11.004> PMID: 22197308
9. Hoshino T. Estimation and analysis of preference heterogeneity in residential choice behaviour. *Urban Studies*. 2011; 48(2):363–382. <https://doi.org/10.1177/0042098010363498> PMID: 21275199
10. Liao FH, Farber S, Ewing R. Compact development and preference heterogeneity in residential location choice behaviour: A latent class analysis. *Urban Studies*. 2015; 52(2):314–337. <https://doi.org/10.1177/0042098014527138>
11. Hensher D, Rose J, Greene W. *Applied choice analysis: A primer*. Cambridge University Press; 2005.

12. Yang L. Modeling preferences for innovative modes and services: a case study in Lisbon. Massachusetts Institute of Technology. Cambridge, MA; 2010.
13. Ding C, Wang D, Liu C, Zhang Y, Yang J. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. *Transportation Research Part A*. 2017; 100:65–80.
14. Doherty E, Campbell D, Hynes S, van Rensburg TM. Examining labelling effects within discrete choice experiments: an application to recreational site choice. *Journal of Environmental Management*. 2013; 125:94–104. <https://doi.org/10.1016/j.jenvman.2013.03.056> PMID: 23648268
15. Bliemer MCJ, Rose JM. Experimental design influences on stated choice outputs: an empirical study in air travel choice. *Transportation Research Part A*. 2011; 45:63–79.
16. Wen CH, Lai SC. Latent class models of international air carrier choice. *Transportation Research Part E*. 2010; 46(2):211–221. <https://doi.org/10.1016/j.tre.2009.08.004>
17. Bekker-Grob ED, Hol L, Donkers B, van Dam L, Habbema DF, van Leerdam ME, et al. Labeled versus unlabeled discrete choice experiments in health economics: an application to colorectal cancer screening. *Value in Health*. 2010; 13(2):315–324. <https://doi.org/10.1111/j.1524-4733.2009.00670.x> PMID: 19912597
18. Louviere JJ, Hensher DA, Swait JD. *Stated Choice Methods: Analysis and Applications*. Cambridge University Press; 2000.
19. Caussade S, Ortúzar J, Rizzi L, Hensher D. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B*. 2005; 39(8):621–640. <https://doi.org/10.1016/j.trb.2004.07.006>
20. Blamey RK, Bennett JW, Louviere JJ, Morrison MD, Rolf J. A test of policy labels in environmental choice modeling studies. *Ecological Economics*. 2000; 32(2000):269–286. [https://doi.org/10.1016/S0921-8009\(99\)00101-9](https://doi.org/10.1016/S0921-8009(99)00101-9)
21. Ortúzar J, Iacobelli A, Valeze C. Estimating demand for a cycle-way network. *Transportation Research Part A*. 2000; 34(5):353–373.
22. Bhat CR, Castelar S. A unified mixed logit framework for modeling revealed and stated preferences: formulation and application to congestion pricing analysis in the San Francisco Bay area. *Transportation Research Part B*. 2002; 36(7):593–616. [https://doi.org/10.1016/S0191-2615\(01\)00020-0](https://doi.org/10.1016/S0191-2615(01)00020-0)
23. Cherchi E, Ortúzar J. Mixed RP/SP models incorporating interaction effects: modelling new suburban train services in Cagliari. *Transportation*. 2002; 29(4):371–395. <https://doi.org/10.1023/A:1016307308723>
24. Hensher DA, Greene WH. Specification and estimation of the nested logit model: alternative normalisations. *Transportation Research Part B*. 2002; 36(1):1–17. [https://doi.org/10.1016/S0191-2615\(00\)00035-7](https://doi.org/10.1016/S0191-2615(00)00035-7)
25. Cantillo V, Ortúzar J. A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B*. 2005; 39(8):641–657. <https://doi.org/10.1016/j.trb.2004.08.002>
26. Chrzan K. Three kinds of order effects in choice-based conjoint analysis. *Marketing Letters*. 1994; 5:165–172. <https://doi.org/10.1007/BF00994106>
27. van der Waerden P, Borgers A, Timmermans H, Béréanos M. Order effects in stated-choice experiments: study of transport mode choice decisions. *Transportation Research Record*. 2006; 1985:12–18. <https://doi.org/10.3141/1985-02>
28. Campbell D, Hutchinson WG, Scarpa R. Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental and Resource Economics*. 2008; 41(3):401–417. <https://doi.org/10.1007/s10640-008-9198-8>
29. Scarpa R, Gilbride T, Campbell D, Hensher D. Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*. 2009; 36(2):151–174. <https://doi.org/10.1093/erae/jbp012>
30. Hensher D, Rose J, Greene W. The implications on willingness to pay of respondents ignoring specific attributes. *Transportation*. 2005; 32(3):203–222. <https://doi.org/10.1007/s11116-004-7613-8>
31. Puckett S, Hensher D. The role of attribute processing strategies in estimating the preferences of road freight stakeholders. *Transportation Research Part E*. 2008; 44(3):379–395. <https://doi.org/10.1016/j.tre.2007.01.002>
32. Hess S, Hensher D. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B*. 2010; 44(6):781–790. <https://doi.org/10.1016/j.trb.2009.12.001>
33. Train K. *Discrete Choice Methods with Simulation*. 2nd ed. Cambridge University Press; 2009.

34. Ding C, Wu X, Yu G, Wang Y. A gradient boosting logit model to investigate driver's stop-or-run behavior at signalized intersections using high-resolution traffic data. *Transportation Research Part C*. 2016; 72:225–238. <https://doi.org/10.1016/j.trc.2016.09.016>
35. Swait J, Louviere JJ. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*. 1993; 30(3):305–314. <https://doi.org/10.2307/3172883>
36. Jiang H, Chen S. Dwelling unit choice in a condominium complex: analysis of willingness to pay and preference heterogeneity. *Urban Studies*. 2015; Published online.
37. Greene W. *Econometric analysis*. International ed. Essex, England: Pearson Education Ltd.; 2012.
38. Ma X, Wu YJ, Wang Y, Chen F, Liu J. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C*. 2013; 36:1–12. <https://doi.org/10.1016/j.trc.2013.07.010>
39. Ma X, Tao Z, Wang Y, Yu H, Wang Y. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C*. 2015; 54:187–197. <https://doi.org/10.1016/j.trc.2015.03.014>
40. Xinhua News Agency. Press release from the State Council executive meeting on September 28; 2015. [http://news.xinhuanet.com/fortune/2015-09/29/c\\_1116716502.htm](http://news.xinhuanet.com/fortune/2015-09/29/c_1116716502.htm).
41. Shaheen SA, Cohen AP. Growth in worldwide carsharing: an international comparison. *Transportation Research Record*. 2007; 1992(1):81–89. <https://doi.org/10.3141/1992-10>
42. Shaheen SA, Cohen AP, Chung MS. North American carsharing: 10-year retrospective. *Transportation Research Record*. 2009; 2110:35–44. <https://doi.org/10.3141/2110-05>
43. Shaheen S, Martin E. Demand for carsharing systems in Beijing, China: an exploratory study. *International Journal of Sustainable Transportation*. 2010; 4(1):41–55. <https://doi.org/10.1080/15568310802273172>
44. Wang M, Martin E, Shaheen SA. Carsharing in Shanghai, China: analysis of behavioral response to a local survey and potential competition. *Transportation Research Record*. 2012; 2319(1):86–95. <https://doi.org/10.3141/2319-10>
45. Ma X, Liu C, Wen H, Wang Y, Wu YJ. Understanding commuting patterns using transit smart card data. *Journal of Transport Geography*. 2017; 58:135–145. <https://doi.org/10.1016/j.jtrangeo.2016.12.001>
46. Kuhfeld W. *Marketing Research Methods in SAS*. SAS Institute; 2010.
47. Breusch TS, Pagan AR. The Lagrange multiplier test and its application to model specifications in econometrics. *Review of Economic Studies*. 1980; 47:239–253. <https://doi.org/10.2307/2297111>