Journal of Choice Modelling 48 (2023) 100426

Contents lists available at ScienceDirect

Journal of Choice Modelling

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

Reducing sample size requirements by extending discrete choice experiments to indifference elicitation

Ambuj Sriwastava *, Peter Reichert [1]

Eawag: Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

ARTICLE INFO

Keywords:
Decision analysis
Preference elicitation
Discrete choice experiments
Indifference elicitation
Bayesian inference

ABSTRACT

Discrete choice (DC) methods provide a convenient approach for preference elicitation and they lead to unbiased estimates of preference model parameters if the parameterization of the value function allows for a good description of the preferences. On the other hand, indifference elicitation (IE) has been suggested as a direct trade-off estimator for preference elicitation in decision analysis decades ago, but has not found widespread application in statistical analysis frameworks as for discrete choice methods. We develop a hierarchical, probabilistic model for IE that allows us to do Bayesian inference similar to DC methods. A case study with synthetically generated data allows us to investigate potential bias and to estimate parameter uncertainty over a wide range of numbers of replies and elicitation uncertainties for both DC and IE. Through an empirical case study with laboratory-scale choice and indifference experiments, we investigate the feasibility of the approach and the excess time needed for indifference replies. Our results demonstrate (i) the absence of bias of the suggested methodology, (ii) a reduction in the uncertainty of estimated parameters by about a factor of three or a reduction of the required number of replies to achieve a similar accuracy as with DC by about a factor of ten, (iii) the feasibility of the approach, and (iv) a median increase in time needed for indifference reply of about a factor of three. If the set of respondents is small, the higher elicitation effort may be worth to achieve a reasonable accuracy in estimated value function parameters.

1. Introduction

Discrete choice (DC) methods (experiments, models and their statistical evaluation) (Ben-Akiva and Lerman, 1985; Train, 1986; Louviere et al., 2000; Train, 2009; Hensher et al., 2015) provide a convenient approach for preference elicitation in many fields of applied economics, such as transport research, marketing, consumer preference research, health economics, and environmental economics (Ben-Akiva and Lerman, 1985; Train, 1986; Tempesta et al., 2019; Clark et al., 2014; Brouwer, 2008, and many more). They are relatively easy to design and easy to explain to stakeholders as, in the simplest setting, the respondents only have to choose the preferred outcomes from multiple hypothetical outcome sets. On the other hand, individual replies to discrete choice questions do not contain much preference information as a single reply does not resolve the differences in preference between the outcomes. This information is obtained indirectly from the set of replies as the probabilities of the possible replies depend on the difference in preference. The small amount of information extracted from each reply is usually compensated by large sample sizes used in the "experiment" which can often relatively easily be achieved, in particular for the elicitation of societal preferences for which the number of stakeholders is beyond critical limits. However, if the problem to be assessed requires the elicitation of preferences

* Corresponding author.
E-mail addresses: ambuj.sriwastava@eawag.ch (A. Sriwastava), peter.reichert@emeriti.eawag.ch (P. Reichert).
[1] Current status: retired from Eawag, see https://peterreichert.github.io for updated information.

https://doi.org/10.1016/j.jocm.2023.100426
Received 24 December 2021; Received in revised form 23 May 2023; Accepted 31 May 2023
Available online 19 June 2023
1755-5345/

of experts, e.g. in environmental management or health economics, it can be difficult to get a sufficiently large sample size for obtaining reliable preference estimates (de Bekker-Grob et al., 2015).

Many approaches have been explored to reduce sample size requirements in discrete choice experiments. These can be grouped into three categories: (1) The most straightforward technique is to optimize the design of choice sets that are presented to the respondents (Huber and Zwerina, 1996; Bliemer and Rose, 2005; Rose and Bliemer, 2013). However, these techniques depend on prior assumptions about the preference model and its parameters. (2) Other techniques try to gain more information from individual replies by using larger sets of potential outcomes and asking for the best and the worst outcome (Finn and Louviere, 1992; Marley and Louviere, 2005; Marley and Pihlens, 2012; Greiner et al., 2014), or for a complete ranking of the potential outcomes (Marley and Pihlens, 2012). Even other techniques elicit the certainty of the replies (Greiner et al., 2014; Dekker et al., 2016; Mattmann et al., 2019) or the frequencies of choosing one or the other outcome. For these techniques, there is a trade-off between the potential gain of information and the higher elicitation effort for each individual reply. Another method of this category is to ask respondents to modify one of the attributes that characterize the outcomes until indifference between two potential outcomes is achieved (Keeney and Raiffa, 1976; Keeney, 1992; Eisenführ et al., 2010). (3) The third category of approaches are based on adaptive designs by modifying new questions partially based on replies to earlier questions instead of fixing the complete design from the beginning (Myung et al., 2013; Cavagnaro et al., 2013).

It has been shown that using adaptive designs can lead to biased results (Bradley and Daly, 2000) which requires caution with the application of approaches of category (3). For this reason, the typical approach in applied economics is of category (1). The interest in techniques of category (2) is that they can be combined with those of category (1) to further decrease sample size requirements.

It is the goal of this paper to quantitatively explore the potential of an indifference or trade-off technique described as the last option under category (2) in comparison to a discrete choice experiment. To be able to do so, we develop a statistical model for indifference replies that allows us to do statistical inference from elicited data in a similar way as it is done with discrete choice replies. For both techniques, we use the same error model for the value (random utility) function, but for the indifference elicitation technique, we extend the model by Haag et al. (2019) by formulating an additional error model for the indifference point specification by the respondents. This allows us to uniquely describe the uncertainties common to discrete choice and indifference elicitation and consider the additional uncertainty in indifference point specification separately. This indifference point specification error model also resolves the issues related to inconsistent replies for choice situations where the respondent is close to indifference (Bostic et al., 1990), as it allows for inconsistencies caused by the indifference point specification error. We believe that using a continuous error model for the indifference point specification improves uncertainty representation compared to a discrete indifference threshold that has been used before to address this issue (Cantillo et al., 2010; Branke et al., 2017). Considering these two error terms at different levels in the model makes the model a hierarchical, probabilistic model for which doing Bayesian inference is not trivial. We introduce a numerical approach to solving this problem using the efficient Hamiltonian Markov Chain Monte Carlo method for sampling from the posterior distribution. We then test the suggested methodology with a case study with synthetically generated data and with an empirical case study. The synthetic case study allows us to investigate a potential bias of the approach and to quantify the gain in accuracy for a wide range of sample sizes and elicitation uncertainties. The empirical case study makes it possible to investigate the practical feasibility of the approach and to estimate the additional effort needed for indifference elicitation. On the other hand, our empirical case study is not large enough to provide an extensive, empirical comparison of the two techniques. It would be interesting to do this in a real application study with more resources for an extensive experiment.

2. Methods

In this section, we briefly review the value/utility function approach for describing preferences (Section 2.1), its application to evaluate the results of discrete choice and indifference elicitations (Sections 2.2 to 2.4), the design of the synthetic and the empirical case studies (Sections 2.5 and 2.6), and our numerical approach for Bayesian inference (Section 2.7).

2.1. Value functions (utilities)

If preferences of decision makers or stakeholders between potential states of a system or potential outcomes of decision alternatives are complete (respondents can always indicate which of two potential outcomes they prefer or whether they are indifferent between them) and transitive (if they prefer outcome A over outcome B and outcome B over outcome C they also prefer outcome A over outcome C), then their preference can be represented by a value function, v, such that preference of outcome A over outcome B is equivalent to $v(\boldsymbol{a}^{(A)}) > v(\boldsymbol{a}^{(B)})$. Here, $\boldsymbol{a}^{(A)}$ and $\boldsymbol{a}^{(B)}$ are attribute values that characterize the outcomes A and B, respectively, and the value function, v, maps the space of potential outcomes to real numbers. Similarly, $v(\boldsymbol{a}^{(A)}) = v(\boldsymbol{a}^{(B)})$ indicates indifference between the outcomes A and B. All such so-called "ordinal" value functions that are related by monotonically increasing transformations equivalently represent the same preference (see Keeney and Raiffa (1976), Eisenführ et al. (2010) or any other textbook on decision analysis for technical details). If also differences in preference are complete and transitive, preferences can be represented by a "cardinal" or "measurable" value function that is unique up to an affine transformation with positive slope (Keeney and Raiffa, 1976; Dyer and Sarin, 1979; Eisenführ et al., 2010, or any other textbook on decision analysis).

In the nomenclature used in decision analysis, a utility function complements preferences of certain outcomes with risk attitudes (Keeney and Raiffa, 1976; Dyer and Sarin, 1982; Eisenführ et al., 2010) to cope with uncertain outcomes characterized by probability distributions. Note that in economics, the term "utility function" is used for both kinds of preference description (Hensher et al., 2015). The methods discussed in this paper are about the elicitation of value functions, preferences for given outcomes that do not consider risk attitudes. This is motivated by the strategy of eliciting a value function first, and adding risk attitudes for the main objective as a second step (Dyer and Sarin, 1982). For this reason, eliciting a value function without risk attitudes remains a key issue in preference elicitation and, when following this strategy, it is this step that covers the trade-offs between sub-objectives (Reichert et al., 2015).

Although discrete choice replies do not provide direct information about the strength of preference, as derived below, the consideration of uncertainty still makes it possible to derive a measurable or cardinal value function as the probabilities for replies depend on the differences in value between the two potential outcomes.

The typical way of considering uncertainty in preferences is to use an additive, random error term to the value function (called random utility in economics) (Ben-Akiva and Lerman, 1985; Train, 1986; Hensher et al., 2015, or any of the other references about discrete choice methods cited above):

$$V(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\psi}) = v(\mathbf{a}, \boldsymbol{\theta}) + E_v(\mathbf{a}, \boldsymbol{\psi}). \tag{1}$$

Here, we use parameterizations of the value function and the error term with parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, respectively. Usually, it is assumed that the random variable characterizing the error to the value function will not depend on the attributes:

$$E_v(\mathbf{a}, \boldsymbol{\psi}) = E_v(\boldsymbol{\psi}). \tag{2}$$

This is also the assumption used in this paper. As we will see later, it is then worth introducing a new random variable for the difference of two independent error terms of the values of different potential outcomes, A and B:

$$E_v^*(\boldsymbol{\psi}) = E_v(\mathbf{a}^{(B)}, \boldsymbol{\psi}) - E_v(\mathbf{a}^{(A)}, \boldsymbol{\psi}). \tag{3}$$

As we assume that the errors are independent of the attributes of the potential outcomes (see Eq. (2)), we omitted the alternatives on the left-hand side of this equation and will also not specify them any more for the individual error terms.

Note that the invariance of the preferences under affine transformations (with a positive slope) of the value function makes the scaling of the uncertainty term dependent on the chosen scaling of the value function or vice versa. This offers two options for formulating the uncertain value or random utility model (1):

(i) The width of the error term can be fixed and the scaling of the value function determines the degree of uncertainty (relative to the width of the uncertainty term).
(ii) The value function can be made unique by fixing the value for two given outcomes and the degree of uncertainty is characterized by a width parameter of the error term (usually its standard deviation).

All decision-related quantities, such as the sign of the difference in value between two outcomes, or marginal substitution rates are independent of the chosen scaling, so that the two options are equivalent and lead to the same results except for the scaling-related parameters of the value function (the first statement follows from $av(\mathbf{a}_1) + b > av(\mathbf{a}_2) + b \Leftrightarrow v(\mathbf{a}_1) > v(\mathbf{a}_2)$ for $a > 0$ and any values of $\mathbf{a}_1$ and $\mathbf{a}_2$, the second follows from the independence of the indifference manifolds of $v(\mathbf{a})$ and $av(\mathbf{a}) + b$ on the values of a and b as these do not depend on $\mathbf{a}$). In economics, discrete choice experiments are typically evaluated according to option (i): The error term is standardized and the relative significance of the error is inferred indirectly through the scaling of the value (called utility) function. This is a convenient choice as no constraints on the parameters of the value function have to be considered. However, with this option the parameters of the value function change with changing degree of uncertainty. This is not a disadvantage if a single discrete choice experiment is evaluated. However, as we plan to do a sensitivity analysis regarding the degree of uncertainty in our synthetic case study, this would be inconvenient as we could not directly compare marginal posterior parameter estimates under different assumptions about the uncertainty in value. Also for our empirical case study we are interested in comparing the uncertainty in value of different respondents which is much easier if this is a model parameter rather than extracting the information indirectly from different scalings of the value functions of different individuals. For these reasons, we will design and analyze our data based on option (ii): We will make the value function unique by specifying the value to be 0 and 1, respectively, for the two most extreme outcomes within the choice setup. We will still allow the value function to extend beyond the interval [0, 1] so that this choice for our analysis does not decrease the generality of the results (see also the discussion of the chosen value functions and their scaling in Section 2.4).

2.2. Discrete choice model

We briefly review the simplest case of a discrete choice (DC) model among two alternatives to introduce our notation and to prepare for the indifference elicitation model in Section 2.3 (Ben-Akiva and Lerman, 1985; Train, 1986; Hensher et al., 2015, or any other description of discrete choice experiment analysis). Given the error model (1), the probability of choosing outcome A over outcome B is given by

$$\begin{aligned} P(\text{choice} = A \mid \boldsymbol{\theta}, \boldsymbol{\psi}, A, B) &= P\Big(V(\mathbf{a}^{(A)}, \boldsymbol{\theta}, \boldsymbol{\psi}) \geq V(\mathbf{a}^{(B)}, \boldsymbol{\theta}, \boldsymbol{\psi})\Big) \\ &= P\Big(E_v(\mathbf{a}^{(B)}, \boldsymbol{\psi}) - E_v(\mathbf{a}^{(A)}, \boldsymbol{\psi}) \leq v(\mathbf{a}^{(A)}, \boldsymbol{\theta}, \boldsymbol{\psi}) - v(\mathbf{a}^{(B)}, \boldsymbol{\theta}, \boldsymbol{\psi})\Big). \end{aligned} \tag{4}$$

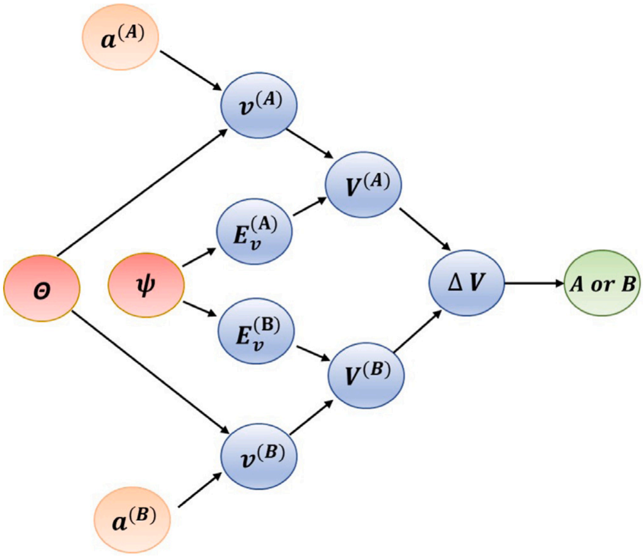


Fig. 1. Schematic diagram of the discrete choice model. The values, v, of both alternatives, A and B, depend on the attributes, $\mathbf{a}$, and the model parameters, $\boldsymbol{\theta}$. Including uncertainty, $\mathbf{E}$, makes them also depend on the error model parameters, $\boldsymbol{\Psi}$. Finally, the decision for A or B depends on the difference in the uncertain values of both outcomes A and B. Orange nodes represent given inputs, red nodes marginal (unconditional) random variables, blue nodes internal (conditional) random variables, and the green node represents the random variable that models the observed outcome. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Under the assumption that the uncertainties in the value functions for the two outcomes are independent of each other and do not depend on the attributes (2), and using the notation introduced in Eq. (3), we get

$$\begin{aligned} P(\text{choice} = A \mid \boldsymbol{\theta}, \boldsymbol{\psi}, A, B) &= P\Big(E_v^*(\boldsymbol{\psi}) \leq v(\mathbf{a}^{(A)}, \boldsymbol{\theta}, \boldsymbol{\psi}) - v(\mathbf{a}^{(B)}, \boldsymbol{\theta}, \boldsymbol{\psi})\Big) \\ &= F_{E_v^*(\boldsymbol{\psi})}\Big(v(\mathbf{a}^{(A)}, \boldsymbol{\theta}, \boldsymbol{\psi}) - v(\mathbf{a}^{(B)}, \boldsymbol{\theta}, \boldsymbol{\psi})\Big), \end{aligned} \tag{5}$$

where $F_{E_v^*(\boldsymbol{\psi})}$ is the cumulative distribution function of the random variable $E_v^*(\boldsymbol{\psi})$. Formulated for an arbitrary choice, c, this leads to

$$P(c \mid \boldsymbol{\theta}, \boldsymbol{\psi}, A, B) = \begin{cases} P(\text{choice} = A \mid \boldsymbol{\theta}, \boldsymbol{\psi}, A, B) & \text{if } c = A \\ 1 - P(\text{choice} = A \mid \boldsymbol{\theta}, \boldsymbol{\psi}, A, B) & \text{if } c = B. \end{cases} \tag{6}$$

Finally, for a set of N discrete choice replies, $\{c^{(i)}\}_{i=1}^N$ corresponding to a set of paired potential outcomes, $\{A^{(i)}, B^{(i)}\}_{i=1}^N$, and under the assumption of independence of the error terms, we get

$$P\big(\{c^{(i)}\}_{i=1}^N \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \{A^{(i)}, B^{(i)}\}_{i=1}^N\big) = \prod_{i=1}^N P\big(c^{(i)} \mid \boldsymbol{\theta}, \boldsymbol{\psi}, A^{(i)}, B^{(i)}\big) \tag{7}$$

and for the joint probability of replies and parameters

$$\begin{aligned} &P\big(\{c^{(i)}\}_{i=1}^N, \boldsymbol{\theta}, \ \boldsymbol{\psi} \mid \{A^{(i)}, B^{(i)}\}_{i=1}^N\big) \\ &\quad = P\big(\{c^{(i)}\}_{i=1}^N \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \{A^{(i)}, B^{(i)}\}_{i=1}^N\big) \cdot f(\boldsymbol{\theta}) \cdot f(\boldsymbol{\psi}) \\ &\quad = \prod_{i=1}^N P\big(c^{(i)} \mid \boldsymbol{\theta}, \boldsymbol{\psi}, A^{(i)}, B^{(i)}\big) \cdot f(\boldsymbol{\theta}) \cdot f(\boldsymbol{\psi}) \end{aligned} \tag{8}$$

with the prior probability densities of the parameters $f(\boldsymbol{\theta})$ and $f(\boldsymbol{\psi})$. This equation is the basis for Bayesian inference for the discrete choice model as the posterior of the parameters is proportional to this joint probability with the actual replies substituted for $\{c^{(i)}\}_{i=1}^N$. Fig. 1 shows a graphical representation of the discrete choice model and the left panel of Fig. 2 illustrates outcomes, value function and the choice decision in a two-dimensional attribute space.

2.3. Indifference/trade-off elicitation

The concept of indifference elicitation (IE), asking the respondent to modify one of the attributes of one of two potential outcomes to get indifference between the outcomes, is very old. It was already described in the textbook by Keeney and Raiffa (1976). IE is an interesting concept as it provides direct information about preferences for trade-offs of the respondent between values of attributes. Nevertheless, IE has not found widespread application within a statistical framework that accounts for uncertainty as it is common

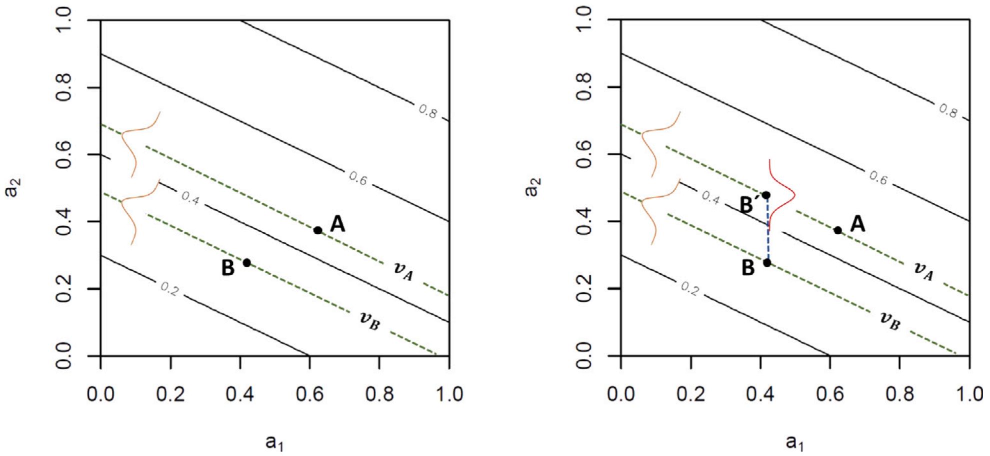


Fig. 2. Illustration of a discrete choice decision (left) and of indifference elicitation (right) for a two-dimensional attribute space. Solid and dashed lines combine points of equal value, the distributions at the left illustrate their uncertainty. For discrete choice, the respondent just specifies preference between the two alternatives, here with high probability in favor of outcome A (as there is only a small overlap of the uncertain values of A and B). For indifference elicitation, the respondent changes the value of a_2 from the value at point B to the value at point B' to be indifferent between A and B'. The distributions at point B' indicates the uncertainty in the specification of this value.

practice for the evaluation of DC replies. In this section, we derive a probabilistic model for indifference replies that accounts for the uncertainty in value (analogous to the DC case) as well as for the uncertainty in the indifference point specification and allows us to do Bayesian (or Frequentist) inference about value function parameters based on elicited indifference replies.

In the following, we focus on binary choice and assume that the attribute to be modified is continuous or, at least, has a high number of discrete states that allow a sufficiently accurate specification of the indifference point. We also assume that its range is large enough that the indifference point can be reached. Note that this makes it impossible to apply this technique to cases in which a change in one attribute cannot be compensated by a change in another attribute (e.g. if persons are asked for commuting preferences and have the choice between using their car or traveling by bus and they prefer one or the other irrespective of travel time, there would be no indifference point regarding travel time for two alternatives that differ in the transportation means). The function that returns the indifference point in the attribute a_j, $\tilde{a}_j$, can be defined as the solution of the implicit equation of equal value of the two potential outcomes, A and B:

$$\tilde{a}_j\left(\boldsymbol{\theta}, \epsilon_v^{(A)}, \epsilon_v^{(B)}, A, B\right) = \mathrm{sol}_x\left(v(\mathbf{a}^{(A)}, \boldsymbol{\theta}) + \epsilon_v^{(A)} = v\left((\mathbf{a}_{-j}^{(B)}, x), \boldsymbol{\theta}\right) + \epsilon_v^{(B)}\right). \tag{9}$$

Here, $\epsilon_v^{(A)}$ and $\epsilon_v^{(B)}$ are realizations of the random variables $E_v^{(A)}$ and $E_v^{(B)}$, sol_x refers to the solution of the equation provided as the argument for the variable x and $(\mathbf{a}_{-j}^{(B)}, x)$ is the vector of attributes $\mathbf{a}^{(B)}$ with the component j substituted by x:

$$(\mathbf{a}_{-j}^{(B)}, x) = (a_1^{(B)}, \dots, a_{j-1}^{(B)}, x, a_{j+1}^{(B)}, \dots, a_n^{(B)}), \tag{10}$$

where n is the number of attributes. If we again assume that the uncertainties in the value functions for the two outcomes are independent of each other and do not depend on the attributes (2), and using the notation introduced in Eq. (3), we get

$$\tilde{a}_j\left(\boldsymbol{\theta}, \epsilon_v^*, A, B\right) = \mathrm{sol}_x\left(\epsilon_v^* = v(\mathbf{a}^{(A)}, \boldsymbol{\theta}) - v\left((\mathbf{a}_{-j}^{(B)}, x), \boldsymbol{\theta}\right)\right), \tag{11}$$

where $\epsilon_v^* = \epsilon_v^{(B)} - \epsilon_v^{(A)}$ is a realization of the random variable defined by Eq. (3). As the specification of this "indifference point" involves additional uncertainty to the uncertainty in value, we assume an additive uncertainty term similar and in addition to the uncertainty in the value function. This leads to the random variable A_j for the response a_j conditional on the error ϵ_v^*

$$A_j \mid \epsilon_v^* = \tilde{a}_j(\boldsymbol{\theta}, \epsilon_v^*, A, B) + E_a(\boldsymbol{\psi}). \tag{12}$$

We denote the probability density of this random variable as follows

$$f(a_j \mid \tilde{a}_j, \boldsymbol{\psi}). \tag{13}$$

This leads to the model

$$f(a_j, \epsilon_v^* \mid \boldsymbol{\theta}, \boldsymbol{\psi}, A, B) = f\left(a_j \mid \tilde{a}_j\left(\boldsymbol{\theta}, \epsilon_v^*, A, B\right), \boldsymbol{\psi}\right) \cdot f(\epsilon_v^* \mid \boldsymbol{\psi}). \tag{14}$$

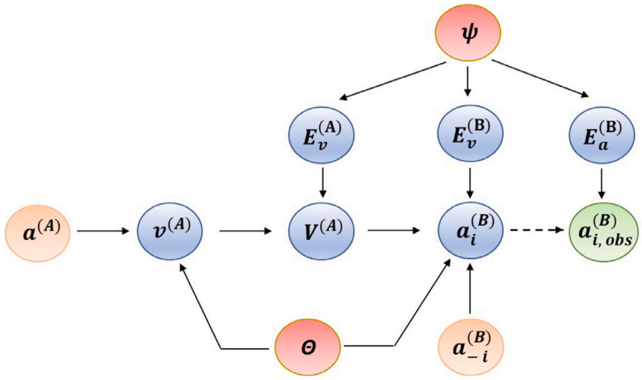


Fig. 3. Schematic diagram of the indifference reply model. The value of the outcome A. $v^{(A)}$, depends on the attributes, $\mathbf{a}^{(A)}$ and the model parameters $\boldsymbol{\theta}$, the value with uncertainty additionally depends on the error term, $\mathbf{E}_v^{(A)}$, and its parameters, $\boldsymbol{\Psi}$. The indifference point, $a_i^{(B)}$ depends on the other attribute components of the outcome B, $\mathbf{a}_{-i}^{(B)}$, on the uncertain value for the outcome A, $V^{(A)}$, and on the error term $\mathbf{E}_v^{(B)}$, and its parameters, $\boldsymbol{\Psi}$, through the implicit Eq. (11). Finally, the elicited indifference point depends on the exact point and its uncertainty, $\mathbf{E}_a^{(B)}$. See Fig. 1 for an explanation of the colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For a set of N replies, $\{a_{j(i)}^{(i)}\}_{i=1}^{N}$, corresponding to the set of potential outcomes $\{A^{(i)}, B^{(i)}\}_{i=1}^{N}$, we then get

$$f\big(\{a_{j(i)}^{(i)}\}_{i=1}^{N}, \{\epsilon_v^{*(i)}\}_{i=1}^{N} \mid \boldsymbol{\theta}, \boldsymbol{\psi}, \{A^{(i)}, B^{(i)}\}_{i=1}^{N}\big) = \prod_{i=1}^{N} f(a_j \mid \tilde{a}_{j(i)}\big(\boldsymbol{\theta}, \epsilon_v^{*(i)}, A^{(i)}, B^{(i)}\big), \boldsymbol{\psi}) \cdot f(\epsilon_v^{*(i)} \mid \boldsymbol{\psi}). \tag{15}$$

This finally leads to the joint probability density

$$f\big(\{a_{j(i)}^{(i)}\}_{i=1}^{N}, \{\epsilon_v^{*(i)}\}_{i=1}^{N}, \boldsymbol{\theta}, \boldsymbol{\psi} \mid \{A^{(i)}, B^{(i)}\}_{i=1}^{N}\big) = \prod_{i=1}^{N} f(a_j \mid \tilde{a}_{j(i)}\big(\boldsymbol{\theta}, \epsilon_v^{*(i)}, A^{(i)}, B^{(i)}\big), \boldsymbol{\psi}) \cdot f(\epsilon_v^{*(i)} \mid \boldsymbol{\psi}) \cdot f(\boldsymbol{\theta}) \cdot f(\boldsymbol{\psi}). \tag{16}$$

Fig. 3 shows a graphical representation of the indifference reply model and the right panel of Fig. 2 illustrates the process of indifference elicitation. Rather than choosing the better alternative, the interviewee is asked to specify the value of the attribute $a_j^{(B)}$ for which he or she would be indifferent between the two alternatives.

Note that the uncertainties in value differences, $\{\epsilon_v^{*(i)}\}_{i=1}^{N}$, lead to a large number of “internal nodes” of a hierarchical model that increase the dimension of the inference problem compared to the non-hierarchical model presented by Haag et al. (2019) that did not explicitly consider these uncertainties but lumped them with the uncertainties in value. As we will usually be interested primarily in the marginal posterior for the “global” parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$, we may want to integrate across these internal variables:

$$f\ \big(\{a_{j(i)}^{(i)}\}_{i=1}^{N}, \boldsymbol{\theta}, \boldsymbol{\psi} \mid \{A^{(i)}, B^{(i)}\}_{i=1}^{N}\big) = \prod_{i=1}^{N} \int f(a_j \mid \tilde{a}_{j(i)}\big(\boldsymbol{\theta}, \epsilon_v^{*(i)}, A^{(i)}, B^{(i)}\big), \boldsymbol{\psi}) \cdot f(\epsilon_v^{*(i)} \mid \boldsymbol{\psi})\, \mathrm{d}\epsilon_v^{*(i)} \cdot f(\boldsymbol{\theta}) \cdot f(\boldsymbol{\psi}). \tag{17}$$

2.4. Distributional assumptions, attribute scaling, and value functions

The theory outlined in Sections 2.2 and 2.3 is independent of specific assumptions about distributional shapes. In this section, we demonstrate how to get the frequently used probit and logit models as special cases. In the applied part, we will use the probit model. This model assumes a Normal distribution for the error term in the value function

$$\mathbf{E}_v(\boldsymbol{\psi}) \sim \mathrm{N}(0, \sigma_v), \tag{18}$$

which leads to

$$\mathbf{E}_v^{*}(\boldsymbol{\psi}) \sim \mathrm{N}(0, \sqrt{2}\sigma_v) \tag{19}$$

for differences in value (see Eq. (3)). Similarly, we can get the logit model, by replacing the normal distribution in Eq. (18) by a Gumbel distribution that we scale to a standard deviation of σ_v to make it comparable to the probit model:

$$\mathbf{E}_v(\boldsymbol{\psi}) \sim \mathrm{Gumbel}\left(0, \frac{\sqrt{6}}{\pi}\,\sigma_v\right). \tag{20}$$

Note that in contrast to the probit approach, the mean of the error term (20) is not zero. This does not affect our analysis, as we only need the differences of two error terms where this cancels out (Train, 2009). We then get

$$\mathbf{E}_{v}^{*}(\boldsymbol{\psi}) \sim \text{Logistic}\left(0, \frac{\sqrt{6}}{\pi}\,\sigma_{v}\right), \tag{21}$$

which is, due to our scaling and as in Eq. (19), a distribution with a standard deviation of $\sqrt{2}\sigma_v$.

We assume a Normal distribution for the error in the indifference replies

$$\mathbf{E}_{a_i}(\boldsymbol{\psi}) \sim \mathrm{N}(0, \sigma_{a_i}). \tag{22}$$

For our synthetic case study we use the same value of σ_a for the standard deviations of the uncertainties in all attributes, σ_{a_i}. This is possible as we generate the data synthetically. In the empirical case study, we have to infer different values of σ_{a_i} for different attributes as they have different units and there would be no reason for identical uncertainties even if there would be multiple attributes with the same units.

With the choice of distributional shapes and parameters given by Eqs. (18) and (22), we thus get the error model parameters $\boldsymbol{\psi} = \sigma_v$ for the discrete choice model and $\boldsymbol{\psi} = (\sigma_v, \sigma_a)$ or $\boldsymbol{\psi} = (\sigma_v, \sigma_{a_1}, \ldots, \sigma_{a_n})$ for the indifference reply model, depending on the need for different uncertainties in different attributes (see discussion above). However, for the latter model, due to its hierarchical structure, we also have the "internal parameters" $\{\mathbf{E}_{v,i}\}_{i=1}^{N}$ which we will usually integrate out, but the posterior distribution of which could also be of interest for the analysis of structural errors of the value function.

Both approaches, discrete choice or indifference elicitation, can be applied with any kind of parameterized value function. In decision analysis, a value function is often constructed by eliciting an objectives hierarchy and parameterizing the value function of the overarching objective by constructing value functions of the lowest level objectives and aggregating them to value functions of higher level objectives (Eisenführ et al., 2010; Reichert et al., 2015; Reichert, 2020). In economics, value functions are usually constructed similarly to constructing statistical models in other contexts by formulating a linear model and adding non-linear and interaction terms as needed (Ben-Akiva and Lerman, 1985; Hensher et al., 2015). Both approaches end with a parameterized value function and both techniques analyzed in this study can be directly applied to the overall value function as well as to value functions of sub-objectives.

To not make our analyses unnecessarily complicated we assume value functions that are monotonic in all attributes (either increasing or decreasing with increasing attribute values). We can then separate the shape of the value function from the specific units and ranges by introducing the transformation

$$\tilde{a}_i = t_i(a_i) = \frac{a_i - a_{i,\mathrm{minv}}}{a_{i,\mathrm{maxv}} - a_{i,\mathrm{minv}}}\ ,\ \mathbf{t} = (t_1, \ldots, t_n)\ ,\ \tilde{\mathbf{a}} = \mathbf{t}(\mathbf{a}). \tag{23}$$

In this equation, $a_{i,\mathrm{minv}}$ and $a_{i,\mathrm{maxv}}$, are the components of the attributes corresponding to the minimum and maximum value in the design outcome space used for the DC or IE inquiry. As we assumed monotonicity of value in all attributes, this transformation transforms the design outcome space to the set $[0,1]^n$. Attribute values outside the design space are still allowed and lead to transformed attributes outside the interval $[0,1]$. This transformation allows us to separate shape and scaling of the value function by formulating the value function as

$$v(\mathbf{a}) = \tilde{v}\big(\tilde{\mathbf{a}} = \mathbf{t}(\mathbf{a})\big), \tag{24}$$

where $\tilde{v}$ maps the set $[0,1]^n$ to the interval $[0,1]$ with increasing values in each attribute (note that we scaled the value function to the interval $[0,1]$ for attributes in the design set, see option (ii) in Section 2.1). We can thus define the shape of the value function independently of the attribute ranges by defining the function $\tilde{v}$.

The most straightforward value function is a linear function of its attributes. With the constraints discussed above (range of transformed attributes [0, 1], range of values [0, 1], increasing value with increasing attributes), this leads to the weighted average of the transformed attributes:

$$\tilde{v}_{\mathrm{add}}(\tilde{\mathbf{a}}, \boldsymbol{\theta}) = \sum_{i=1}^{n} w_i \tilde{a}_i\ ,\ \boldsymbol{\theta} = (w_1, \ldots, w_n)\ ,\ w_i \geq 0\ ,\ \sum_{i=1}^{n} w_i = 1. \tag{25}$$

Substituting the transformation (24) into this equation leads to the value function as a function of the original attributes

$$v_{\mathrm{add}}(\mathbf{a}, \boldsymbol{\theta}) = \sum_{i=1}^{n} w_i t_i(a_i) = \sum_{i=1}^{n} w_i \frac{a_i - a_{i,\mathrm{minv}}}{a_{i,\mathrm{maxv}} - a_{i,\mathrm{minv}}} = \sum_{i=1}^{n} c_i a_i + c \tag{26a}$$

with

$$c_i = \frac{w_i}{a_{i,\mathrm{maxv}} - a_{i,\mathrm{minv}}} \quad \text{and} \quad c = -\sum_{i=1}^{n} \frac{w_i a_{i,\mathrm{minv}}}{a_{i,\mathrm{maxv}} - a_{i,\mathrm{minv}}}. \tag{26b}$$

The non-dimensional parameters w_i can be useful to compare the relative importance of attributes given their ranges, whereas the parameters c_i are needed to calculate marginal rates of substitution $(-c_i/c_j)$ that have to be expressed in the original units.

As mentioned in Section 2.1, we add an uncertainty term to the value function with a standard deviation σ_v. Note that the equivalent formulation used in economics would be to use a standardized uncertainty term and scale the value function relative to the standardized uncertainty. Both of these descriptions are equivalent and connected as follows (we have to divide the uncertain value by σ_v to standardize the uncertainty term and we ignore the constant offset in Eq. (26)):

$$v_{\mathrm{add,alt}}(\mathbf{a},\boldsymbol{\theta}) = \sum_{i=1}^{n} \beta_i a_i = \sum_{i=1}^{n} \frac{c_i}{\sigma_v}\ a_i\ ,\ \boldsymbol{\theta} = (\beta_1,\dots,\beta_n)\ ,\ \beta_i = \frac{c_i}{\sigma_v}. \tag{27}$$

Note that we will use the formulation ((25), (26)) for our synthetic experiment because it allows us to parameterize the uncertainty explicitly with the parameter σ_v and thus allows us to analyze the sensitivity to the magnitude of uncertainty while keeping the value function parameters constant. Nevertheless, the two approaches are equivalent and choosing the formulation (27) instead would also have been fine, just a bit less convenient for the discussion of the sensitivity to the magnitude of uncertainty in value. Note also that, for this simple linear value function model, marginal rates of substitution are given by (negative) ratios of the coefficients which are identical ($-c_i/c_j = -\beta_i/\beta_j$) as the standard deviation of the error term used in Eq. (27) to convert the parameters from one approach to the other, cancels out.

In the following, we will discuss alternative shapes of value functions in the form of the function $\tilde{v}$. In many cases, a linear value function (25) or (27) is not sufficient to satisfyingly describe preferences. In economics, to test for nonlinearity, often interaction terms in the form $\beta_{i,j}^{\mathrm{int}} a_i a_j$ and/or nonlinear terms, e.g. $\beta_i^{(2)} a_i^2$, are added to the linear function (27) to test for nonlinearity. In decision analysis, often other nonlinear relationships are used (Langhans et al., 2014; Haag et al., 2019; Reichert et al., 2019). A simple example of such a nonlinearity is the weighted average between additive and minimum functions:

$$\begin{aligned}\tilde{v}_{\mathrm{addmin}}(\tilde{\mathbf{a}},\boldsymbol{\theta}) &= (1-\alpha)\sum_{i=1}^{n} w_i \tilde{a}_i + \alpha \min(\{\tilde{a}_i\}_{i=1}^{n}) \\ &\qquad \boldsymbol{\theta} = (w_1,\dots,w_n,\alpha)\ ,\ w_i \ge 0\ ,\ \sum_{i=1}^{n} w_i = 1\ ,\ \alpha \in [0,1].\end{aligned} \tag{28}$$

We can also get the reverse "curvature" of the isolines or iso-surfaces by using the maximum

$$\begin{aligned}\tilde{v}_{\mathrm{addmax}}(\tilde{\mathbf{a}},\boldsymbol{\theta}) &= (1-\alpha)\sum_{i=1}^{n} w_i \tilde{a}_i + \alpha \max(\{\tilde{a}_i\}_{i=1}^{n}) \\ &\qquad \boldsymbol{\theta} = (w_1,\dots,w_n,\alpha)\ ,\ w_i \ge 0\ ,\ \sum_{i=1}^{n} w_i = 1\ ,\ \alpha \in [0,1].\end{aligned} \tag{29}$$

In our study, we combine these two value functions by using negative values of α to describe the mixture with the maximum with a weight $|\ \alpha\ |$:

$$\begin{aligned}\tilde{v}_{\mathrm{addminmax}}(\tilde{\mathbf{a}},\boldsymbol{\theta}) &= \begin{cases} (1-\alpha)\sum_{i=1}^{n} w_i \tilde{a}_i + \alpha \min(\{\tilde{a}_i\}_{i=1}^{n}) & \text{for } \alpha \ge 0 \\ (1+\alpha)\sum_{i=1}^{n} w_i \tilde{a}_i - \alpha \max(\{\tilde{a}_i\}_{i=1}^{n}) & \text{for } \alpha < 0 \end{cases} \\ &\qquad \boldsymbol{\theta} = (w_1,\dots,w_n,\alpha)\ ,\ w_i \ge 0\ ,\ \sum_{i=1}^{n} w_i = 1\ ,\ \alpha \in [-1,1].\end{aligned} \tag{30}$$

This value function has the advantage that the inference process can decide whether a mixture of the additive model with the minimum model or the maximum model is more appropriate as the parameter α allows for a continuous transition between the two approaches. Figure SI.1 in the Supporting Information provides an overview of typical shapes of this value function in two dimensions. Note that for $\alpha = 0$ we get the additive (linear) model.

As we are eliciting a measurable value function (see Section 2.1), we should provide the possibility of a nonlinear transformation to account for varying distances of the indifference lines or indifference sub-manifolds. We do this by allowing for the following transformation:

$$\begin{aligned}\tilde{v}_{\mathrm{trans}}(\tilde{\mathbf{a}},\boldsymbol{\theta}) &= \begin{cases} \dfrac{\exp\big(\beta\ \tilde{v}_{\mathrm{addminmax}}(\tilde{\mathbf{a}},\boldsymbol{\theta})\big)-1}{\exp(\beta)-1} & \text{for } \beta \ne 0 \\ \tilde{v}_{\mathrm{addminmax}}(\tilde{\mathbf{a}},\boldsymbol{\theta}) & \text{for } \beta = 0 \end{cases} \\ &\qquad \boldsymbol{\theta} = (w_1,\dots,w_n,\alpha,\beta)\ ,\ w_i \ge 0\ ,\ \sum_{i=1}^{n} w_i = 1\ ,\ \alpha \in [-1,1]\ ,\ \beta \in \mathbb{R}.\end{aligned} \tag{31}$$

This value function is illustrated in Figure SI.2 in the Supporting Information for $\alpha = 0$ and different values of β.

2.5. Design of the synthetic case study

The goal of the synthetic case study is to test the suggested methodology for potential bias and to explore the gain in uncertainty reduction compared to discrete choice for a wide range of numbers of replies and elicitation uncertainties. Both can only be done with synthetically generated data. The test for bias can only be done if the true values are known and testing for many different numbers of replies and uncertainties would be practically infeasible in an empirical case study.

Table 1
Parameter default values, sensitivity ranges, and priors for the synthetic experiment. The chosen Dirichlet distribution corresponds to a uniform distribution of the weights on a standard simplex, the Lognormal distribution is parameterized with its true mean and standard deviation, NormalTrunc is a Normal distribution with given mean and standard deviation truncated to the interval given as the third argument. The parameters are defined in sections 2.4 and N is the number of replies. See also the overview of notation in the appendix.

Parameter	Default value(s)	Sensitivity range	Prior
$\mathbf{w}$	$\left(\frac{1}{3}, \frac{2}{3}\right)$		Dirichlet (1, 1)
	$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$		Dirichlet (1, 1, 1)
	$\left(\frac{1}{6}, \frac{1}{6}, \frac{1}{3}, \frac{1}{3}\right)$		Dirichlet (1, 1, 1, 1)
σ_v	0.05	[0.05, 0.4]	Lognormal (0.05, 0.025)
σ_a	0.05	[0.05, 0.4]	Lognormal (0.05, 0.025)
α	−0.5, 0, 0.5		NormalTrunc (0, 0.5, [−1, 1])
β	−1, 0, 1, 2		Normal (0, 1)
N		[16, 4225]	

Table 2
Range of attributes considered for decision problems A and B.

	(A) Buying a car			(B) Renting a flat		
	Price (CHF)	FuelCons (L/100 km)	NOxEmiss (%)	Rent (CHF/month)	Travel time (min.)	EnergyCons (kWh/m^2/y)
Minimum	19 500	3	10	1350	0	30
Maximum	31 500	9	100	2550	60	150

We choose the simplest analysis layout for the comparison of the two methods with not distinguishing different respondents and thus generating data with global value function parameters and we will consequently also infer global parameter values. The case study or the application of the technique could easily be extended to consider different respondent characteristics, but the simplest case seems to be sufficient to analyze the accuracy of inferred parameters for a given number of replies or the required number of replies to achieve a certain accuracy. However, we will test linear and nonlinear value functions as outlined in Section 2.4. For the synthetic case study, we chose the attribute ranges to be $[0, 1]$ and thus the transformation (24) corresponds to the identity. The selected default parameter values and sensitivity ranges for generating synthetic data and the priors for re-inferring the parameters are given in Table 1.

Uncertainty analysis and sensitivity analysis regarding the number of replies was done with data generation using the default parameter values and re-inferring the parameters. Sensitivity analyses regarding σ_v and σ_a were done by generating data with varying these parameters while keeping the other parameters at their default values. All priors were chosen to be wide to have a minor effect on the inference results. The priors were always kept the same also when parameters values were varied for sensitivity analysis to avoid a prejudice by the known value.

All synthetic cases were based on a full factorial design of the discrete choice and indifference questions. We did not use a D-efficient design (Federov, 1972; Bliemer and Rose, 2010; Huber and Zwerina, 1996) for our analyses to be independent of prior structural assumption of the value function (Walker et al., 2018).

2.6. Design of the empirical case study

For the empirical case study we designed two simple decision problems for laboratory-scale choice or indifference experiments: (A) buying a car, and (B) renting a flat. In both cases, the alternatives were characterized by three attributes assuming other aspects to be similar. For the decision problem A the attributes are price of the car (in Swiss francs CHF), Fuel consumption (in liters per 100 km), and NOx emission (expressed as a percentage of the regulatory limit). For the problem B the attributes are monthly rent of a 1 bedroom apartment in Zurich (in CHF per month), time required for traveling to the work place (in minutes), and the energy consumption for heating the apartment (in kWh per square meter living area per year). The attribute ranges for both case study problems are given in Table 2.

We collected preference data for both problems A and B from 6 respondents. For each choice situation, the respondents identified their preferred alternative and indicated their indifference reply as a value of a predefined attribute. To analyze the relative difficulty in answering an indifference question compared to a discrete choice question, we also measure the time taken by the respondents to answer both type of questions.

After scaling the attribute ranges to the interval [0, 1], we used the same priors for the parameters $\mathbf{w}$, σ_v and σ_{a_i} as were used in the synthetic case study given in Table 1. However, we distinguished different uncertainties for different attributes as there is no reason that the uncertainty relative to the interval given in Table 2 should be the same.

Similar to the synthetic case study, we applied the full factorial design to generate the choice sets. For the decision problem A, we divided the *Price* attribute into 4 levels and the other two attributes into 3 levels each. Similarly for the problem B, we divided the *Rent* attribute into 4 levels and the remaining two attributes into 3 levels each. Consequently for both the problems, we generate choice sets comprising $4 \times 3 \times 3 = 36$ discrete choice and indifference reply questions. Each individual replied to all of these 36 decision situations. The number of replies of 36 was chosen as a compromise between the achievable accuracy of inferred parameters and a potential deterioration of the results by fatigue of the respondents. The first parts of two questionnaires are provided as examples in section SI.4 of the supporting information. All complete questionnaires are provided in our institutional repository https://doi.org/10.25678/0007TW. As different individuals have different preferences regarding buying a car or renting a flat, we did not merge the elicited replies of the 6 respondents to a single sample but just present them as 6 examples of individual preferences.

2.7. Numerical algorithms and implementation

The discrete choice model leads to an analytical expression for the joint probability of replies and parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ given by Eq. (8). The posterior probability density of the model parameters is proportional to this expression if the actual observations are substituted into this function. This leads to a low-dimensional inference problem that can efficiently be solved numerically e.g. with any Markov Chain Monte Carlo (MCMC) sampler. On the other hand, due to the hierarchical nature of the indifference specification model, we can only derive a (relatively) simple expression for the joint probability density of replies, value function error differences, $\{\epsilon_v^{*(i)}\}_{i=1}^N$, and parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ (Eq. (16)). Due to the presence of the value function differences in this expression that cannot easily be integrated out, this leads to a much higher-dimensional inference problem (this adds unknowns of the dimension of the number of replies, N). Samples from the marginal distributions of the parameters according to Eq. (17) can easily be obtained from a sample of the joint posterior of errors and parameters by ignoring the information from the error terms. To account for this high dimension of the inference problem, we decided to use the Hamiltonian Monte Carlo (HMC) approach to sample from the posterior distribution of the parameters and internal nodes (Neal, 2011). More specifically, the algorithmic parameters were chosen by using the No-U-Turn Sampler (NUTS) (Hoffman and Gelman, 2014) as implemented in the Stan software, https://mc-stan.org (Carpenter et al., 1985).

Data generation, analysis and post-processing was done in R, https://r-project.org (R Core Team, 2020). The full factorial designs were generated using the R package `AlgDesign`, https://cran.r-project.org/package=AlgDesign. Bayesian inference was implemented by accessing the Stan software using the R package `rstan`, https://cran.r-project.org/package=rstan (Stan Development Team, 2020).

Examples of our code are provided in section SI.3 of the Supporting Information, the complete code can be found in our institutional repository at https://doi.org/10.25678/0007TW.

3. Results and discussion

3.1. Synthetic case study

3.1.1. Computational effort

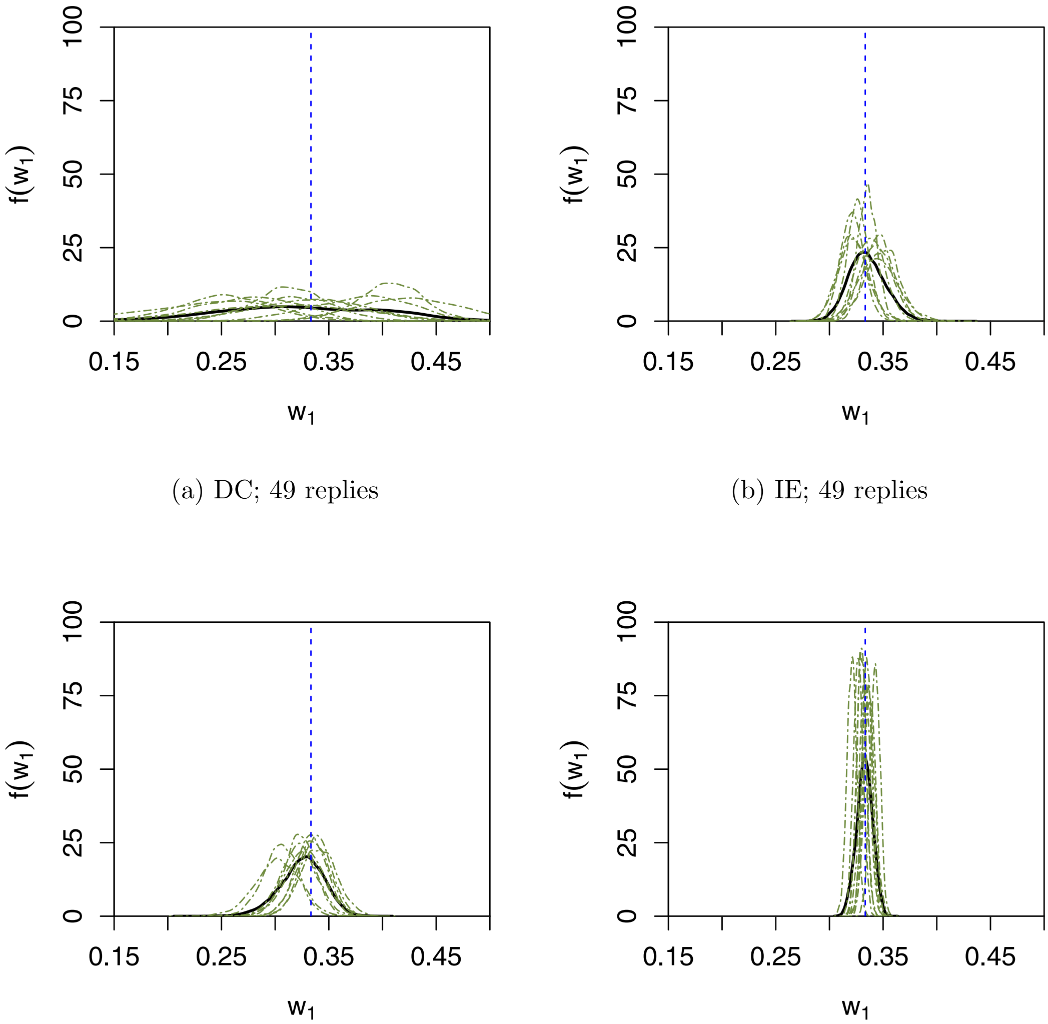
As outlined in Section 2.7, we have a much higher-dimensional inference problem for indifference elicitation (IE) than for discrete choice (DC). This can be coped with by using the very efficient Hamiltonian Monte Carlo sampler (see Section 2.7), but it leads to longer computation times for IE compared to DC. As a representative example, inference (a single repetition) using DC for the additive function with 2 attributes took about 18 s for 100 replies and about 42 s for 400 replies. The corresponding inference times for IE were 66 s and 429 s respectively. The inference was done using 4 Markov chains with the length of each chain being 1000. These simulation times are for a computer equipped with Intel(R) Core(TM) i7-7500U CPU processor with 12 GB RAM and 4 cores. Despite the large relative increase in computation time for IE compared to DC, which is even increasing for larger numbers of replies, these times are not at all limiting the evaluation of IE data. As the inference problem remained feasible for even much larger numbers of replies, the computational demand is not a crucial factor for the decision about which method to apply.

3.1.2. Dependence of inference results on the number of replies

Fig. 4 shows the posterior marginals of the parameter w_1 of the additive value function with two attributes estimated using DC and IE. The experiment and parameter inference was repeated 10 times to test the method for potential bias. The green dashed curves display each repetition and, as evident from the figure, these individual estimates differ from each other reflecting variability inherent to finite sample sizes. The solid black curve represents the parameter estimate combining samples generated in all the 10 repetitions. The true value of the parameter is shown as the blue dotted line. Combining the 10 repetitions results in parameter estimates with little or no bias while the individual repetitions demonstrate the variability that can be expected from a single experiment with the indicated number of replies. Figure SI.4 in the supporting information complements Fig. 4 with the same results for the parameter σ_v.

Fig. 5 compares marginal posterior characteristics of the parameter w_1 of the additive value function with two, three and four attributes for discrete choice and indifference elicitation data. We obtain unbiased parameter estimates with decreasing uncertainty for increasing numbers of replies. On the other hand, the marginal posteriors for IE are about three times less uncertain than those for DC with only minor dependence on the number of replies.

SI.2.1.1 in the Supporting Information expands the results shown above to the parameters w_1, α, β, σ_v and σ_a of the additive (Figs. SI.5 to SI.7), the additive-minimum–maximum (Figs. SI.8 to SI.11) and the transformed value functions (Figs. SI.12 to SI.15). The ratios of the posterior standard deviations between discrete choice and indifference estimation (“accuracy ratios”) are mostly in the range between 2 and 4 for all tested value functions and numbers of replies. They are somewhat smaller for the standard



(a) DC; 49 replies (b) IE; 49 replies

(c) DC; 400 replies (d) IE; 400 replies

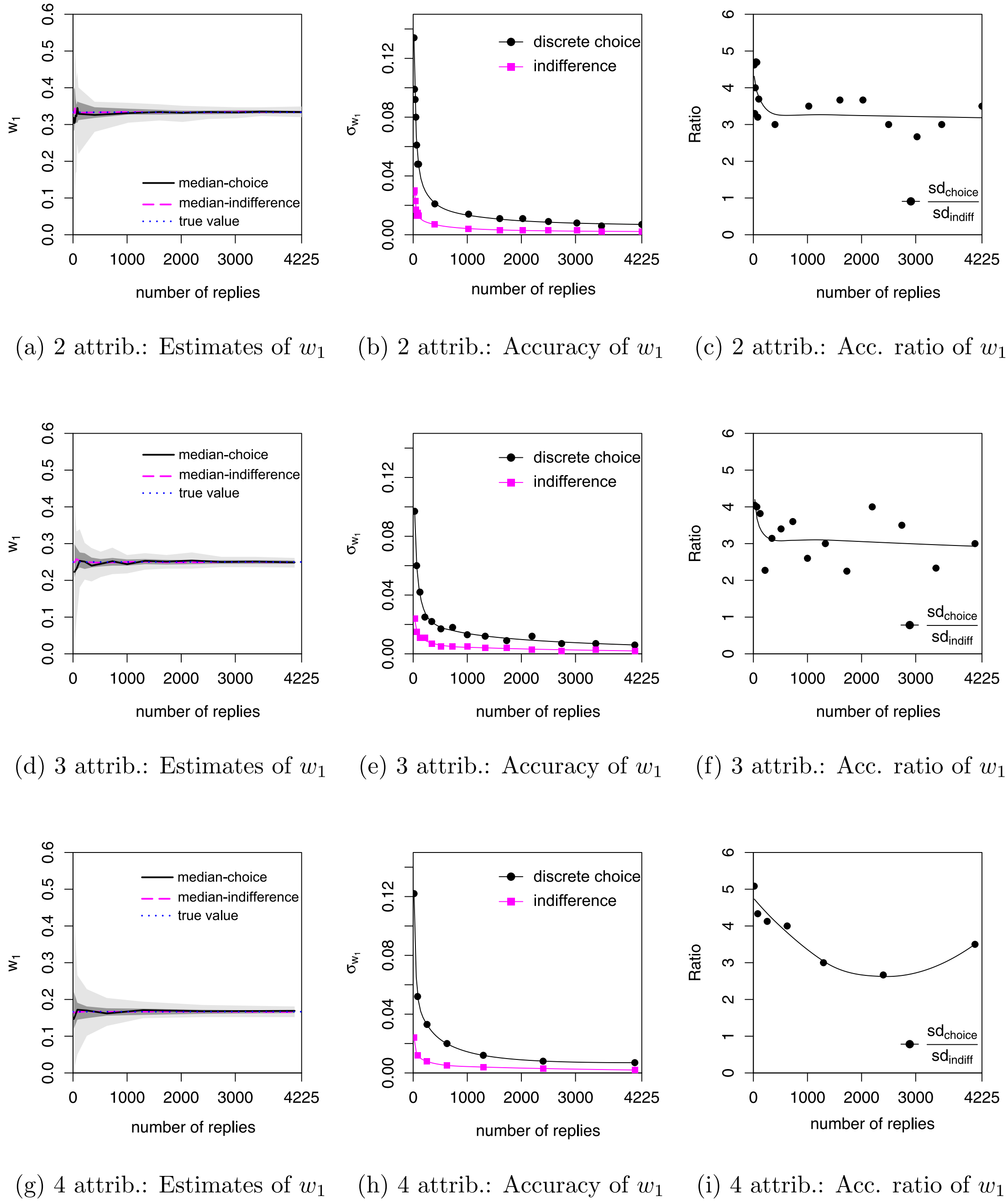
Fig. 4. Estimates of the posterior marginals of the parameter w_1 of the additive value function with two attributes for two different numbers of replies using Discrete Choice (a and c) and Indifference Elicitation (b and d). The solid black line shows the parameter estimate when combining 10 repetitions of the experiment while the green dashed lines represent the posterior marginals of individual repetitions. The blue dotted line represents the true value of the parameter used to generate the synthetic data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

deviation σ_v and the parameter β of the transformed-additive value function than for the weight parameter w_1 and the nonlinearity parameter α. This factor of 3 in posterior uncertainty reduction indicates that indifference elicitation leads to a similar posterior uncertainty with a ten times smaller number of replies (and still with a four times smaller number of replies for the parameters that are only reduced by a factor of two).

3.1.3. Sensitivity to uncertain preference and indifference point specification

This section presents results for varying degrees of uncertainty in preference modeled by the uncertainty in value quantified by σ_v (see Eqs. (1) and (18)) and of uncertainty in the specification of the indifference point quantified by σ_a (see Eqs. (12) and (22)). All analyses in this section are done with value functions with two attributes.

Sensitivity to σ_v: Fig. 6 shows how the accuracy in the estimation of parameter w_1 changes when the value of σ_v is increased from 0.05 to 0.40 for 49, 400 and 4225 replies. For the additive value function, the increase in the standard deviation of parameter estimates is considerably higher for DC when the random error to the value function σ_v is increased. Note that a σ_v value of 0.4 indicates a very large uncertainty as the values are scaled to the interval [0, 1] for attributes within the design ranges. But even for such a high uncertainty, IE performs significantly better than DC.



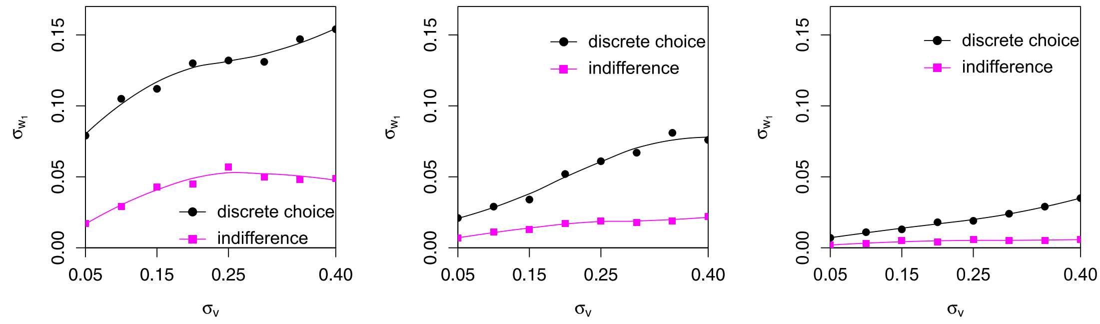
(a) 2 attrib.: Estimates of w_1 (b) 2 attrib.: Accuracy of w_1 (c) 2 attrib.: Acc. ratio of w_1

(d) 3 attrib.: Estimates of w_1 (e) 3 attrib.: Accuracy of w_1 (f) 3 attrib.: Acc. ratio of w_1

(g) 4 attrib.: Estimates of w_1 (h) 4 attrib.: Accuracy of w_1 (i) 4 attrib.: Acc. ratio of w_1

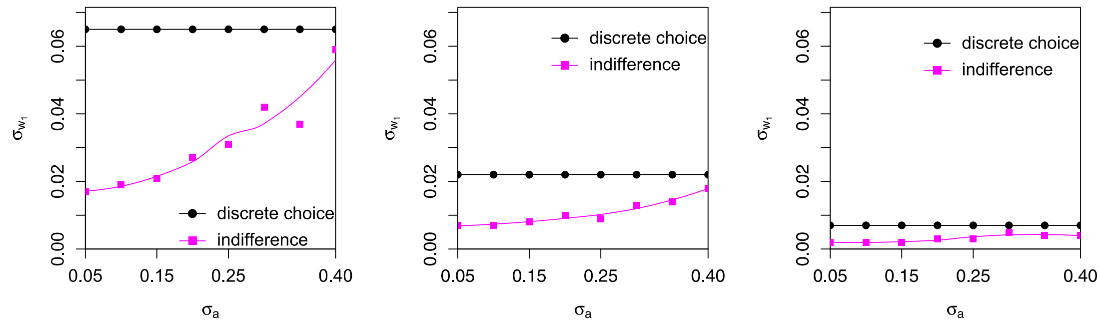
Fig. 5. Comparison of marginal posteriors of the parameter w_1 of the additive value function with two attributes (a), (b), (c); three attributes (d), (e), (f); and four attributes (g), (h), (i) for DC and IE as functions of the number of replies. In each row, the left panel shows the posterior median and 95% credibility interval for DC (light gray) and for IE (dark gray), the middle panel shows the standard deviations of the marginal posteriors and the right panel shows the ratio of standard deviations of DC to IE. In the middle and right columns, the markers represent values obtained for the corresponding number of replies, the lines are smoothed relationships to support interpretation.

Sensitivity to σ_a: For three different numbers of replies, 49, 400 and 4225, sensitivity to σ_a was analyzed by inferring the model parameters for 8 different values of σ_a: 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35 and 0.40 (note that the attribute ranges for the design layouts were scaled to the interval [0, 1]). Since the indifference point specification error, which is characterized by the parameter σ_a, is only present for indifference elicitation, discrete choice experiment results are not sensitive to σ_a. As expected and shown in Fig. 7, the posterior standard deviation of w_1, σ_{w_1}, is increasing with increasing values of σ_a as we gain less information with increasing elicitation uncertainty. Nevertheless, even for the very large value of $\sigma_a = 0.4$ it is still smaller than for DC. For more



(a) 49 replies (b) 400 replies (c) 4225 replies

Fig. 6. Sensitivity to σ_v: posterior standard deviation of w_1 for the additive value function as a function of the value of σ_v used to generate the synthetic data. Markers and lines have the same meaning as in Fig. 5.



(a) 49 replies (b) 400 replies (c) 4225 replies

Fig. 7. Sensitivity to σ_a: posterior standard deviation of w_1 for the additive value function as a function of the value of σ_a used to generate the synthetic data (note that σ_a is the uncertainty in specifying the indifference point and is thus is only needed for IE).

realistic elicitation uncertainties of σ_a of up to 0.15 (an approximate 95% credibility interval would cover more than half of the attribute range) we maintain the approximate gain in accuracy of about a factor of three.

3.1.4. *Sensitivity to model structure errors*

We test the performance of discrete choice and indifference elicitation for situations in which the assumption is wrong that the parameterized family of value functions contains the value function that represents the decision maker's preference perfectly. Our test consists of two cases: inferring parameters of a linear (additive; Eq. (25)) value function when the preference data was generated by using a non-linear (additive-minimum–maximum; Eq. (30)) value function and vice-versa.

Additive value function inferred for data generated with an additive-minimum–maximum value function: We compare DC and IE for inferring the parameters of the additive function with preference data generated with the additive-minimum function. Fig. 8 shows the results for the parameter w_1 of the additive value function. As this figure shows, we cannot recover the true value of w_1, which is a natural result when inferring parameters with a wrong model. It is an interesting result that the ratio of the standard deviations of DC vs IE is smaller than for any of the inference results shown in the preceding sections. This is an indication that IE is more sensitive, or also more discriminatory, regarding the shape of the value function.

Additive-minimum–maximum value function inferred for data generated with an additive value function: Figure SI.16 in the Supporting Information shows the results for the parameters w_1 and α of the additive-minimum–maximum function. The results confirm the consistency of both approaches, DC and IE: Since the additive function is a special case of the additive-minimum–maximum function when α equals zero, both approaches recover this value of zero and we re-gain the uncertainty ratio of about a factor of three for the posterior standard deviations.

Given the difficulty of checking the validity of the model, it may be the best strategy to analyze the data with different, nested model structures to identify the need for a higher model complexity. Besides using information criteria, such as AIC or BIC, as it is often done in economics, it is then possible to check whether the coefficients are significantly different from their value for the simpler model structure (in most cases, this means significantly different from zero). Our analyses demonstrated that this strategy works also for indifference elicitation for the value functions used in this study. We may speculate that the ratio of the

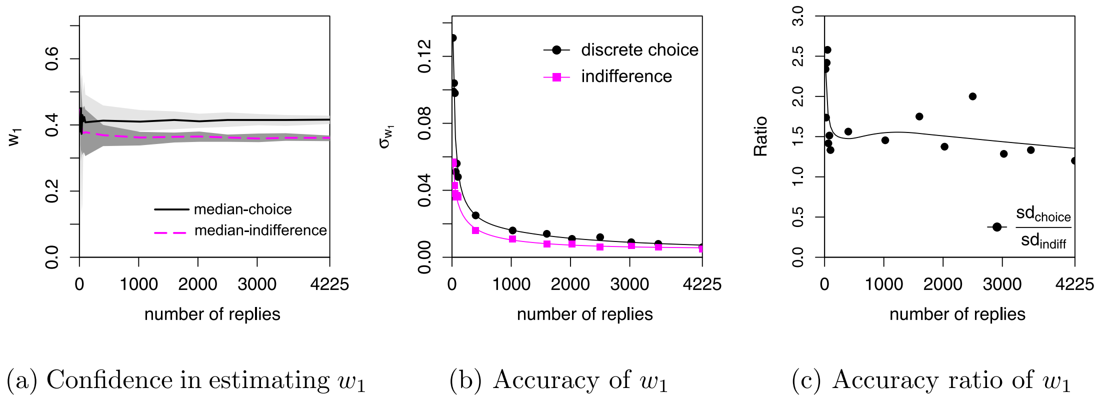


Fig. 8. Comparison of DC and IE for the parameter w_1 of the additive value function when inferred using data generated with the additive-minimum value function.

standard deviations between DC and IE evaluations of the same data may even be an indicator of structural model deficits. However, confirming this speculation would need more investigation.

3.2. Empirical case study

For the decision problems A and B described in Section 2.6 we collected preference data from six respondents R1, …, R6. For each choice question comprising two alternatives, the respondents were first asked to indicate their preferred alternative. Subsequently the respondents were asked to adjust the (randomly selected) marked attribute to the value so that they are indifferent between the two presented alternatives (see section SI.4 in the supporting information for examples from the questionnaires or our institutional repository for the full versions https://doi.org/10.25678/0007TW). This data was used to compare DC and IE by inferring the parameters of linear value functions. The results of parameter inference are presented in Section 3.2.1. In addition to preference data we also recorded the time required for answering discrete choice and indifference reply questions. The time taken for answering the questions are used to analyze the mental difficulty of processing the provided information and choosing an alternative or providing an estimate of the indifference point. The timer was started when a question was presented to the respondent and once the respondent made the choice of the preferred alternative the timer was stopped. This time was recorded as t_{DC}. The timer was re-started from 0 and once the respondent provides the indifference reply the timer was stopped again. This time was recorded as t_{IE}. The time t_{IE} alone would be an underestimation of the time the respondent takes to answer an indifference reply because the respondent became already familiar with the decision problem while choosing an alternative and he or she anyway needs to implicitly make a choice of the preferred alternative to specify the indifference point. Hence, the total time $t_{\mathrm{DC}} + t_{\mathrm{IE}}$ was used as a conservative estimate of the time required for answering the indifference reply question. Results analyzing the difficulty of DC and IE questions relative to each other are presented in Section 3.2.2.

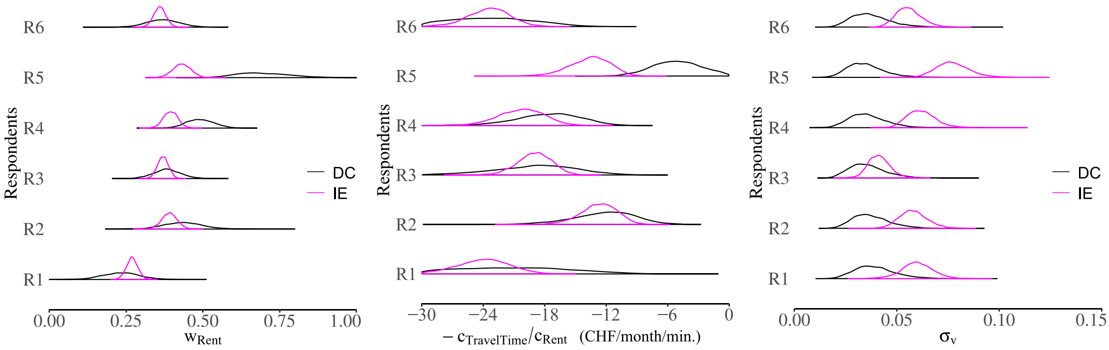
3.2.1. DC vs IE: Parameter inference

Fig. 9 shows the comparison of posterior marginals of the weight parameter w_{Rent}, of the marginal substitution rate $-c_{\mathrm{Traveltime}}/c_{\mathrm{Rent}}$, and for the standard deviation of the uncertainty in value, σ_v, inferred from the preference data collected from the respondents for the decision problem B, renting a flat, described in Section 2.6. The results for the remaining parameters and for the decision problem A, buying a car, are included in section SI.2.2.1 in the Supplementary Information.

The weight parameter indicates the relative importance of rent compared to the other two attributes of travel time to the work place and energy consumption for heating for the given ranges of these attributes. It is very important to note that this relative importance is not universal but that it depends on the attribute range; it decreases with decreasing length of the attribute range. As the parameters of the respondents represent individual preferences, we do not expect them to be equal across respondents. However, we expect the results of DC and IE to be consistent for each individual respondent. Considering the large uncertainty of DC, this seems to be the case at least for all respondents except respondent R5. However, having in mind the large variability in inferred posteriors for small numbers of replies (see dashed lines in Fig. 4(a)), this could still be a random result. It would be interesting to do such a comparison with a larger number of replies to investigate whether there are systematic differences between the results of the different elicitation techniques.

The marginal substitution rate has the direct behavioral interpretation by how much the respondents are willing to increase their rent to save a minute of travel time. Again, we see indications for potential inconsistent replies only for respondent R5.

The results for the standard deviation of the uncertainty in value, σ_v, indicate a larger uncertainty for IE compared to DC. This could result from a larger sensitivity of IE to the shape of the value function that may not be ideally covered by the linear model. A comparison of the ratios of the marginal posterior standard deviations shown in Fig. 10 with the larger ones for the synthetic case study based on the correct model (Fig. 5) and the smaller ones for inference with an incorrect model (Fig. 8) would be a further



(a) Weight parameter (b) Marginal subst. rate (c) Uncertainty

Fig. 9. Posterior marginals inferred for w_{Rent}, for the marginal substitution rate $-c_{TravelTime}/c_{Rent}$, and for σ_v of the additive value function for the six respondents for the decision problem B: Renting a flat based on the replies for DC and IE.

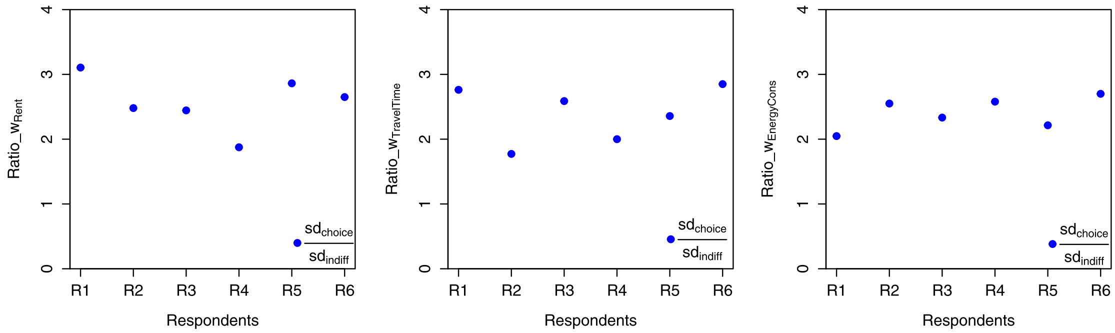


Fig. 10. Ratios of standard deviations of posterior marginals of the weight parameters for DC and IE for the decision problem B: Renting a flat.

indication in support of this hypothesis (see also discussion in Section 3.1.4). However, the nonlinear additive-minimum–maximum model did also not lead to much better results. Our number of replies is too small to clearly identify a potentially needed nonlinear value function.

3.2.2. *Difficulty of choice questions*

Fig. 11 shows the ratio of $t_{\mathrm{DC}}+t_{\mathrm{IE}}$ to t_{DC} for the decision problem B: Renting a flat (see Fig. SI.25 in the Supporting Information for the results for both decision problems). Here t_{DC} is the time needed to chose the preferred alternative and t_{IE} is the time taken by the respondent to answer the indifference reply after he or she has already made a choice of the preferred alternative. Hence we use $t_{DC}+t_{IE}$ as a conservative estimate of time taken for the indifference reply. Overall, indifference replies took approximately 2–4 times more time than discrete choice replies. This supports the hypothesis that indifference replies generally take longer time than replies to discrete choice questions. However, the results shown in Fig. 11 also provide a quantitative measure of the degree to how much more difficult indifference reply questions are in comparison to discrete choice questions. For both the empirical case studies, the accuracy ratio was found to be between 2–3 i.e. around 4–9 times potential reduction in the required number of replies across different respondents. Based on the collected preference data from our empirical case study, it can be concluded that the higher elicitation effort for IE is well compensated by the higher achieved accuracy.

Additional analysis was done by identifying questions for which it is particularly easy for the respondents to choose the preferred alternative. If all attribute comparisons between the two presented alternatives individually lead to favoring the same alternative, it is particularly easy to select the preferred alternative, as no trade-off across the attribute dimensions is needed. We call these kind of questions "easy" for DC and all the other ones that require such trade-offs "difficult". We expect that the response times for DC would be shorter for such "easy" questions compared to "difficult" questions, but that there is no similar effect for IE, as trade-offs are always needed for IE. Our results, presented in section SI.2.2.2 of the Supporting Information confirm this expectation. These kind of "easy" questions are a particular example of why we gain more information from IE than for DC as these DC replies do not provide any information about the shape of the value function for value functions that are monotonic in all attributes once it is known for all attributes whether an increase or decrease in the attribute leads to an increase in the value function. In such cases, these questions could be eliminated from the design to increase the efficiency of DC.

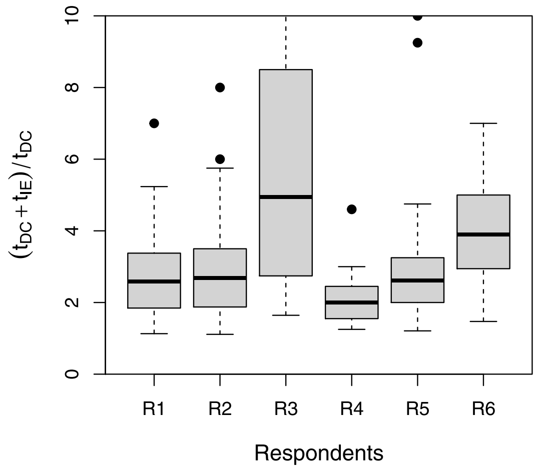


Fig. 11. Ratio of total time $t_{DC}+t_{IE}$ to t_{DC} as a representative measure of relative difficulty of an IE question to a DC question for the decision problem B: Renting a flat.

4. Conclusions

We investigated the extension of discrete choice inquiries to the specification of the indifference point between two presented outcomes by varying one attribute of one of the outcomes. This indifference elicitation concept has been described in textbooks already many decades ago (Keeney and Raiffa, 1976) but has not found widespread application in studies that are evaluated within rigorous statistical frameworks. We developed a statistical model for indifference elicitation that allows us to use this concept within a similar statistical framework as it is common practice for discrete choice inquiries. We used the same description of uncertainty in the value function as it is done with discrete choice methods. However, as we need additional uncertainty in the specification of the indifference point, the model becomes hierarchical and thus more demanding for inferring model parameters from elicited data. To cope with this problem, we implemented Bayesian inference with a Hamiltonian Markov Chain Monte Carlo approach for sampling from the posterior. This limits the computational burden to an amount not relevant as a criterion for the selection of the elicitation methodology. We tested the suggested approach with a case study with synthetically generated data to check for potential bias and to quantify the gain in uncertainty reduction when using the same number of replies or in reducing the number of replies for achieving a similar uncertainty as when applying the discrete choice approach. In addition, we did a small empirical case study to learn about the feasibility of the approach and the increase in elicitation effort for each question.

Our results demonstrate (i) the absence of bias of the suggested methodology, (ii) a reduction in the uncertainty of estimated parameters by about a factor of three or the reduction of the required number of replies to achieve the same accuracy as with discrete choice by about a factor of ten (exact values depend on the value function, the number of replies, and value function and indifference point estimation uncertainties, but those orders of magnitude are quite stable within the investigated ranges of numbers of replies and uncertainty; see extensive sensitivity analyses in the results section and in the Supporting Information), (iii) the feasibility of the approach, and (iv) a median increase of time needed to reply to the indifference question of about a factor of three. If the set of respondents is small, e.g. in the case of preference elicitation from experts in environmental management or health economics, or if it is intended to be small to get prior information for designing an effective sampling strategy for a larger scale study, the higher elicitation effort may be manageable and results in a similar accuracy of results with about one-tenths of the number of replies compared to discrete choice replies or higher accuracy for smaller numbers of replies reductions.

Key limitations of the indifference elicitation approach are that it requires at least one continuous attribute, the range of which is large enough so that the indifference point to the other presented alternative can be reached and the willingness to reply to indifference questions by the respondents.

To gain more experience with the suggested approach, a larger scale study with the application of discrete choice and indifference elicitation would be very useful. Such a study could also make it possible to explore potential bias introduced by the difficulties of human judgment regarding indifference point specification.

CRediT authorship contribution statement

Ambuj Sriwastava: Writing – review & editing, Methodology, Software, Formal analysis, Validation, Visualization. **Peter Reichert:** Writing – original draft, Writing – review & editing, Supervision, Funding acquisition, Conceptualization, Methodology, Software.

Declaration of competing interest

The authors declare that there is no conflict of interest involved in the submission of the manuscript titled "Reducing Sample Size Requirements by Extending Discrete Choice Experiments to Indifference Elicitation".

Acknowledgments

We thank the Swiss National Science Foundation for funding AS (grant 200021_192122), and Nele Schuwirth, Judit Lienert, Ivana Logar, Marco Baity Jesi and Andreas Scheidegger for stimulating discussions and comments to draft versions of this paper.

Appendix A. Notation

$\mathbf{a}$: attributes that characterize a system regarding the properties relevant for a decision.
$\mathbf{a}^{(A)}$: attributes corresponding to the outcome A.
$\mathbf{E}_v(\mathbf{a}, \boldsymbol{\psi})$: additive uncertainty contribution to the value function.
f_R: probability density function (pdf) of the continuous random variable R. Note that R is sometimes omitted if clear from the context.
F_R: cumulative distribution function (cdf) of the random variable R. Note that R is sometimes omitted if clear from the context.
n: number of attributes (dimension of the attribute space).
N: number of replies to choice or indifference questions. (For the statistical evaluation, this would be called sample size, however, in discrete choice analysis sample size denotes the number of respondents. For the joint analysis of respondents the number of replies is the sample size times the number of replies by each respondent. For individual evaluation of respondents (as it is done in our empirical case study) the number of replies is equal to the number of replies by each respondent.)
P_R: probability distribution function of the discrete random variable R. Note that R is sometimes omitted if clear from the context.
$v(\mathbf{a})$: value function representing preferences of a decision maker or stakeholder.
$v(\mathbf{a}, \boldsymbol{\theta})$: parameterized value function; parameters will be estimated from replies of the decision maker or stakeholder to match his or her preferences as closely as possible.
$V(\mathbf{a}, \boldsymbol{\theta}, \boldsymbol{\psi})$: uncertain value function ($= v(\mathbf{a}, \boldsymbol{\theta}) + \mathbf{E}_v(\mathbf{a}, \boldsymbol{\psi})$).
$\mathbf{w}$: weight parameters of all value functions.
α: parameter of the additive-minimum–maximum value function (30).
β: parameter of the transformed value function (31).
$\boldsymbol{\theta}$: parameters of non-specific value functions.
$\boldsymbol{\psi}$: parameters of non-specific uncertainty models of the value function and the attribute specification.
σ_a: standard deviation of the chosen normally distributed error model of the attribute specification.
σ_v: standard deviation of the chosen normally distributed error model of the value function.

Appendix B. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jocm.2023.100426. The complete code and results can be found in our institutional repository at https://doi.org/10.25678/0007TW.

References

Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge, Massachusetts, USA.
Bliemer, M.C.J., Rose, J.M., 2005. Efficiency and Sample Size Requirements for Stated Choice Studies. Working paper itls-wp-05-08, Institute of Transport and Logistic Studies, The University of Sydney, Sydney, Australia.
Bliemer, M.C.J., Rose, J.M., 2010. Construction of experimental designs for mixed logit models allowing for correlation across choice observations. Transp. Res. B 44, 720–734.
Bostic, R., Herrnstein, R.J., Luce, R.D., 1990. The effect on the preference-reversal phenomenon of using choice indifferences. J. Econ. Behav. Organ. 13, 193–212.
Bradley, M., Daly, A., 2000. New analysis issues in stated preference research. In: Ortuzar, J.d.D. (Ed.), Stated Preference Modelling Techniques: A Compilation of Major Papers Selected from PTRC's Vast Bank of Meeting and Conference Material. In: PTRC Perspectives v. 4, PTRC Education and Research Services Ltd., London.
Branke, J., Corrente, S., Greco, S., Gutjahr, W., 2017. Efficient pairwise preference elicitation allowing for indifference. Comput. Oper. Res. 88, 175–186.
Brouwer, R., 2008. The potential role of stated preference methods in the water framework directive to assess disproportionate costs. J. Environ. Plan. Manag. 51, 597–614.
Cantillo, V., Amaya, J., Ortúzar, J.d.D., 2010. Thresholds and indifference in stated choice surveys. Transp. Res. B 44, 753–763.
Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M.A., Guo, J., Li, P., Riddell, A., 1985. Stan: A probabilistic programming language. J. Stat. Softw. 76 (1), 1–32.
Cavagnaro, D.R., Gonzales, R., Myung, J.I., Pitt, M.A., 2013. Parameter-free elicitation of utility and probability weighting functions. Manage. Sci. 59 (2), 358–375.

Clark, M.D., Determann, D., Stravos, P., Moro, D., de Bekker-Grob, E.W., 2014. Discrete choice experiments in health economics: a review of the literature. PharmacoEconomics 32, 883–902.
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. Patient 8, 373–384.
Dekker, T., Hess, S., Brouwer, R., Hofkes, M., 2016. Decision uncertainty in multi-attribute stated preference studies. Resour. Energy Econ. 43, 57–73.
Dyer, J.S., Sarin, R.K., 1979. Measurable value functions. Oper. Res. 27 (4), 810–822.
Dyer, J.S., Sarin, R.K., 1982. Relative risk aversion. Manage. Sci. 28 (8), 875–886.
Eisenführ, F., Weber, M., Langer, T., 2010. Rational Decision Making. Springer, Berlin.
Federov, V., 1972. Theory of Optimal Experiments. Academic Press, New York.
Finn, A., Louviere, J.J., 1992. Determining the appropriate response to evidence of public concern: the case of food safety. J. Public Policy Mark. 11, 12–25.
Greiner, R., Bliemer, M., Ballweg, J., 2014. Design considerations of a choice experiment to estimate likely participation by north Australian pastoralists in contractual biodiversity conservation. J. Choice Model. 10, 34–45.
Haag, F., Lienert, J., Schuwirth, N., Reichert, P., 2019. Identifying non-additive multi-attribute value functions based on uncertain indifference statements. Omega 95, 49–67.
Hensher, D.A., Rose, J.M., Greene, W.H., 2015. Applied Choice Analysis, second ed. Cambridge University Press, Cambridge, UK.
Hoffman, M.D., Gelman, A., 2014. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. J. Mach. Learn. Res. 15, 1351–1381.
Huber, J., Zwerina, K., 1996. The importance of utility balance and efficient choice designs. J. Mar. Res. 33 (3), 307–317.
Keeney, R.L., 1992. Value-Focused Thinking. Harvard University Press, Cambridge.
Keeney, R.L., Raiffa, H., 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Wiley.
Langhans, S.D., Reichert, P., Schuwirth, N., 2014. The method matters: A guide for indicator aggregation in ecological assessments. Ecol. Indic. 45, 494–507.
Louviere, J.J., Hensher, D.A., Swait, J.D., 2000. Stated Choice Methods. Cambridge University Press, Cambridge, UK.
Marley, A.A.J., Louviere, J.J., 2005. Some probabilistic models of best, worst, and best–worst choices. J. Math. Psych. 49, 464–480.
Marley, A.A.J., Pihlens, D., 2012. Models of best–worst choice and ranking among multiattribute options (profiles). J. Math. Psych. 56, 24–34.
Mattmann, M., Logar, I., Brouwer, R., 2019. Choice certainty, consistency, and monotonicity in discrete choice experiments. J. Environ. Econ. Policy 8 (2), 109–127.
Myung, J.I., Cavagnaro, D.R., Pitt, M.A., 2013. A tutorial on adaptive design optimization. J. Math. Psych. 57, 53–67.
Neal, R.M., 2011. MCMC using Hamiltonian dynamics. In: Brooks, S., Gelman, A., Jones, G.L., Meng, X.-L. (Eds.), Handbook of Markov Chain Monte Carlo. CRC Press, Boca Raton, pp. 113–162.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Reichert, P., 2020. Towards a comprehensive uncertainty assessment in environmental research and decision support. Water Sci. Technol. 81 (8), 1588–1596.
Reichert, P., Langhans, S.D., Lienert, J., Schuwirth, N., 2015. The conceptual foundation of environmental decision support. J. Environ. Manag. 154, 316–332.
Reichert, P., Niederberger, K., Rey, P., Helg, U., Haertel-Borer, S., 2019. The need for unconventional value aggregation techniques: experiences from eliciting stakeholder preferences in environmental management. EURO J. Decis. Process 7, 197–219.
Rose, J.M., Bliemer, M.C.J., 2013. Sample size requirements for stated choice experiments. Transportation 40, 1021–1041.
Stan Development Team, 2020. RStan: the R interface to Stan. R package version 2.21.2.
Tempesta, T., Vecchiato, D., Nassivera, F., Bugatti, M., Torquati, B., 2019. Consumers demand for social farming products: An analysis with discrete choice experiments. Sustainability 11 (23), 6742.
Train, K., 1986. Quantitative Choice Analysis - Theory, Econometrics, and an Application To Automobile Demand. The MIT Press, Cambridge, Massachusetts, USAA.
Train, K.E., 2009. Discrete Choice Methods with Simulation, second ed. Cambridge University Press, New York.
Walker, J.L., Wang, Y., Thorhauge, M., Ben-Akiva, M., 2018. D-efficient or deficient? A robustness analysis of stated choice experimental designs. Theory and Decision 84, 215–238.