

Is ‘referencing’ a remedy to hypothetical bias in value of time elicitation? Evidence from economic experiments

Lars Hultkrantz¹  · Selen Savsin¹

Published online: 11 August 2017

© The Author(s) 2017. This article is an open access publication

Abstract This paper demonstrates that commonly used methods for eliciting value of time can give downward bias and investigates whether this can be reversed by ‘referencing’ as has been suggested (e.g., by Hensher in *Transp Res B* 44:735–752, 2010), i.e. with attributes of choice alternatives pivoted around a recently made journey. Value-of-time choice experiments were conducted in two rounds. In the first round, real and hypothetical purchases of performance of a simple time-consuming task were done to assess hypothetical bias; in the second round, participants were asked to do hypothetical travel choices with and without ‘referencing’ to a specific occasion, to be able to test ‘referencing’ as a remedy to the bias confirmed in the first round. A negative hypothetical bias was found for allocation of time at another occasion than the present (but not for a decision concern allocation of time ‘here and now’). A striking result was held from the second round experiments: ‘referencing’ indeed affects responses, but by reducing the elicited implicit value of time, so any negative hypothetical bias that would exist without ‘referencing’ would have been further magnified by the ‘referencing’ design.

Keywords Value of time · Stated choice · Certainty calibration · Referencing

Introduction

In valuation of travel time, studies surprisingly have found evidence of negative hypothetical bias (Brownstone and Small 2005; Isacsson 2007), i.e., the value of time revealed by real choice is higher than the value that is estimated from a hypothetical choice survey. Hensher (2010) has suggested that a remedy to such bias is to make “choice experiments

✉ Lars Hultkrantz
lars.hultkrantz@oru.se

Selen Savsin
selen.savsin@oru.se

¹ Örebro University School of Business, Örebro University, 701 82 Örebro, Sweden

with ‘referencing’ back to a real market activity”, i.e., to ask respondents whether they would have preferred some hypothetical alternative in an already experienced real choice situation. In this study we conduct a series of simple time-use choice experiments that on the one hand confirm presence of a negative hypothetical bias, on the other hand indicate that ‘referencing’ results in even lower estimates of the value of time.

Small (2012) refers to the value of travel time as “incompletely understood and ripe for further theoretical and empirical investigation” (pp. 2–3). One important approach for eliciting the value of time is based on stated preferences of individuals. Yet, in several studies stated preference (SP) methods are found to generate a lower value of travel time than what is found by revealed preference (RP) approaches (see reviews by Shires and deJong 2009; Abrantes and Wardman 2011). Both methods are regularly used in value of travel-time elicitation but since RP studies rely on non-experimental and often incomplete observational data, SP values drawing on data from choice experiments are often preferred. However, these SP experiments are framed in a hypothetical choice context and are therefore susceptible to various biases that can arise in such a setting, one of them being hypothetical bias.

The prevalence of a positive bias that leads to overestimation of the actual (real) value of public and private goods in hypothetical valuation surveys is well documented (Cummins et al. 1995; Harrison 2006a, b; Murphy et al. 2005; Harrison and Rutström 2008; Svensson 2010; Loomis 2011). It is therefore surprising that for travel time SP estimates seem to be lower than RP values. As pointed out by Carlsson (2010), some factors suggested by Levitt and List (2007) for explaining behavioural differences between laboratory experiments and real world can be relevant in comparing surveys and actual behaviour. Carlsson (2010) listed them as scrutiny, context, stakes, selection of participants and restrictions on the time horizon and on the choice set. More specific to our context; participants making hypothetical choices between different goods or bundles of attributes of one good, tend to ignore that in a real choice expenditure on other items requires more detailed consideration. Therefore the participants might allocate too much of a given budget to the goods or attributes that are in focus in the hypothetical choice experiment. Then, in a similar manner, participants making hypothetical choices concerning travel time may disregard alternative time use (i.e., underestimate the opportunity cost of the time constraint) leading to a difference between real and hypothetical travel time choices. As a remedy, Hensher (2010) has recently advocated the use of a ‘referencing’ approach implemented in the SP approach used in value of time studies. In this approach, the SP choice alternatives are framed as hypothetical variations of a reference journey that the respondent recently has conducted. The idea is to make the choice scenarios more relevant by linking them to a situation in which the respondent is aware of all considerations that could affect the choice; such as specific time-use alternatives. Given that ignorance of such specification in a standard SP study leads to a lower implicit value of time than what is revealed in a real choice setting, Hensher’s suggestion implies that the implicit value of time is higher with ‘referencing’ than without.

In this study we have tested that hypothesis. We proceeded in two steps. *First*, in a round of choice experiments with 358 participants we investigated whether the acceptance rates to value of time bids, and thus the implicit value of time, differ between a specific time-allocation situation (‘here and now’) and a hypothetical choice of time use in a less precise future situation (‘later’). The choice experiments were conducted with university students that at the end of the last lecture of the day were individually asked whether they would be willing to do a paid simple but time-consuming task. Based on a random split, some participants were asked whether they would undertake the task ‘here and now’, while

others were asked whether they were willing to undertake the same task at a similar occasion and place but at another, not precisely defined, day ('later'). Moreover, the participants that were asked whether they were willing to undertake the task 'here and now', were randomly assigned as either a real or a hypothetical choice situation. Our results confirmed that the elicited value of time is lower in the hypothetical choice of time use in a less precise future situation ('later'). *Second*, in another round of choice experiments with 300 new participants we investigated whether responses to a simple hypothetical travel choice SP survey are affected by whether or not reference is made to a recent travel occasion. The choice experiments were made in a similar class-room setting with two randomly distributed versions of a simple travel-choice SP survey with or without 'referencing' to the subject's "most recent longer trip".

The study makes two contributions. First, we have replicated parts of a study made by Isacsson (2007) to investigate whether there is a hypothetical bias. Similar results were found regarding the direction of hypothetical bias in value-of-time elicitation, at least for an allocation of time to be made at another day. Second, we have tested whether 'referencing' changes the estimated value-of-time in the opposite direction to such a bias. It is shown that 'referencing' does not necessarily reduce the bias by finding significantly a lower implicit value of time when a neutral reference to the previous travel experience has been made by the participants.

The outline of the paper is the following: Next section ("[Background](#)" section) gives a more thorough account of previous literature on hypothetical bias, in particular in the case of the value of travel time. The "[Theoretical and statistical approach](#)" section presents the basic theory on how to estimate the value of time and our study design. "[Study design](#)" section gives the results and "[Results](#)" section discussion and conclusions.

Background

Hypothetical bias in SP valuation of private and public goods

The degree of hypothetical bias in stated preference studies has been investigated by comparisons of hypothetical choice experiments to economic experiments involving choice that results in a real economic transaction. It is found that the degree of the bias varies considerably and seems to depend on context; and cannot be eliminated by a rule-of-thumb approach such as division with some specific number. Murphy et al. (2005) find that the mean of the ratio of a hypothetical to a real payment in 83 observations of 28 previously published studies is 2.6. In a recent extension of this study with 71 observations including some from more recent work, Svensson (2010) estimates the mean bias at 3.4. Resolving this problem is therefore essential in making survey-based methods valid tools for practical use, such as estimation of value parameters in cost–benefit models for evaluation of investments and policies in transport, health, environmental protection, and other sectors.

Regarding the causes of hypothetical bias, several suggestions have been presented. Murphy et al. (2005) reveal some regularities contingent on the choice elicitation mechanism, the kind of participants (students/non-students) and the nature of the goods (public/private). More recently, Murphy et al. (2010) did a valuation study of their own comparing hypothetical bias when participants were provided with values of the goods ("induced" value) or had to assess values themselves ("home grown" values). With "induced" values

there was no hypothetical bias, whereas the bias was around two in the case of “home grown” values. The authors conclude that it seems that the problem arises in the value formation phase, not in the elicitation phase. Yet, there exists no widely accepted theory explaining the bias (Loomis 2011).

Hypothetical bias and scheduling constraints

As mentioned, reviews of the value of travel time have observed that SP values tend to be lower than RP values. However, experiments that allow direct comparisons are rare.¹ Brownstone and Small (2005) compare results from Stated Preference (SP) studies to Revealed Preference (RP) observational data for the valuation of travel time in the case of High-Occupancy Vehicle (HOV) Lanes in the U.S. HOV lanes offer a natural experiment situation as tolled less congested lanes are provided along toll-free highway lanes. It was found that the implicit real average value of travel time was double in magnitude compared to what SP studies indicated. Wardman et al. (2016), also find significantly larger RP valuations than SP valuations both for business and non-business trips in their extensive meta-analysis of values of time with 3190 monetary values from 389 European studies from 1963 to 2011. More similar to the studies reviewed by Murphy et al. (2005), Isacsson (2007) collected experimental data in two studies among students doing real time-use choices with economic consequences. The first study was a value-of-travel-time quasi-experiment where some students were given a choice between a slow bus and a fast bus to come to a lecture in another city, while other students responded to a hypothetical SP survey for a similar choice.² The second study was a random split-sample experiment where students were rewarded if they volunteered to fill in a lengthy research questionnaire. In both cases the estimated value of time of a real choice was double the value of time in a hypothetical setting.

Brownstone and Small (2005) suggest two explanations for the difference in the SP and RP values of travel time; neglected time constraint and misperceptions of travel time. The first is that participants to hypothetical choice questions have a tendency to consider only broad regular time constraints and to oversee more occasional time constraints that must be considered in a real choice setting. The second possible explanation is that drivers may overestimate the time loss due to congestion, and therefore be willing to pay more to avoid the congested traffic lanes than what is implied by their responses to choice questions with explicit travel-time differences.³

Regarding the first of these explanations, hypothetical value-of-travel-time questions are often not very precise about when the choice is to be made. For instance, the question may refer to a specific trip made by the subject and then ask “if you would do this trip again and you had a choice between two travel alternatives so and so, which would you chose”, i.e. the choice is not to be made ‘here and now’ but ‘later’ (i.e., the next time a similar trip will be done). Variable or occasional scheduling constraints ‘later’ may be difficult to consider and even to imagine, which implies that a hypothetical choice may be restricted by less compelling scheduling constraints (i.e., have a lower opportunity cost of time) than a real choice. In fact, there is a colloquial notion for this: “time optimism”, meaning that plans for future time-consuming activities tend to be too ambitious, possibly because of this bias.

¹ Beck et al. (2016), Fifer et al. (2014), Hossan et al. (2016) are some examples of recent studies combining RP and SP data to be able to assess potential bias in transportation research.

² Some of these students were asked another day than the day of the lecture.

³ Hensher (2010) finds this explanation less convincing.

Hensher (2010) has suggested that hypothetical bias in value-of-time studies can be overcome by the referencing, or pivot-based, designs that already are sometimes used in such studies (for instance in the recent Swedish and Norwegian value-of-time studies, Johansson et al. 2010; Ramjerdi et al. 2010), but not by Brownstone and Small (2005) or Isacsson (2007). In such designs of value-of-time studies, respondents are interviewed about the attributes (such as time and cost) of a reference trip that they already have made (or are undertaking), and are then asked to participate in a stated choice experiment with choice alternatives that are constructed as different versions of the reference trip. In this way, it may be possible to capture the context and occasion dependent constraints the individual is facing in a specific choice situation (a choice to be made ‘later’). Hensher emphasizes the habitual feature of many travel choices and claims that the reference alternative (an already experienced trip) “has important information on the marginal disutility of attribute levels associated with the experienced alternative” (Hensher 2010).

Other remedies to hypothetical bias

Besides Hensher’s suggestion that pivot-based design may reduce hypothetical bias, a range of other remedies have been explored in the literature. These include *ex ante* measures affecting how data is collected and *ex post* measures affecting the analysis of data. Some *ex ante* countermeasures, such as making respondents aware of budget constraints, are now standard procedure and others, such as so-called cheap talk scripts, have been tried with mixed results (Cummings and Taylor 1999; Little and Berrens 2004). *Ex-post* mitigation tries to remove the hypothetical bias by instrument or statistical calibration. Instrument calibration uses additional information obtained in the survey to calibrate results, while in statistical calibration this is made with bias functions derived from other samples, showing how the bias varies with different characteristics of the participants (Blackburn et al. 1994).

In instrument calibration, the most prevalent approach in recent studies is to re-code “yes” responses into “no” for respondents who reveal some degree of less than full confidence to their statement in a follow-up question (Johannesson et al. 1998, 1999; Blumenschein et al. 1998, 2008; Blomquist et al. 2009; Hedemark Lundhede et al. 2009). Another possible approach is “restricting”, that is, estimation based on the sub-sample of fully confident (or close to fully confident) respondents (Hultkrantz et al. 2006; Svensson 2009 and Sund 2009). Both approaches are based on a presumption that responses from individuals who are more certain about their stated responses (intentions) are better predictors of real behaviour, but diverge on how to treat uncertain responses. “Re-coding” assumes that uncertain yes-responses are false (while uncertain no-responses are always true), whereas the “restricting” method ignores all uncertain (yes and no) responses.⁴ The latter alternative thus has some resemblance with inclusion of a “don’t know” choice option. While “re-coding” is necessarily conservative in the sense that the share of yes-responses is decreased, “restricting” can go in either way. Also, research on the effects from having a “don’t know” option in dichotomous SP surveys does not support the assumption that all or even most uncertain respondents would reject a real offer (Balcombe and Fraser 2009).

⁴ Intermediate approaches are possible in which yes and/no responses are weighted differently depending on the degree of self-reported confidence. The Asymmetric Uncertainty Model (ASUM) multiplies the Yes (=1)/No (=0) variable by the certainty score (scaled 0–1). The Symmetric Uncertainty Model (SUM) instead adds or subtracts the certainty score to the Yes/No variable, which therefore makes a 0.5 confident Yes equal to a = .5 confident No.

A related but less recognised issue is to what extent answers to such a question really measures preference uncertainty.⁵ Possibly, some individuals always are more confident than others, irrespective of whether they made hypothetical or real choices. Furthermore, it could be that respondents report the degree of cognitive effort they used to answer the choice question, instead of the strength of belief that they would actually do as they have stated. Drawing on Kahneman's (2003) dual-process theory of decisional thinking, Regier et al. (2014) suggest that self-reported certainty can be used as a proxy for whether a respondent has used the fast, intuitive System 1 processing or the deliberate, rational System 2 processing. If so, it could be that uncertain responses are more useful as they indicate that these responses are more deliberate.

Conjectures

In our study we investigated two issues related to hypothetical bias in a value-of-time survey. First, we wanted to see whether there is a hypothetical bias when participants were given hypothetical and real willingness-to-accept, WTA, choices for "selling" time. From previous studies (Brownstone and Small 2005; Isacsson 2007), we expected that there is a negative bias at least when the hypothetical choice concerns a task to be done 'later'. We will also examine whether there is such a bias when both the real task and the hypothetical task is to be conducted now (i.e., with equal individual scheduling constraints that are known by all participants) in contrast to when the task is to be done at a similar place and situation (in a classroom at the end of a day's lectures) but at a vaguely specified future date (i.e., when the participants are likely to be more uncertain about their scheduling constraints that day). Second, after having found that there is a negative hypothetical bias in the latter case, we investigated whether a 'referencing' design of a value-of-travel-time elicitation choice experiment leads to a higher value than if no reference is made to a specific travel occasion.

Theoretical and statistical approach

The value of time is usually defined within the context of household production theory (Becker 1965; Jara-Diaz 2007). Using Becker's classic model, an individual allocates time between paid work, non-paid work and leisure and gets utility from consumption of goods (purchased or produced by non-paid work) and leisure. If for instance non-paid work is commute travel to work, then the value of a travel time change per hour will equal the after tax (net) wage rate. Instead assuming that both work time and transport time are also included in the utility function and that consumption is constrained not just by income from work but also from leisure time, the value of time will equal the net wage rate plus the difference between the marginal disutility of work and the marginal disutility of travel (both evaluated by the marginal utility of income) (see Jara-Diaz 2007). Thus value of time can be both higher and lower than the net wage.

In the first round of choice experiments we have done, we offered participants a payment for performing a 15 min task. This is thus much like offering work so if participants

⁵ In a review of empirical studies on preference uncertainty in contingent valuation, Akter et al. (2008) reflect in a final remark that "the fundamental issue that needs to be addressed at this point in the development of preference uncertainty research is whether or not respondent uncertainty can be measured accurately".

expected the task to be no more or no less discomforting than “regular work” we would expect to estimate an average value of time that is close to the net wage rate of (part-time) work available for students. However, as the task we provided (filling in a questionnaire for a student’s thesis) may be considered as an act of social responsibility, there is possibly a positive lump-sum utility component. Thus the estimated value of time may possibly be lower than the opportunity cost of time (i.e., the net wage rate), even negative.

For a long time, empirical analysis of the value of time was based on the random utility (RU) approach developed by McFadden (1974). In this, utility, U , is assumed to be linearly dependent on choice attributes, such as cost, C , and time, T . Taking the difference between two alternatives (for instance between two alternative travel modes or routes in an urban road network) and adding a stochastic term ε we have:

$$dU = \beta_T dT + \beta_C dC + \varepsilon \quad (1)$$

With appropriate statistical distribution assumptions, this can be estimated with logit or probit regression, and the value of time can then be calculated from the quota of the regression coefficients β_T/β_C , corresponding to the marginal rate of substitution. However, the linear functional form and the statistical distribution assumptions are quite restrictive, so much work has been put into elaboration of the functional form (e.g., Gaudry et al. 1989; Jara-Diaz and Videla 1989; Hensher 1997; Hultkrantz and Mortazavi 2001) and/or distributional assumptions (in recent years, predominantly by development of the mixed-logit model of McFadden and Train 2000).

In spite of such improvement of the random utility model, it remains a problem that the value of time is derived indirectly, from a quota of regression coefficients that is strongly sensitive to model misspecification. During the latest years, value-of-time research (Hultkrantz et al. 1996; Fosgerau 2007; Börjesson 2010; Johansson et al. 2010; Ojeda-Cabral et al. 2016) has therefore turned to direct estimation of value of time with the so called random valuation (RV) method. As shown by Ojeda-Cabral et al. (2016) while the random component of the RU model relates to the difference between the utilities of travel alternatives, in the RV model it relates to the difference between the value of travel time and a suggested bid value. Both Fosgerau (2007) and Ojeda-Cabral et al. (2016) find a consistent superiority of the RV approach in applications to datasets from the UK and Denmark. The RV approach (originally suggested by Cameron 1988) estimates the willingness to accept (WTA) or willingness to pay (WTP) (or the Hicksian variation) of a time change in “bid space”, i.e., from yes or no answers to a price bid (Euro per hour). This can be done both non-parametrically and parametrically when there is only one attribute (here time) besides the price (the offered payment).⁶ The simplest parametric specification of the systematic part of such a model, which was used in this study, is a first-order willingness-to-accept function:

$$dV = \alpha_0 + \alpha_1 P + \mu \quad (2)$$

where dV is the change of indirect utility from accepting the offer, P is the price bid and μ is a stochastic term. A respondent was assumed to accept the offer when $dV > 0$, hence the ratio $-\alpha_0/\alpha_1$ is the value of time (the minimum monetary compensation for the sacrifice of an hour).⁷

⁶ Or if other attributes provided in SC survey are ignored.

⁷ Note that to estimate this model we only need to vary the price bid, i.e., it is not necessary to vary time and monetary compensation separately (although in an elaborated value-of-travel-time study one may want to do that to be able to exam for instance how WTA varies with the time increment).

We further assumed that the binary response variable *acceptance* (Yes = 1, No = 0) is one when the change of indirect systematic utility plus an i.i.d. error term is positive and zero otherwise. We estimated the model as a probit model

$$A_i = \begin{cases} 1 & \text{if } \lambda_1 \text{Hyp}_i^{\text{Now}} + \lambda_2 \text{Hyp}_i^{\text{Later}} + \pi \text{Price}_i + \theta \text{Gender}_i + \epsilon_i > 0 \\ 0 & \text{if } \lambda_1 \text{Hyp}_i^{\text{Now}} + \lambda_2 \text{Hyp}_i^{\text{Later}} + \pi \text{Price}_i + \theta \text{Gender}_i + \epsilon_i \leq 0 \end{cases} \quad (3)$$

where A_i is the acceptance of the participant i , $\text{Hyp}_i^{\text{Now}}$ and $\text{Hyp}_i^{\text{Later}}$ are variables indicating the treatments that the participants received as hypothetical questionnaire suggesting the task to be performed ‘here and now’ or ‘later’. The participants who randomly received the ‘real’ treatment with real payments at the spot for undertaking the task, were the reference group. Thus, the key parameters to this study are λ_1 and λ_2 which show the mean acceptance rate differences from the ‘real’ treatment. *Price* is coded as a dummy variable indicating low (5 SEK/15 min) and high (30 SEK/15 min) bids. Due to the fact that the participants’ socio-demographics such as age and marital status show almost no variation, they were not controlled in the regressions except gender. ϵ_i is the error term assumed to have zero mean and unit variance.

In another specification, we investigated high and low prices’ interactions with the ‘hypothetical now’ and ‘hypothetical later’ treatments. The interaction model again with probit specification was estimated as:

$$A_i = \begin{cases} 1 & \text{if } \pi \text{Price}_i + \gamma_1 \text{Price}_i \text{Hyp}_i^{\text{Now}} + \gamma_2 (1 - \text{Price}_i) \text{Hyp}_i^{\text{Now}} + \\ \gamma_3 & \text{Price}_i \text{Hyp}_i^{\text{Later}} + \gamma_4 (1 - \text{Price}_i) \text{Hyp}_i^{\text{Later}} + \theta \text{Gender}_i + \epsilon_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where γ_1 gives the acceptance difference between ‘hypothetical now’ and ‘real’ treatments for the high price and γ_2 gives the acceptance difference for the low price. Similarly, to investigate the price heterogeneity for the ‘hypothetical later’ treatment case, we estimated γ_3 for the high price and γ_4 for the low price. This specification makes possible cross comparison of the responses to the treatments for the low and high bids. This is potentially important as the slopes and curvatures of the survival curves (showing how the acceptance rate declines with the bid level) may differ between the two cases.

Study design

The choice experiments were made in two rounds. The first round consisted of choice experiments made in three waves among first year Business Administration and Economics students at Örebro University in 2009. The experiment aimed to identify hypothetical bias across the alternative price values on the bidding space. Later, we conducted the second round consisting of another choice experiment aiming to relate hypothetical bias with referencing to previous experience. The second experiment was conducted in five waves with first year Business Administration, Economics, or Social Work students at Örebro University in 2015.

First round of choice experiments

Study design

The choices of the first round were made on a paper sheet distributed to all students in a class room. Students were informed by the text on the real choice sheet explaining that we

were studying response behaviour to questionnaires and that the results would be used for a master thesis in economics; that participation was voluntary and not related to the course the student was studying; that the task was to fill in one out of two questionnaires; that this was expected to take 10 min to perform, but those who accepted to perform the task would not be allowed to leave the room until 15 min have passed; and that a specific monetary compensation (one out of two or three bid levels) would be paid immediately after that.

All students were asked three questions to be answered individually: (1) Would you be willing to perform a quarter-of-an-hour task in exchange for a monetary reward (specified as a certain value, i.e., a bid, that was varied across the participants; (2) How certain are you about this decision; and (3) Are you male or female. The task was specified as filling in a one of two questionnaires. The given alternative responses to the first question was Yes and No, and to the second question a position on a Likert scale going from 1 (low certainty) to 10 (high certainty).

The first question is a standard discrete choice WTA question. Such questions are regularly asked in value of travel time studies, where for instance a commuter is asked whether she would accept a travel alternative with longer travel time than a reference alternative. Value of travel-time studies are often based on representative samples of residents or commuters, but this study is limited to a specific subpopulation. The design of the study (the first question and the task) resembles the one used by Isacsson (2007).⁸

We used a three-level bid vector at 5, 15, and 30 Swedish krona (SEK), randomly distributed among the individual students for the ‘real’ and ‘hypothetical (here and now)’ cases and 5 SEK and 30 SEK for the ‘hypothetical (later)’ case.⁹ The mid-level bid (15 SEK) was selected to approximate the wage per hour of a simple part-time work. We performed a pilot study in January on a group education students (at the start of their lunch break) using a real setting offering the high price (30 SEK). In the pilot, we noticed that all students accepted this offer and we also found that it was possible for all to fill in one of the questionnaires (both were used) within 15 min.

We only asked one question on individual characteristics (i.e., on gender), because we wanted to keep the survey very short to not miss participants with a high value of time. Also, participants were homogenous with respect to age (most were 20–30 years old) and it is difficult to get useful and comparable responses on income from students (some live with their parents, many have seasonal work, etc.), so we did not expect age and income to have much explanatory power. All were students of Business Administration classes during the two first semesters. Participation was voluntary, but the participation rate during these two first semesters was high. We kept responses from different days separate so that in the statistical analysis we could also differentiate with respect to survey day heterogeneity.

⁸ In fact, one of the tasks that we used was the same as in Isacsson’s study. This consisted of filling in a lengthy questionnaire designed by a PhD student in traffic psychology for learning about how the respondent would act in different traffic situations. However, to distract attention from the task as such, we used two different questionnaires for Swedish students (while Isacsson used only one), the second being to fill in a questionnaire that is used to assess how medical patients value their overall quality of life. Participants were given no more information on these questionnaires than that one was about “traffic safety” and the other about “quality of life”.

⁹ Skipping the mid bid in this case was unintended, but in fact it may improve the efficiency of the value-of-time estimates as it is more important to get observations of the tails of the distribution.

Treatments

With this generic value-of-time survey, we investigated three treatments by varying the phrasing of the WTA-question across participants. First, this question was framed as either a hypothetical choice, with no further consequences, ('hypothetical (here and now)') or as a real choice that was immediately followed by a 15 min task and monetary reward ('real'). In the 'hypothetical (here and now)' case, the participants were told that they were not among those that had been selected for the real task, but asked how they would have answered if the question was for real. Second, the hypothetical choice was rephrased in later separate sessions ('hypothetical (later)') as concerning a task to be after a lecture in the same class at the same time of the day but later during the semester. Thus, the three treatment are 'real' 'hypothetical (now)', and 'hypothetical (later)'.

Procedure

All participants were recruited without previous notice in a regular class-room where it was possible to get response from everyone. The teachers had been contacted in advance, but had been asked to not mention about the experiments to the students before or during the lecture. By surprising participants in this way, we wanted to avoid that results could be affected by scheduling or re-scheduling before or during class.

We came to each class a few minutes before a lecture was over. When the teacher had finished we immediately asked the students, while still seated, to fill in the one-page paper sheet with the three questions. This task took at most 2 min. In sessions that mixed 'real' and 'hypothetical (here and now)' choices, students that had been given a real offer and had accepted were asked to stay while other students (including those who had made a hypothetical choice) left the room. Thus, unlike in Isacsson (2007), in these sessions respondents to both real and hypothetical choices were in the same room, and therefore similarly exposed to any open or subtle signals from peers and instructors. The students who randomly received the experimental sheet with real payment were then given the 15 min task.¹⁰ Finally, when 15 min had passed, all questionnaires were collected and the students were paid an amount equal to the bid they had been offered individually. In each session, there was a senior researcher instructing the students. He was assisted by three or four persons so that distribution and collection of questionnaires, and subsequent payments, could be made very quickly, to keep additional time above the stated 15 min at a negligible magnitude. Additional sessions with only 'real' and 'hypothetical (later)' choices were organized in the same way.

Participants and sessions

The choice experiments were conducted in three days; the first two days were in March, 2009 with 260 students and the last day was in November 2009 with 98 students. The students were in Economics and Business Administration classes.¹¹ For the 'hypothetical

¹⁰ Students were asked to raise a hand if they did *not* have a driving license and those that did this were give the Quality of Life questionnaire, while the others got the Traffic Safety questionnaire. The attention requirement for both of the questionnaires was low as the main aim was to keep the students in the classroom for 15 min.

¹¹ The first sessions were done in three first-year student classes on March 8 and a second semester class on March 23. On March 8, two classes ended at 3:00 p.m. and the third one at 4:00 p.m. On March 23, the class

(here and now)’ choices we ended up with 96 observations in total, for the ‘real’ choices we got 112 observations, and for ‘hypothetical (later)’ we collected 150 observations, see Table 7 in the “Appendix”.

Second round of choice experiments

Study Design

The second round of choice experiments were made in a similar but simpler manner. Five student classes (here denoted Day 1–5) were visited during 2015.¹² Two versions of a simple generic value of travel-time elicitation choice survey, with or without referencing, were distributed randomly in each of the occasions. The students were asked to individually choose between a given travel alternative with train or coach for a long trip (enduring at least 1 h one way) and another alternative with the same mode with a somewhat shorter (longer) travel time and a higher (lower) ticket price.

Treatments

The basic treatments were ‘referencing’ versus no ‘referencing’ in a choice between two otherwise similar travel alternatives that differed by 15 min in travel time and 15 SEK in ticket price. Thus the break-even value of travel-time for these alternatives is 60 SEK/h.

The no-referencing question was framed in the following way:

Consider that you were to make a long trip (at least one hour one way) with train or coach. Assume that you are offered two alternatives for this trip, one standard alternative and another alternative with a longer (shorter) travel time. The only differences between these alternatives are travel time and ticket price. Would you choose the slower (faster) alternative if it would mean 15 min longer (shorter) travel time, 15 SEK lower (higher) price?

The response alternatives were Yes and No, so there was no opt-out alternative.

The ‘referencing’ question was introduced in a different manner, starting with “We ask you to consider the latest somewhat longer trip (at least 1 h one way) that you have done with train or coach.”¹³ Then four questions were asked about this trip, concerning the day of the week, the total travel time (in hours and minutes) from door to door, the travel mode and the one way ticket price. A fifth question asked whether another travel mode could have been chosen (e.g., car) (Yes/No) and, if the answer was Yes, why the train or coach alternative had been chosen (Less expensive, More convenient, Other explanation).

Footnote 11 continued

ended at 3:00 p.m. The final extra session was done November 9, in a first semester class ending at 3:00 p.m. There were no other scheduled university activities afterwards.

¹² We also checked for possible differences between WTP and WTA responses by using the latter format in one class (Day 1). Further, to check for sensitivity to the levels of the two attributes, a second choice question with another pair of attribute levels was posed in two classes (Day 1 and 2). In this the travel alternatives differed by 30 min in travel time and 60 SEK in ticket price. Thus the break-even value of time in this case is 120 SEK/h.

¹³ The reasons for the 1 hour limit was that we wanted the memory of the trip to be vivid and the travel time change to be proportionate.

The hypothetical choice was then framed in this way:

Consider that you had been able to make the same trip with the same travel mode, but with longer (shorter) travel time and lower (higher) travel cost. Would you have chosen this alternative if had meant 15 min longer (shorter) travel time, 15 SEK lower (higher) price?

Again the response alternatives were Yes and No.

Finally, as in the first round experiments we controlled for gender.

Procedure

The procedure of the second round was in most relevant aspects the same as in the first round. The main difference being that it was not necessary to control for differences between classes regarding the time of the day, so the classes were interrupted at the end of a lecture hour, just before the time of a break.

Participants and sessions

The participants were students at Örebro University in various classes.¹⁴ In total, 138 responses for the ‘referencing’ group and 162 responses for ‘no referencing’ were collected (300 observations), see Table 7 in the “Appendix”.

Results

Descriptive results

In Table 1 we summarize the data from the first and second rounds, split on the three treatment groups in the first round and the two treatment groups in the second round. It can be noted that the distribution of sexes is quite even. In the first round, the acceptance rate was highest in the ‘hypothetical (later)’ sub-sample (0.67). In the second round, the acceptance rate was higher in the ‘referencing’ subsample.

A further description of the acceptance indication is provided by Fig. 1, showing the share of yes responses in the three treatment groups in first round and the two treatment groups in second round Day 1 and 2. It shows that the acceptance rate increases in all samples monotonously with the size of the payment. In the first round one-third of the participants rejected the highest bid when it was for real. We therefore lack observations from the high end of the value-of-time distribution, and we have no information on response rate for negative bids. More to the point of our study, we see that the acceptance rate for ‘hypothetical (later)’ is above the corresponding rate for the ‘real’ treatment, while the acceptance rates for ‘hypothetical (here and now)’ and the ‘real’ treatments are close. In the second round, introducing referencing resulted in higher acceptance rate as seen in the acceptance rate for ‘referencing’ group being above to the one for the ‘no reference’ case.

¹⁴ The five experiments were held on January 30 (at 14.00), February 19 (at 14.00), April 13 (at 12.00), April 23 (at 12.00) and May 19 (at 10.15). The classes were Economics A-level, Business administration B-level, Business administration A-level, (Ekonomie Kandidat programme), Social work-A level, and Business administration A-level, (Civilekonom programme), respectively.

Table 1 Descriptive statistics of the first and second rounds' data

Variable	Real	Hypo Now	Hypo Later	Reference	No reference	Min	Max
Price	16.61 (10.47)	16.51 (10.24)	17.5 (12.54)	15 (0.00)	15 (0.00)	5	30
Female	0.40 (0.49)	0.50 (0.50)	0.55 (0.51)	0.48 (0.50)	0.54 (0.50)	0	1
Accept	0.51 (0.50)	0.51 (0.50)	0.67 (0.47)	0.53 (0.50)	0.31 (0.46)	0	1
No. of obs	112	96	150	138	162		

Unweighted averages of each variable with standard errors in parentheses

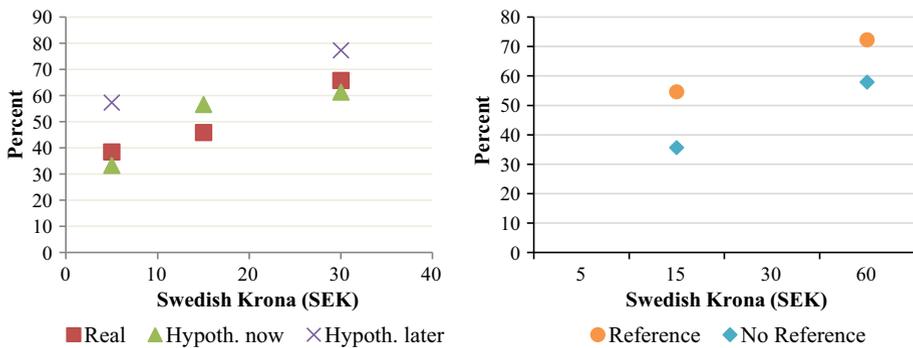


Fig. 1 Acceptance rates (percent) at different bid levels (SEK) for the first and second rounds, respectively

First round results on hypothetical bias

Table 2 shows the probit estimates of three versions of the model including the gender covariate and the survey day controls.¹⁵ The survey day variables are not significant for any of the estimations. The *price* variable is positive and significant in all versions, indicating as expected an upward sloping WTA (supply) relation. Further we find that there is no significant difference between the ‘hypothetical (here and now)’ and the ‘real’ treatments. In the basic model without interactions (column 1), the ‘*hypothetical later*’ variable has a significant and positive coefficient suggesting that there is a negative hypothetical bias in the ‘later’ treatment. Further in the basic model version the variable *female* is not significant.

The second column shows results with interactions between the treatment variables and the high and low end of the bid space. These results suggest that hypothetical bias in the ‘later’ context is an issue for low bid levels, not for high bids, while the comparison between ‘hypothetical (here and now)’ and ‘real’ treatments does not show any sensitivity to the high and low end of the price space. The third column shows results for interactions with the gender variable, indicating that the results for the bias in the ‘hypothetical (later)’ case are driven by the response from females.

¹⁵ The model with interaction variables is split into two versions (columns 2 and 3) due to collinearity.

Table 2 Probit model of estimations for the first round of experiments

Variables	(WTA) (1)	(WTA) (2)	(WTA) (3)
Hypo (now)	-0.105 (-0.384)		
Hypo (later)	0.491** (2.017)		
Price	0.651*** (4.188)	0.703** (2.385)	0.650*** (4.153)
Female	0.180 (1.172)	0.181 (1.174)	-0.144 (-0.482)
2.sday	-0.052 (-0.161)	-0.054 (-0.165)	-0.030 (-0.091)
3.sday	-0.159 (-0.686)	-0.160 (-0.696)	-0.166 (-0.713)
L_price*Hypo (now)		-0.119 (-0.344)	
H_price*Hypo (now)		-0.089 (-0.252)	
L_price*Hypo (later)		0.547* (1.863)	
H_price*Hypo (later)		0.427 (1.361)	
Female*Hypo (now)			-0.122 (-0.321)
Male*Hypo (now)			-0.045 (-0.137)
Female*Hypo (later)			0.860*** (2.707)
Male*Hypo (later)			0.195 (0.674)
Constant	-0.337* (-1.895)	-0.363* (-1.686)	-0.215 (-1.073)
No. of obs.	290	290	290
Pseudo R ²	0.0722	0.0727	0.0855

Robust z-statistics are in parentheses and ***, **, * refer to 1, 5 and 10% significance levels

Second round results on ‘referencing’

The main aim of the second round of experiments was to study whether responses to hypothetical questions for eliciting the value of time are affected by whether reference is made to a specific situation. For this purpose we compared responses to two alternative versions of a standard hypothetical travel-choice experiment survey, with or without ‘referencing’ to a specific previous trip made by the participant. The probit estimation results from the second round are found in Tables 3 and 4, referring to the basic and interaction model specification, respectively. In Table 3 we make comparisons between WTA and WTP results possible. The results are shown separately for Day 1 (WTP)

(column 1), Day 2–5 (WTA) (column 2), and for all days (column 3), where all responses have been coded as WTA responses (i.e., by changing signs of the two attribute variables). When the signs are reversed, there are no significant differences in responses between Day 1 and the other sessions. All three columns show that the ‘referencing’ design leads to a significantly higher acceptance rate. The gender covariate is not significant.¹⁶

In Table 4, where we explore first the interactions between gender and the ‘referencing’ (column 2) and also the survey day heterogeneity (column 3), we find that the positive significance of the ‘referencing’ treatment holds and is driven by the responses from male participants.

Check for other possible remedies to hypothetical bias

Based on the first round of experiments a check of the effects of other possible remedies than ‘referencing’ was made by employing two forms of instrument calibration from the responses to the certainty variable. To demonstrate these effects the implicit values of times in the three treatment groups were calculated from each treatment group. The values of time in the three treatment groups are shown in Table 5. The estimated ‘real’ case value of time is 63 SEK per hour. Although this mean estimate should not be taken too seriously given that the tails of the distribution are not fully covered by the bid vector, we notice that it is close to the current after tax minimum wage of work that is common among students as part-time work, for instance work as shop assistant,¹⁷ which suggests that the students on average were regarding the questionnaire responding task as equivalent to such work.

The sample of respondents that got the ‘hypothetical (now)’ survey on average revealed a value of time of 62 SEK per hour, i.e., just one percent below the value from the ‘real’ choice. A different picture is given by the ‘hypothetical (later)’ sample. The point estimate of the Swedish corresponding value is SEK -13 per hour.¹⁸

Table 6 then shows results from instrument calibration with the two methods. For fully confident respondents (level 9 and 10 or just 10) both methods indicate a negative value of time. These results are difficult to interpret as the standard deviations are very high.¹⁹

Discussion

In digestion of the results, we begin to notice that we found no hypothetical bias in the first round when respondents to a real and hypothetical value-of-time WTA question were in an equal choice situation because both the real and the hypothetical task would be conducted at once and at the same location. These sessions were made at the end of the day when

¹⁶ There were no significant differences between Day 1–5 sessions.

¹⁷ The minimum wage is SEK 87.44 per hour; after tax (marginal tax rate 26%) amount is SEK 64.71 per hour.

¹⁸ As the task provided (filling in a questionnaire for a student’s thesis) may be considered as an act of social responsibility there might be a positive lump-sum utility component. Thus the estimated value of time may possibly be lower than the opportunity cost of time (i.e., the net wage rate), even negative. This may be of interest for future research.

¹⁹ For the first round, we conclude that none of the methods seem to work. For the second round, we found that 55 percent of the “Reference” group remembered all of the details that they were asked about their most recent trip. As a robustness check we rerun the second round’s estimation both by restricting the “Reference” group to these participants (combined with a 55 percent random sample from “No reference” group) and recoding the “Reference” group answers in the case of any missing information among the recent trip’s reminder questions. The results were found robust.

Table 3 Probit model of responses for the second round of experiments

Variables	(WTP) Day 1 (1)	(WTA) Day 2–5 (2)	(WTA) Day 1–5 (3)
Referencing	−0.525** (−2.024)	0.582*** (3.179)	0.570*** (3.8344)
Female	0.161 (0.616)	0.036 (0.196)	−0.033 (−0.225)
Constant	0.359 (1.630)	−0.547*** (−3.434)	−0.481*** (−3.754)
No. of obs.	101	199	300
Pseudo R ²	0.0377	0.0384	0.0371

Robust z-statistics are in parentheses and ***, **, * refer to 1, 5 and 10% significance levels. Columns refer to different grouping of the experiment days. Column 1 results are obtained from Day 1 only with WTP questions and Column 2 results are from Day 2–5 only with WTA questions. Column 3 refers to all days (1–5) with WTA questions including the modified Day 1, from WTP to WTA

Table 4 Interaction model specifications for the probit model of responses of the second round of experiments

Variables	(WTA) (1)	(WTA) (2)	(WTA) (3)
Referencing	0.570*** (3.834)		0.572*** (3.802)
Female	−0.033 (−0.225)	0.274 (1.310)	−0.104 (−0.666)
2.sday			0.101 (0.531)
3.sday			−0.481* (−1.772)
4.sday			0.363 (1.445)
5.sday			−0.478** (−2.011)
Female*Referencing		0.265 (1.280)	
Male*Referencing		0.900*** (4.136)	
Constant	−0.481*** (−3.754)	−0.653*** (−4.139)	−0.407** (−2.507)
No. of obs.	300	300	300
Pseudo R ²	0.0371	0.0482	0.0688

Robust z-statistics are in parentheses and ***, **, * refer to 1, 5 and 10% significance levels

most students leave the university. They were not informed in advance and some of them wanted to rise from their seats in the same moment as the teacher finished and before we had asked them to fill in the questionnaire. When they finally left some seemed to be on a rush. This indicates that some of the students indeed had “scheduling constraints” and that there were individual differences in the value of time. However, when participants considered their alternative uses of the time that was asked for, by our design there should

Table 5 Value of time (SEK/hour), standard errors in parentheses

	Real	Hypo now	Hypo later
VOT (probit)	62.95 (17.16)	61.88 (20.11)	-12.72 (35.31)

VOT is calculated from estimations of probit models (without covariates) on each separate sample. Standard errors are computed by the delta method

Table 6 Certainty calibration of the “Hypothetical (later)” sample using the “re-coding” and “restricting” methods

	<i>Recoding</i>		<i>Restricting</i>	
	VOT (SEK/hour)	No. of obs.	VOT (SEK/hour)	No. of obs.
Estimated real	62.95		62.95	
≥9	-229.15 (237.40)	150	-17.95 (61.67)	43
10	-276.65 (266.08)	150	-66.04 (282.7)	28

Standard errors within parentheses (delta method). First row denotes results based on the “Real” sample and the rest are from “Hypothetical (later)” sample, at different cut-off points of the ‘certainty’ variable

have been no systematic differences between participants responding to a real or a hypothetical choice question as all had to consider doing the task immediately and at the same location. This finding is therefore consistent with the finding by Murphy et al. (2010) that hypothetical bias is not connected to the value elicitation, i.e., given that both the real and the hypothetical treatment groups were in equal circumstances, no differences in average responses are found.

In contrast, respondents asked to consider a hypothetical choice for a task to be made at a later occasion answered differently. This underlines that a major challenge in designing a hypothetical choice survey is to ascertain that respondents do not systematically consider a different choice set than that of a real choice. The second noticeable finding from our experiments is therefore that we, as Brownstone and Small (2005) and Isacsson (2007), do find a *negative* hypothetical bias in stated hypothetical choice related to time. In the first round we conducted the two ‘hypothetical (later)’ sessions with two other classes than the ‘real’ and ‘hypothetical (here and now)’ sessions, but the conditions were very similar (business administration students of the first two semesters at the same university and at the same time of the day) and there were no significant differences between the two ‘hypothetical (later)’ sessions.

So called certainty calibration is of no use as a remedy to a negative hypothetical bias as a re-coding of uncertain Yes responses to No would only aggravate the bias by further reducing the value of time in the hypothetical case. Instead, we designed the second round choice experiment to investigate the claim that a ‘referencing’ design would help in such a case. As a precondition for using such a design is that the ‘referencing’ can be made to a previous experience; we framed the second round in a standard value of time elicitation setting as a choice between two travel alternatives. As this is different from the framework of the first round, no direct comparisons can be made between acceptance rates or the

implicit value of time. However, the results clearly indicate that the ‘referencing’ treatment increased the acceptance rate (when corresponding to a WTA choice) and thus would have reduced the estimated implicit value of time. Thus our results do not support that ‘referencing’ reduces a negative hypothetical bias from stated-preference based value of time estimation.

An interesting finding is that these results are driven by different gender groups. Similar findings have been made by Mitanis and Flores (2009) in a study of hypothetical bias in elicitation of the WTP for a public good. These authors report that females are more likely to truthfully reveal their value than males through hypothetical payments, but gender is not significant for truthfully revealing their value through real payments.

A final remark is that the literature based on incentivized and hypothetical stated preferences experiments discusses the effects of price levels on decisions (Morkbak et al. 2010; Cooper and Loomin 1992; Carlsson and Martinsson 2008). When we investigated this aspect in the first round of experiments, no sensitivity was traced for the comparison between ‘hypothetical (here and now)’ and ‘real’ treatments to whether price was high, low or in between. For this reason we proceeded in the subsequent rounds with the mid bid (15 SEK for 15 min) corresponding to the average hourly wage rate. However, we detected hypothetical bias in the ‘later’ treatment for the low bid. It is documented in stated-choice value of time studies that there may exist non-linearities, which depend on the size of the cost/time savings relative to the total cost/time of the journey (Daly et al., 2014). In future studies it may therefore be worthwhile to explore the effect of alternative bid spaces.

Conclusions

From these results we infer that negative hypothetical bias is a potentially important issue in value of time elicitation. The results therefore point to the importance of decision and information timing in value of time elicitation.

The main conclusion for value-of-time research from our study is that hypothetical bias should be as much a concern to designers of surveys and users of results from these surveys as it already is in other fields of non-market valuation, in particular environmental economics. It also highlights the important role of timing of choice in eliciting value of time and time reliability, as timing issues affect both the real and the perceived opportunity cost of time/scheduling constraints (as suggested by Börjesson et al. 2010).

While the claim that a ‘referencing’ design is a remedy to negative hypothetical bias gets no support from this study we do not think that this study can be a base for drawing definite conclusions on this method. Further studies are needed for the exploration of the possible heterogeneity within the ‘referencing’ treatment.

Acknowledgements We are grateful to Shengcong Xue for research assistance and for constructive comments from Gunnar Isacsson, Glenn Harrison, Elisabet Rutström, Jonas Eliasson and two anonymous reviewers on previous versions of this paper.

Author contribution Both authors participated in the collection of data for this paper and in the writing. Savsin conducted the statistical analysis. Shengcong Xue assisted in the first round of the data collection.

Compliance with ethical standards

Conflict of interest On behalf of both authors, the corresponding author states that there is no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Table 7.

Table 7 Collection of the data: number of observations per session and the treatment

	Round 1			Round 2					Total
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	
Real	112								112
Hypo now	45	51							96
Hypo later	52		98						150
Ref				46	40	14	18	20	138
No ref				55	43	16	20	28	162
Total	209	51	98	101	83	30	38	48	658

References

- Abrantes, P.A.L., Wardman, M.R.: Meta-analysis of UK values of travel time: an update. *Transp. Res. Part A* **45**(1), 1–17 (2011)
- Akter, S., Bennet, J., Akhter, S.: Preference uncertainty in contingent valuation. *Ecol. Econ.* **67**, 45–351 (2008)
- Balcombe, K., Fraser, I.: Dichotomous-choice contingent valuation with ‘don’t know’ responses and misreporting. *J. Appl. Econom.* **24**(7), 1137–1152 (2009)
- Beck, M.J., Fifer, S., Rose, J.M.: Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty. *Transp. Res. B* **89**, 149–167 (2016)
- Becker, S.G.: A theory of the allocation of time. *Econ. J.* **75**, 493–517 (1965)
- Blackburn, M., Harrison, G.W., Rutström, E.E.: Statistical bias functions and informative hypothetical survey. *Am. J. Agric. Econ.* **76**, 1084–1088 (1994)
- Blomquist, G.C., Blumenschein, K., Johannesson, M.: Eliciting willingness to pay without bias using follow-up certainty statements: comparisons between probably/definitely and a 10-point certainty scale. *Environ. Resour. Econ.* **43**, 473–502 (2009)
- Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N., Freeman, P.: Eliciting willingness to pay without bias: evidence from a field experiment. *Econ. J.* **118**, 114–137 (2008)
- Blumenschein, K., Johannesson, M., Blomquist, G.C., Liljas, B., O’Conor, R.M.: Experimental results on expressed certainty and hypothetical bias in contingent valuation. *South. Econ. J.* **65**, 169–177 (1998)
- Börjesson, M.: Swedish Values of Travel Time and Their Application in Appraisal, Royal Institute of Technology, Centre of Transport Studies, Stockholm, Working Paper (2010)
- Börjesson, M., Eliasson, J., Franklin, J.P.: Valuations of Travel Time Variability in Scheduling Versus Mean-Variance Models, Royal Institute of Technology, Centre for Transport Studies, Stockholm, working paper (2010)
- Brownstone, D., Small, A.K.: Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transp. Res. Part* **39**, 279–293 (2005)
- Cameron, T.A.: A new paradigm for valuing non-market goods using referendum data: maximum likelihood estimation by censored logistic regression. *J. Environ. Econ. Manag.* **15**, 355–379 (1988)
- Carlsson, F.: Design of stated preference surveys: is there more to learn from behavioral economics. *Environ. Res. Econ.* **46**, 167–177 (2010)

- Carlsson, F., Martinsson, P.: How much is too much? An investigation of the effect of the number of choice sets, context dependence and the choice of bid vectors in choice experiments. *Environ. Res. Econ.* **40** (2), 165–176 (2008)
- Cooper, J.C., Loomis, J.B.: Sensitivity of willingness to pay estimates to bid design in dichotomous choice contingent valuation models. *Land Econ.* **68**(2), 221–224 (1992)
- Cummings, R.G., Harrison, G.W., Rutström, E.: Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive compatible? *Am. Econ. Rev.* **85**(1), 260–266 (1995)
- Cummings, R.G., Taylor, L.O.: Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *Am. Econ. Rev.* **89**(3), 649–665 (1999)
- Daly, A., Tsang, F., Rohr, C.: The value of small time savings for non-business travel. *J. Transp. Econ. Pol.* **48**, 205–218 (2014)
- Fifer, S., Rose, J., Greaves, S.: Hypothetical bias in stated choice experiments: is it a problem? And if so, how do we deal with it? *Transp. Res. Part A* **61**, 164–177 (2014)
- Fosgerau, M.: Using nonparametrics to specify a model to measure the value of travel time. *Transp. Res. Part A* **41**, 842–856 (2007)
- Gaudry, M., Jara-Diaz, S.R., Dios Ortúzar, J.D.: Value of time sensitivity to model specification. *Transp. Res. B* **23**, 151–158 (1989)
- Harrison, G.W.: Experimental evidence on alternative environmental valuation methods. *Environ. Res. Econ.* **34**, 125–162 (2006a)
- Harrison, G.W.: Making choice studies incentive compatible. In: Kanninen, B.J. (ed.) *Valuing Environmental Amenities Using Stated Choice Studies*. Springer, Berlin (2006b)
- Harrison, G.W., Rutström, E.: Experimental evidence on the existence of hypothetical bias in value elicitation methods. In: Plott, C., Smith, V.L. (eds.) *Handbook of experimental economics results*. Elsevier Science, New York (2008)
- Hedemark Lundhede, T., Olsen, S.B., Jacobsen, J.B., Thorsen, B.L.: Handling respondent uncertainty in Choice Experiments: evaluating recoding approaches against explicit modeling of uncertainty. *J. Choice Model.* **2**(2), 118–147 (2009)
- Hensher, D.A.: Behavioral value of travel time savings in personal and commercial. In: Greene, D.L., Jones, D.W., Delucchi, M.A. (eds.) *The Full Costs and Benefits of Transportation*, pp. 245–279. Springer, Berlin (1997)
- Hensher, D.A.: Hypothetical bias, choice experiments and willingness to pay. *Transp. Res. B* **44**, 735–752 (2010)
- Hossan, M.S., Asgari, H., Jin, X.: Investigating preference heterogeneity in value of time and value of reliability estimation for managed lanes. *Transp. Res. Part A* **94**, 638–649 (2016)
- Hultkrantz, L., Li, C., Lindberg, G.: Some problems in the consumer preference approach to multimodal transport planning. Presented at the conference Transport and Environment, FEEM, Venice, 9–10 Nov 1995, Dalarna University, CTS Working Paper, 1996:5 (1996)
- Hultkrantz, L., Lindberg, G., Andersson, C.: The value of improved road safety. *J. Risk Uncert.* **32**, 151–170 (2006)
- Hultkrantz, L., Mortazavi, R.: Anomalies in the value of travel-time change. *J. Transp. Econ. Pol.* **35**(2), 285–300 (2001)
- Isacsson, G.: The tradeoff between time and money: Is there a difference between real and hypothetical choices? In: Swedish National Road and Transport Research Institute Transport Economics Research Unit (available at the Scandinavian Working Papers in Economics web site) (2007)
- Jara-Diaz, S.R.: Allocation and valuation of travel-time savings, Chapter 18 in Hensher. In: Button, D.A.J. (ed.) *Handbook of Transport Modelling*, 2nd edn. Elsevier, New York (2007)
- Jara-Diaz, S.R., Videla, J.: Detection of income effect in mode choice. *Transp. Res. B* **23**, 393–400 (1989)
- Johansson, A., Algers, S., Börjesson, M., Sundbergh, P., Byström, C., Almström, P.: Trafikanter värdering av tid—Resultat från den nationella tidvärdesstudien 2007/08, (Transport users valuation of time—results from the national value of time study 1997/08), WSP Sweden, Analys och Strategi Rapport, 11 (in Swedish) (2010)
- Johannesson, M., Liljas, B., Johannesson, P.-O.: An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions. *Appl. Econ.* **30**, 643–647 (1998)
- Johannesson, M., Blomquist, G.C., Blumenschein, K., Johannesson, P.-O., Liljas, B., O’Conor, R.M.: Calibrating hypothetical willingness to pay responses. *J. Risk Uncertain.* **18**, 21–32 (1999)
- Kahneman, D.: Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* **93**(5), 1449–1475 (2003)
- Levitt, S.D., List, J.A.: What do laboratory experiments measuring social preferences reveal about the real world. *J. Econ. Persp.* **21**(2), 153–174 (2007)

- Little, J., Berrens, R.: Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. *Econ. Bull.* **3**(6), 1–13 (2004)
- Loomis, J.: What's to know about hypothetical bias in stated preference studies? *J. Econ. Surv.* **25**(2), 363–370 (2011)
- McFadden, D.: The measurement of urban travel demand. *J. Publ. Econ.* **3**, 303–328 (1974)
- McFadden, D., Train, K.: Mixed MNL models for discrete response. *J. Appl. Econom.* **15**(5), 447–470 (2000)
- Mitani, Y., Flores, N.: Demand revelation, hypothetical bias, and threshold public goods provision. *Environ. Res. Econ.* **44**(2), 231–243 (2009)
- Morkbak, M.R., Christensen, T., Gyrd-Hansen, D.: Choke price bias in choice experiments. *Environ. Res. Econ.* **45**(4), 537–551 (2010)
- Murphy, J.J., Allen, P.G., Stevens, T.H., Darryl, W.: A meta-analysis of hypothetical bias in stated preference valuation. *Environ. Res. Econ.* **30**, 313–325 (2005)
- Murphy, J.J., Stevens, T.H., Yadav, L.: A comparison of induced value and home-grown value experiments to test for hypothetical bias in contingent valuation. *Environ. Res. Econ.* **47**, 111–123 (2010)
- Ojeda-Cabral, M., Batley, R., Hess, S.: The value of travel time: random utility versus random valuation. *Transportmetrica A Transp. Sci.* **12**(3), 230–248 (2016)
- Ramjerdi, F., Flügel, S., Samstad, H., Killi, M.: Value of time, safety and environment in passenger transport-time. Transportøkonomisk institutt (www.toi.no), Oslo, Publication 1053b (2010)
- Regier, D.A., Watson, V., Burnett, H., Ungar, W.J.: Task complexity and response certainty in discrete choice experiments: an application to drug treatments for juvenile idiopathic arthritis. *J. Behav. Exp. Econ.* **50**, 40–49 (2014)
- Shires, J.D., de Jong, G.C.: An international meta-analysis of values of travel time savings. *Eval. Progr. Plan.* **32**, 315–325 (2009)
- Small, K.A.: Valuation of travel time. *Econ. Transp.* **1**(1–2), 2–14 (2012)
- Sund, B.: Certainty calibration in contingent valuation—exploring the within-difference between dichotomous choice and open-ended answers as a certainty measure. Örebro University, Dept. of Econ. Working Paper 2009:1 (2009)
- Svensson, M.: The value of a statistical life in Sweden estimates from two studies using the certainty approach calibration. *Acc. Anal. Prev.* **41**(3), 430–437 (2009)
- Svensson M.: Hypothetical bias in stated preference methods—how big of a problem and what can be done? In: Kinell, G., Söderqvist, T., Hasselström, L. (eds.) Default monetary values for environmental change, *Swedish Environmental Protection Agency*, Report #6323 (2010)
- Wardman, M., Chintakayala, V.P.K., De Jong, G.: Values of travel time in Europe: review and meta-analysis. *Transp. Res. Part A* **94**, 93–111 (2016)

Lars Hultkrantz is a professor of economics at Örebro University. His current research interests are transport economics, health economics and cost–benefit analysis.

Selen Savsın is a PhD student in economics at Örebro University. She is conducting research within international economics and behavioural economics.