



Travellers' perceptions of travel time reliability in the presence of rare events

Emily K. M. Moylan¹ · Michiel C. J. Bliemer¹ · Taha Hossein Rashidi²

Accepted: 25 June 2021 / Published online: 10 November 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Travellers account for variability in transport system performance when they make choices about routes, modes and destinations. Modellers try to quantify travel time reliability through various dispersion measures, most commonly the standard deviation of travel time. However, standard deviation is only one attribute of the nuanced travel time distribution. This paper considers whether standard deviation is sufficient to describe the travellers' understanding and value of travel time reliability and how we might include other aspects of variability such as the frequency of exceeding a lateness threshold or the likelihood of rare events. Car drivers in New South Wales, Australia, were asked to reconstruct the distribution of their commuting time and identify a lateness threshold. Further, we asked them about their preferences in a series of stated choice experiments using three representations of travel time reliability pivoted around their regular commute. The results show reliability ratios consistent with those in the literature for all three presentations. Moreover, the standard error of the estimated coefficient on the risk of rare events indicates that standard deviation alone may not sufficiently capture travellers' preferences towards travel time reliability.

Keywords Reliability · Rare events · Travel behaviour · Discrete choice model · Stated preference

Background

Inherent variability in outcomes is recognised through the inclusion of risk attitude to help explain observed heterogeneity in behavioural responses (Barsky et al. 1997; Drottz-Sjöberg 1991; Ida and Goto 2009). Risk attitude describes an individual's

✉ Emily K. M. Moylan
emily.moylan@sydney.edu.au

Michiel C. J. Bliemer
michiel.bliemer@sydney.edu.au

Taha Hossein Rashidi
rashidi@unsw.edu.au

¹ University of Sydney, Sydney, Australia

² UNSW Sydney, Sydney, Australia

propensity to avoid or seek low-probability outcomes. Decision making under risk occurs in many disciplines, including health (e.g., choosing treatments with possible side-effects) and marketing (e.g., choosing products with different failure rates). In transport, one application where risk enters into behavioural modelling is the inclusion of travel time reliability in route, destination, departure time, or mode choice models. In this context, expected travel time and the reliability of the travel time are both factors in the decision. Significant work has been done to establish frameworks for incorporating travel time reliability and measuring its value (Carrion and Levinson 2012a).

In the transport context, the role of risk attitude in valuing reliability has been explicitly studied (Beaud et al. 2016; Chen et al. 2002; Hensher et al. 2013). However, some types of unreliability are riskier than others. By focusing on the mean–variance model, value-of-reliability studies may be missing the value that individuals place on other attributes of the travel time distribution such as asymmetry or the importance of the distribution’s tail. The rare events associated with the tail present several challenges for both the decision maker and the modeller.

Travel time distributions can be unreliable in many ways. For example, the level of service of a facility downstream of a bottleneck might be resilient to fluctuations in demand but may have an important tail of long travel times associated with incidents or adverse weather conditions. Another facility might have a wide distribution of travel times because it usually operates at or near capacity where the travel time is most sensitive to demand fluctuations. Yet another facility might regularly operate in one of two modes associated with two traffic states, e.g., due to level rail crossings or tidal-flow lanes—this could result in a bimodal distribution of travel times. A popular measure to describe unreliability is the standard deviation of the travel time distribution, although other measures such as variance, percentiles, and range have been proposed to capture more nuance of the distribution. de Jong and Bliemer (2015) reported that the standard deviation is the most preferred measure among experts to determine the value of reliability in cost–benefit analysis. They also summarise the arguments for and against standard deviation to which we add the advantage that it can be calculated from sparse data.

The standard deviation of travel time for the three facilities described above may be the same, but travellers may value the corresponding reliability differently. Furthermore, by including only one reliability measure in utility functions to describe travel behaviour, we are unable to determine whether travellers value that metric independently from the others or as a proxy for other forms of unreliability. This insufficiency in the current methods has been approached in a number of ways including separating out components of unreliability (Soza-Parra et al. 2019, 2021), introducing a heterogeneous traveller population (Kato et al. 2020) and stochastic traffic assignment that does not require the reconstruction of the travel time or demand distributions (Zhu et al. 2021).

Other measures of reliability capture the variety of ways that travel can be unreliable. Generally speaking, this could relate to any aspect of the trip such as the reliability of finding nearby parking, the reliability of getting a seat on the train, or the reliability of being involved in a crash or breakdown. With respect to travel time reliability, there is the day-to-day variation most often captured by the standard deviation and there is also the likelihood of experiencing an extreme travel time. Due to the asymmetry of the travel time distribution with firm lower bounds imposed by speed limits, extremely long travel times are more relevant than extremely short ones. In contrast to standard deviation of travel time, this type of reliability needs to be described by two attributes: how long the travel time is (severity) and how often it occurs (probability).

The inverse correlation between probability and severity for typical travel time distributions means that travellers have comparatively little experience with or information about rare events. The peak-and-end rule (Kahneman et al. 1993) describes the importance of the most extreme and most recent events in assigning utility to alternatives. As specific events are the drivers of the evaluation, peak-and-end rule decisions relate to *episodic memory* which is the type of explicit long-term memory that pertains to memories of specific events or episodes (Tulving 1972). This framework minimises the role that either typical travel time or variation in travel time play in travel choices, and these attributes are examples of *semantic memory* which is the type of explicit long-term memory that pertains to concepts based on knowledge learned in the past (Tulving 1972). Geng et al. (2013) show that the peak-and-end rule explains retrospective hedonic evaluations over short retention intervals (3–7 weeks) whereas longer retention intervals are dominated by semantic memory. Based on the frequency of rare events, they might contribute to episodic memory if they occurred recently (for example, 5% of weekdays or once every 4 weeks on average) or semantic memory if they are sufficiently rare to have not occurred in the last 3–7 weeks (for example, 2% of weekdays or once every 10 weeks on average).

De Palma et al. (2014) discuss that rare events might be under or overvalued if they cannot be communicated succinctly. In stated preference experiments, novel designs for succinctly communicating trade-offs in risk and severity have been used to value rare events in health contexts (de Bekker-Grob et al. 2013; Determann et al. 2014). For unscheduled modes like driving or walking, succinct communication about rare events can be difficult because the vocabulary associated with being on-time is not appropriate. This requires a presentation that simultaneously defines failure and describes its likelihood.

The nuances in how variable performance impacts travellers can be captured with the appropriate models, but previous research has focused on simpler models that were appropriate for a data-sparse environment. Due to the growing availability of data, travel time distributions are known in greater detail than ever.

A preliminary exploration of the value of rare events in travel choices was presented by Moylan et al. (2019). The aim of this paper is to understand whether standard deviation of travel time is sufficient to describe the travellers' understanding and value of travel time reliability and how we might include other aspects of variability such as the frequency of exceeding a lateness threshold or the likelihood of rare events. This investigation requires that survey respondents are presented with more information and in different ways than past studies in order to separate these aspects. The contributions of this paper include an evaluation of three designs for the presentation of travel time distributions accounting for rare events and preliminary evidence for the importance of alternative measures of travel time variability.

The section below presents a dataset and methodology for exploring the travellers' understanding of their travel time distributions with rare events and the role unreliability plays in their behaviour. This is followed by the results of the survey, including comparisons of alternative presentations. The results of the models are presented in the next section. The findings are discussed and placed into context in the final section.

Data collection methods

This study is based on an online survey among car commuters in New South Wales, Australia, consisting of revealed preference and stated preference questions. The survey was custom built to provide interactive maps for route selection and generation of statistically consistent personalised alternatives. The design of the study and the survey instrument was reviewed and approved by the University of Sydney Human Research Ethics Committee (Application 2019/437). It was distributed to the participants through a global market research firm. In the following subsections we provide an overview of the survey, the identification and use of a lateness threshold, the design of the choice experiment, and the presentation styles.

Survey overview

The overall flow of the survey is summarised in Fig. 1. First, in the revealed preference part, respondents are asked about their usual commuting route in order to later construct personalised choice tasks in the stated preference part. The respondents use a map interface as shown in Fig. 2a in order to indicate their typical preferred route. Respondents are prompted to reconstruct their travel time distribution using travel times associated with 20 trips as shown in Fig. 2b.

Lateness threshold

In order to personalise the trade-off between severity and risk, the respondents are asked to provide perceived realised travel times as well as their assessment of what travel times should be considered delayed. Tolerance to variation has been captured as a threshold in the literature (Taylor 2013) including as an explicit value for freight planning (Zhang et al. 2017), as a means of segmenting respondents based on flexibility of arrival time (Asensio and Matas 2008), as a parameter in probabilistic reliability measures (e.g. van Lint et al. 2008) or as a budget/buffer used by travellers when they plan (e.g. Chen and Zhou 2010).

Using attributes of the reported distribution, the respondents are asked to identify a lateness threshold using binary search (Knuth 1998). Binary search has previously been applied to lateness thresholds by Zhang et al. (2014), and it is used here because it can identify the threshold faster than a linear search, which makes it less onerous for the respondent. Respondents are asked if a travel time one standard deviation above their typical travel time would be within the normal day to day variation. Based on their answer, the journey time is either raised or lowered until a lateness threshold can be identified within a single histogram bin. The process is illustrated in Fig. 3 where the respondent's interface is shown with three iterations of the binary search procedure. The search continues until the size of the jump is smaller than the size of the histogram bin and the last step, shown in lighter blue in Fig. 3a, is inferred from the answer to the previous questions, see Fig. 3b.

The final piece of information needed to personalise the trade-off between risk and severity is how often travel time exceeds the lateness threshold. The respondents are asked to report this as a percentage. As shown in Fig. 4, the reported numerical value is shown side-by-side with a graphical display—this display is used later in one of the choice task presentations and also assists respondents who prefer a graphical layout to conceptualise likelihood.

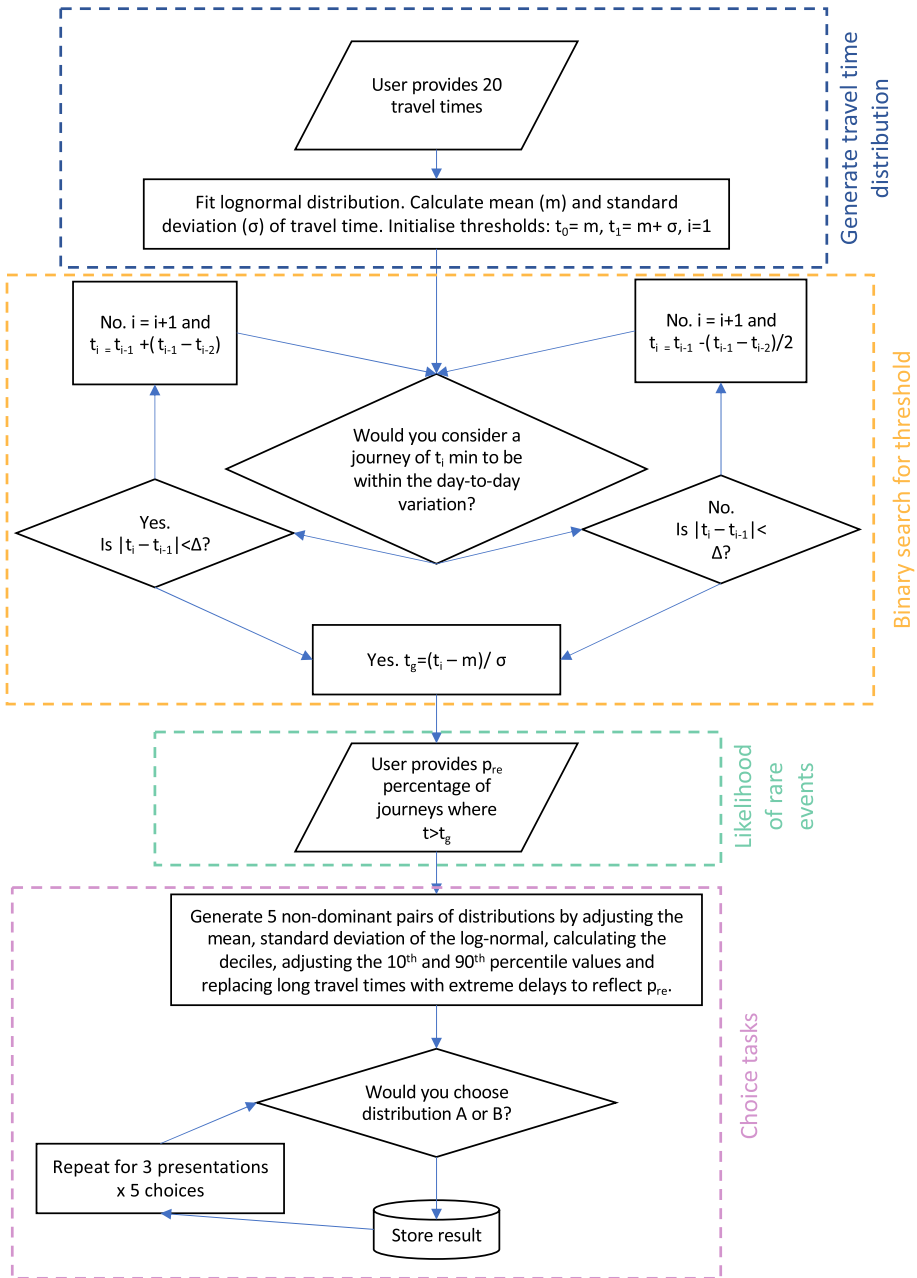


Fig. 1 Flow chart of the survey showing the 4 modules for establishing the travel time distribution, identifying the lateness threshold, querying the likelihood of rare events and presenting the 15 choice tasks

Consider your daily commute. Please enter the origin and destination addresses. Route options will appear on the map. Click to select your typical route on the map below:

If your exact route is not shown, tick this box and choose the closest option.

When do you typically make this trip?
 Check all that apply.

Monday Tuesday Wednesday Thursday Friday Saturday Sunday

Time of day for departure:

Hour	Minute	AM/PM
7	40	AM

How much do you pay in tolls to make this trip?

\$

(a) Map layout to determine the respondent's usual travel behaviour

If you were to make this trip 20 times, how often would you expect to experience each travel time?

16 min or less	
17-19 min	
20-22 min	2
23-25 min	10
26-28 min	4
29-30 min	2
31-33 min	
34-36 min	1
37-39 min	
40 min or more	1
Total times	20

Travel Time Range (min)	Frequency
16 min or less	0
17-19 min	0
20-22 min	2
23-25 min	10
26-28 min	4
29-30 min	2
31-33 min	0
34-36 min	1
37-39 min	0
40 min or more	1

(b) Interface for reconstructing a travel time distribution. Respondents enter numbers in the table and bars appear on the histogram.

Fig. 2 Selected screenshots of the survey instrument. **a** Map layout used to determine the respondent's usual travel behaviour. **b** Interface for reconstructing the travel time distribution from 20 trips. The respondent enters numbers in the table on the left and the orange bars appear on the histogram

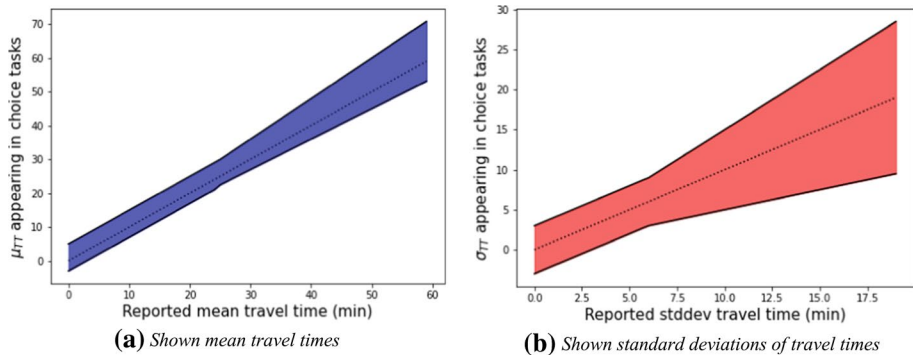


Fig. 5 Range of values that could appear in the presented alternatives based on reported mean and standard deviation of travel time

expected travel time, the standard deviation of travel time, the likelihood of experiencing a delay above the lateness threshold, the 10th percentile and the 90th percentile. To address dependencies between the attributes, the lateness threshold is normalised to the number of standard deviations above the mean. In general, relative pivots (using percentages around the reported reference levels) are applied, except when the reference travel times are short (mean travel times less than 25 min) and/or reliable (standard deviation less than 7 min) in which case absolute pivots are applied as illustrated in Fig. 5. Choice situations with dominant alternatives respective to mean and standard deviation of travel time are automatically removed based on the approach of Bliemer et al. (2017). This approach ensures that respondents can make trade-offs between travel time and travel time unreliability, however, it does not rule out choice situations where an alternative is *stochastically* dominant. An alternative is stochastically dominant if the cumulative distribution function of travel time is always above the other, see for example the choice situation in Figs. 6, 7 and 8. This type of choice situation (which occurred in our survey in 31% of the cases) is included to capture responses where people prefer a more stable travel time even if they have to accept a longer travel time. It is expected that most, but not all, people would select Trip B in the choice situation shown.

Presentation styles

Each of the five choice situations are presented in three ways to create 15 choice tasks per respondent. The three presentation styles are: (i) Colour-Coded Values, (ii) Descriptive Table, and (iii) Graphical Values and Risk. Since the underlying choice situations are identical, the impact of presentation style on choice behaviour can be compared. The order of the presentation styles and choice situations within each presentation style are randomised between respondents to control for learning and fatigue.

The Colour-Coded Values presentation style, shown in Fig. 6, uses a conventional format of sorted numerical values representing the nine deciles of the distribution and a rare event associated with that distribution. Although this style of presentation is common in the literature (Small et al. 1999; Tseng et al. 2008), it is often used with fewer (for example, five) values. Using a larger number of values allows us to better represent the width of the distribution, convey asymmetry and to succinctly communicate less-common events. To



Fig. 6 Colour-Coded Values presentation style with two sets of sorted and colour coded travel times representing the travel time distribution

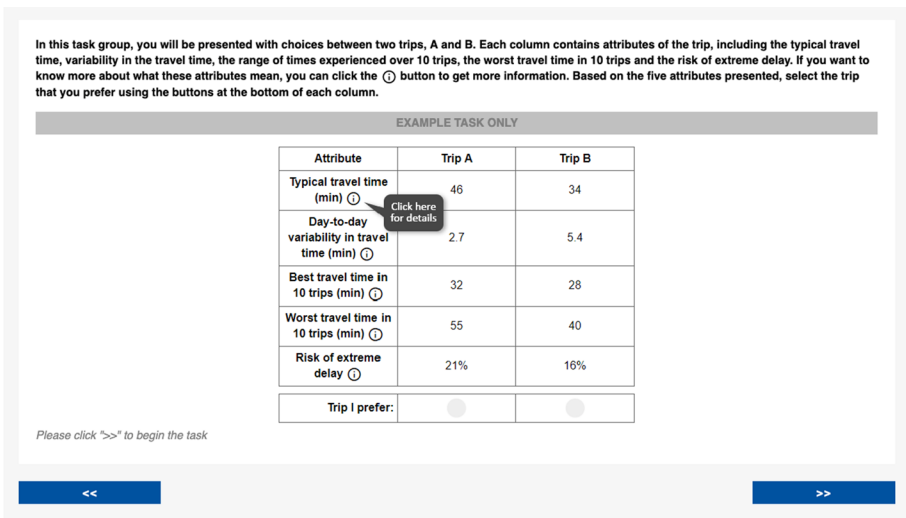


Fig. 7 Descriptive Table (DT) presentation style uses descriptive statistics of the travel time distribution

make comparing a larger number of values across two trips easier for the respondent, we used colour-coding to indicate low (green), medium (yellow), and high (red) travel time.

The second set of choices is presented using statistical measures of the travel time distribution in a Descriptive Table as shown in Fig. 7. As the respondents are not expected to be familiar with the statistical vocabulary, the labels have been rephrased and clicking on the information symbol gives more details. Unlike the other presentations, this layout provides no visual cues besides the numerical values. It also includes the standard deviation of the underlying travel time distribution which the respondents would not be able to recover

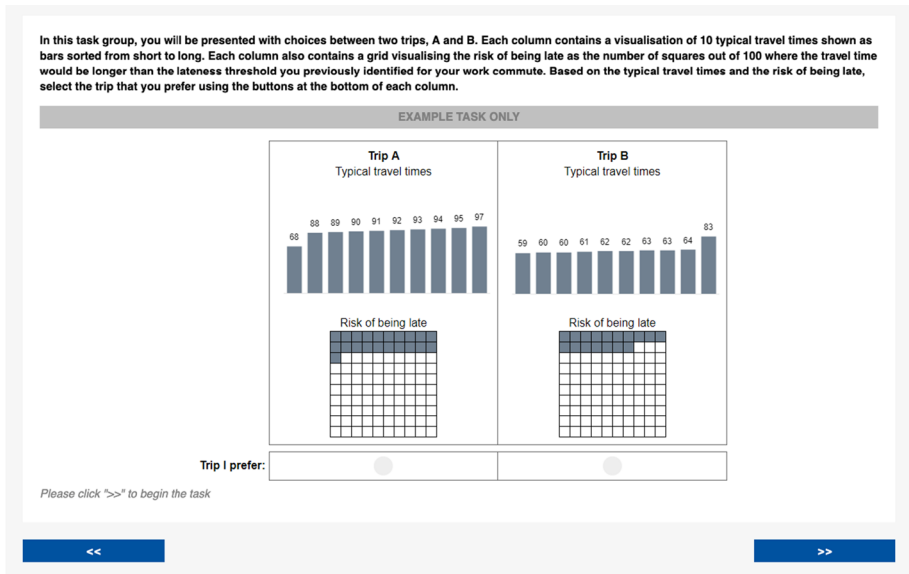


Fig. 8 Graphical Values and Risk presentation style uses a graphical representation to communicate the travel time distribution including the deciles (shown with bars) and rare events (shown with the grid)

precisely from the visual presentations. This presentation includes more information than either of the others, however it places a heavy cognitive load on the respondents who may be unfamiliar with either the vocabulary or the concepts used in this presentation.

The third set of choice tasks uses a Graphical Values and Risk presentation style based on the bar format described by Li et al. (2010), accompanied by a novel visualisation of risk of extreme delay as a number of shaded squares out of 100. The top half of this graphical presentation (Fig. 8) includes the same information as the Colour-Coded Values Presentation (Fig. 6). The bottom half gives information about the likelihood of a delay over the threshold which is unavailable in the first presentation. The grid presentation echoes the graphic from the question about the likelihood of rare events in Fig. 4, so the respondents are familiar with it.

Data analysis methods

In addition to analysing lateness thresholds, we aim to analyse the impact of presentation style. Given that the underlying choice situations of each presentation style are identical, we can directly compare reported difficulty and enjoyability of each presentation style as well as choice consistency. Further, we investigate the impact of presentation style on choice behaviour by estimating behavioural parameters in route choice models based on random utility and expected utility theory. Differences in the model parameters, especially regarding the inclusion of rare events in the choice situation, would indicate the importance of rare events in route choice and would suggest the simultaneous importance of multiple attributes of the travel time distribution.

Random utility model

For each presentation style, models based on random utility theory are estimated using attributes of the route alternatives, such mean travel time and standard deviation of travel time. The subjective utility of respondent n of route alternative j in choice situation s , denoted by U_{nsj} , is assumed to be the sum of systematic utility V_{nsj} and a random error term ε_{nsj} , where V_{nsj} is described by a linear function of attribute vector $\mathbf{x}_{nsj} = [x_{nsjk}]_{k=1,\dots,K}$,

$$V_{nsj} = \sum_{k=1}^K \beta_k x_{nsjk}, \quad (1)$$

where $\boldsymbol{\beta} = [\beta_k]_{k=1,\dots,K}$ is a vector of unknown (and to be estimated) behavioural parameters. We assume that ε_{nsj} are independently and identically extreme value type I distributed such that choice probabilities are described by a logit model.

There are three model specifications for attribute levels \mathbf{x}_{nsj} , namely *traditional*, *underlying* and *underlying +*. For the Colour-Coded Values and Graphical Values and Risk presentations styles only, models are estimated using the central tendency and standard deviation computed from the values presented to the respondent. These models, labelled *traditional*, provide a consistency check with, models are estimated using the central tendency and standard deviation computed from the values presented to the respondent. These models, labelled *traditional*, provide a consistency check with previous work. All three presentations also allow the estimation of models based on the underlying (designed) central tendency and standard deviation of travel time, which we label as *underlying*. Except for the Descriptive Table presentation style, the respondents are not presented with this information directly, but it is used in generating the attribute levels of the alternatives in the choice task. Finally, for each presentation style, additional models are estimated using more nuanced attributes of the travel time distributions which are labelled as *underlying +*. Of particular interest is the coefficient on the risk of extreme delay since it quantifies the impact of experiencing travel times above the lateness threshold.

Past work has demonstrated the importance of travel time reliability to travel choices through the reliability ratio. This describes the equivalence of travel time savings to a reduction in unreliability, and it is calculated as the ratio of the coefficient on the standard deviation of travel time, β_σ , to the coefficient on the central tendency travel time, β_μ . Central tendency is generally the mean of the travel time distribution, but in this work, we have calculated it with both the mean and median travel time. The *underlying +* models also allow the calculation of a risk ratio, which is the ratio of the coefficient of risk of extreme delay to the coefficient on the central tendency. This represents how the respondents are willing to exchange an increase in the expected travel time for a decrease in the risk of experiencing an extreme delay.

2.6 Expected utility model

Attitudes towards risk can be estimated using expected utility theory in the choice models for the presentation styles that show a travel time distribution (i.e., Colour-Coded Values, and Graphical Values and Risk). Rather than using central tendency and standard deviation directly as attributes in the utility function, expected utility theory considers a transformation of travel times in the distribution to reflect perceived risk.

To derive the expected utility function, let V_{nsj}^m be the utility of respondent n in choice situation s for alternative j when a positive outcome of a single attribute with level $m \in \{1, \dots, M\}$ (e.g., a certain prize in a lottery) is realised, and let p_{nsj}^m denote the probability of this outcome. Then the expected utility is.

$$E(V_{nsj}) = \sum_{m=1}^M p_{nsj}^m V_{nsj}^m. \quad (2)$$

If utility increases linearly with the outcome of the single attribute, i.e., if $V_{nsj}^m = \beta x_{nsj}^m$ with $\beta > 0$, then one implicitly assumes risk neutrality. Assuming constant relative risk aversion (CRRA), utilities can be written as

$$V_{nsj}^m = \begin{cases} \frac{(\beta x_{nsj}^m)^{1-\alpha}}{1-\alpha}, & \text{if } \alpha \neq 1; \\ \ln(\beta x_{nsj}^m), & \text{if } \alpha = 1, \end{cases} \quad (3)$$

where α represents the attitude towards risk. If $\alpha = 0$ then decisions are made in a risk-neutral fashion. If $\alpha > 0$ then individuals are risk-averse, which means that people would rather receive \$5 with certainty than gambling between \$0 and \$10 with equal probability, i.e., the expected utility for \$5 is larger than expected utility for making the gamble between \$0 and \$10. Similarly, if $\alpha < 0$ then individuals are risk-seeking. In our case, x_{nsj}^m represent levels of travel time, $m = 1, \dots, 10$, with probabilities $p_{nsj}^m = \frac{1}{10}$. Since utility decreases with increasing travel times, we need to consider negative utilities, $-\sum_{m=1}^M V_{nsj}^m$ and the interpretation of α reverses, namely $\alpha > 0$ refers to risk-seeking and $\alpha < 0$ to risk-averse. Assuming $\alpha \neq 1$ this results in the following expected utility under CRRA:

$$E(V_{nsj}) = - \sum_{m=1}^M p_{nsj}^m \frac{(\beta x_{nsj}^m)^{1-\alpha}}{1-\alpha}, \quad (4)$$

Note that other formulations of risk aversion have appeared in the literature, such as in Li and Hensher (2020) where coefficient β is taken outside the transformation:

$$E(V_{nsj}) = -\beta \sum_{m=1}^M p_{nsj}^m \frac{(x_{nsj}^m)^{1-\alpha}}{1-\alpha}, \quad (5)$$

To maintain comparability of travel time coefficient β with the random utility models, we use expected utility formulation (4). It should be noted that the travel time coefficient β in our expected utility model should be interpreted as being negative instead of positive.

Results

Survey sample

The survey received 1001 responses with a 20% incidence rate. Responses with unusually long (greater than 3 h 20 min) response times were removed. Irreconcilable mismatches between respondent-reported travel times and typical travel times estimated by Google Maps were also removed, resulting in a final sample size of 914 respondents. The typical travel time represents the Google Maps free-flow travel time used while the participant

was taking the survey, and real-time travel time predictions from the correct time of day were subsequently collected for validating the results. A mismatch is defined as those with reported travel times more than three times longer than the Google Maps free-flow estimate or less than half of the Google Maps free-flow estimate, provided that the mismatch is greater than 10 min in both cases.

The final sample is representative of the population with respect to age (except for minors), gender and income where the distribution shows a peak at the NSW median household income of around AUD 77,000 (Australian Bureau of Statistics 2016), see Fig. 9a–c. In New South Wales, 65% of residents commute to work by car either as the driver or the passenger (Australian Bureau of Statistics 2016). Approximately two thirds of the respondents report using technology to navigate at least sometimes, indicating that route choices may be made accounting for real-time conditions rather than memories of past experiences alone. Fewer older drivers use technology (Google Maps, Waze, etc.) to navigate than younger drivers, see Fig. 9a.

The distributions of reported, free-flow and real-time travel times are shown in Fig. 9d. The distribution of reported travel times is similar to the distributions of reported *duration* and *duration in traffic* from the Google Maps API. The distribution of reported typical travel times shows the expected asymmetry with a long tail of extremely long commuting times going out to 300 min (censored in the figure). There is some inconsistency between the reported typical travel time and the Google Maps free-flow time, partly due to imprecise origin and destination addresses. Many respondents only provided a suburb for origin and destination. This was allowed because restricting to street addresses prevents those working in named places (hospitals, office parks, universities, etc.) from using

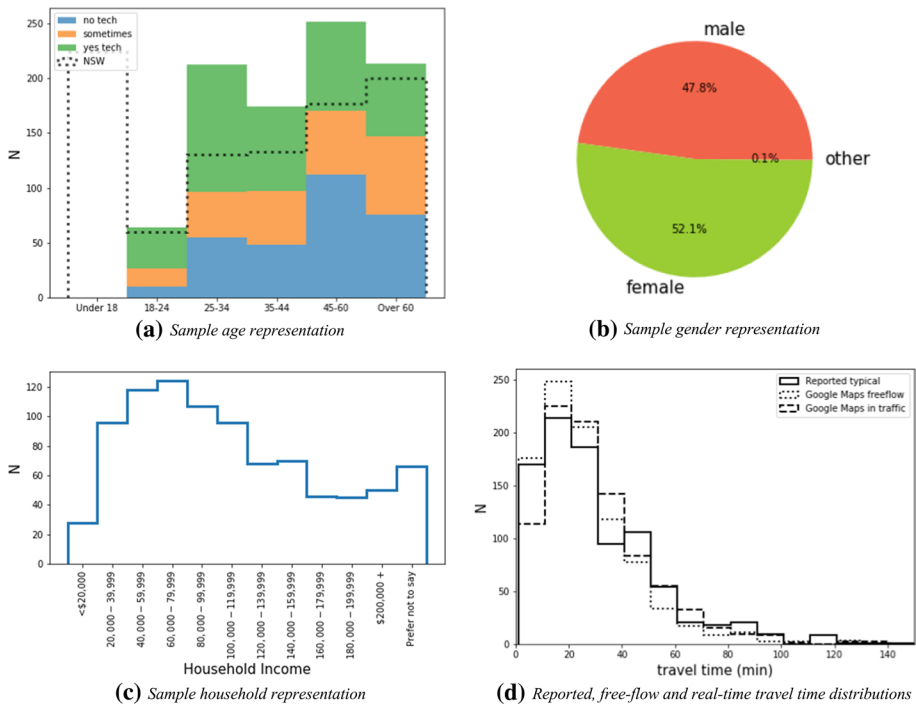


Fig. 9 Attributes of the respondents

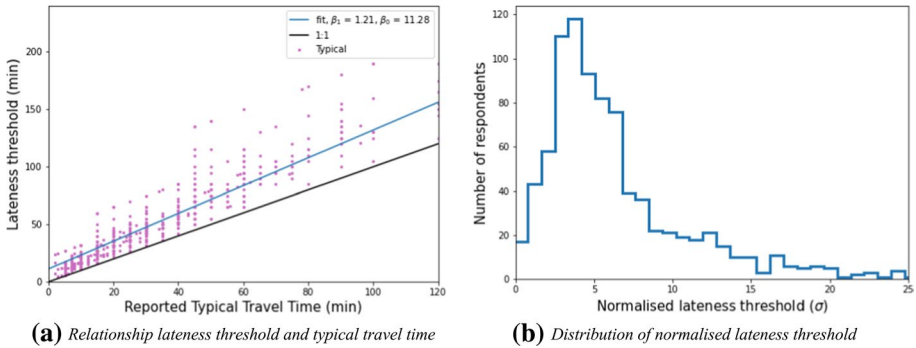


Fig. 10 Respondent-reported thresholds of lateness

point-of-interest (POI) information in the maps database. This granularity can make Google Maps and respondent-reported travel times differ, especially for shorter trips.

With respect to observations in the choice experiment, in choice situations containing a stochastically dominant alternative, most respondents selected the trip that had strictly faster travel times, but 86 (6%) chose the option with slower, more stable travel times.

Attitudes towards risk

Thresholds of lateness were ascertained using binary search and compared to the reported typical travel times in Fig. 10. Each respondent's threshold is normalised by subtracting the mean and dividing by the standard deviation of travel time from their reported distribution. The normalised threshold is a measure of the risk sensitivity of the respondents.

Overall, the reported thresholds are 21% larger than the typical travel times plus an 11 min buffer, which suggests that lateness for short trips is dominated by an absolute margin (11 min) but for long trips the acceptable delay increases. The high values (the peak of the distribution is around 5 standard deviations above the typical travel time) suggests that many respondents are comfortable with a high degree of variability in their travel times. This plot does not show the tail of the histogram with several respondents with thresholds that are up to 100 standard deviations above the typical travel time—it is not expected that a commuter would observe an event this rare in their lifetime.

The attitude towards risk captured in the threshold is an important factor in understanding the respondents' value of reliability in the models. It shows heterogeneity in the population that is not well explained by basic demographic information alone, as shown in Fig. 11. Respondents over 45 report both the smallest and largest thresholds in the sample, but the average threshold is similar for all age groups. There is no clear trend between income and normalised threshold.

Risk acceptance in travel times might also be associated with inattention to travel times. Comparing the reported typical travel time to the mean of many time-of-day specific results from Google Maps API's *duration in traffic* gives a measure of how well the respondents know their own commute. Outliers have already been filtered as described above, and the remaining responses show both under and overestimation. The normalised threshold is based on the mean and standard deviation of the 20 reported travel times. As shown in Fig. 12 there is a weak but statistically significant correlation between inaccuracy

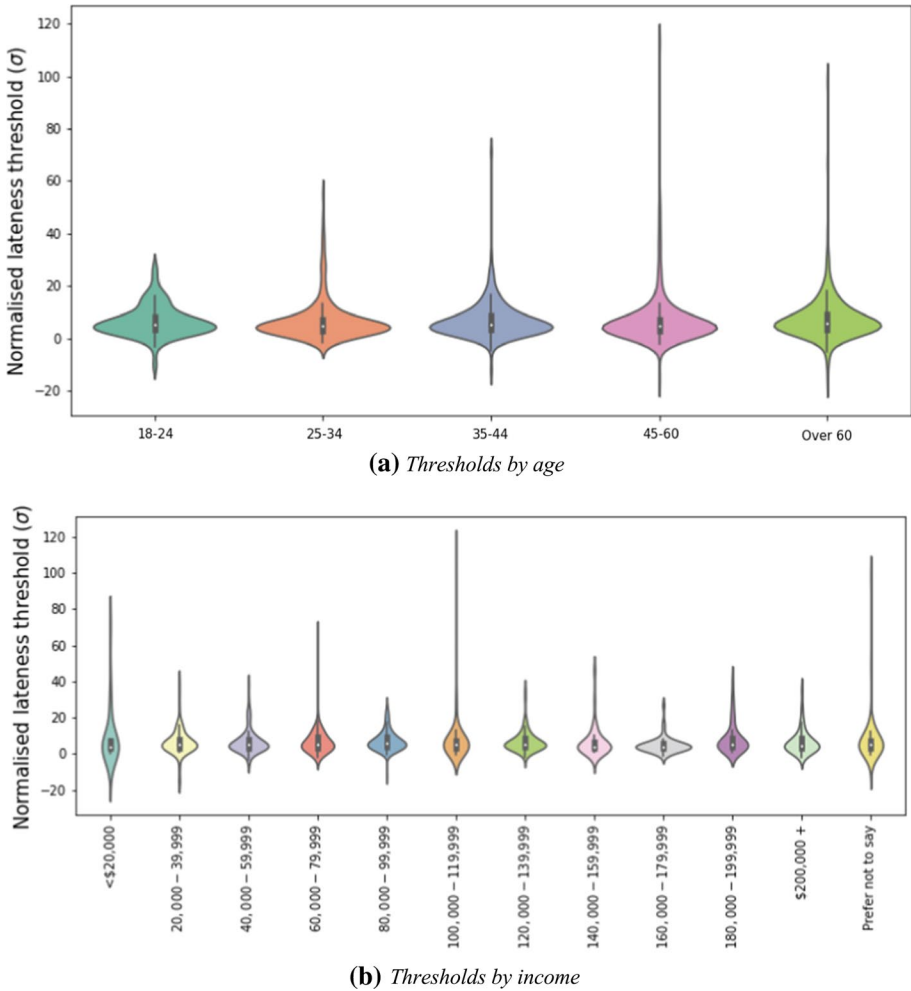


Fig. 11 The relationship between reported lateness threshold and key demographic attributes

and threshold where respondents with lower normalised lateness thresholds report inaccurate understandings of their typical travel time.

Difficulty and enjoyability of presentation styles

In order to test for fatigue and confusion, after each set of choice tasks, the respondents were asked how difficult and enjoyable that task group was. They were asked to rate on a 7 point Likert scale, ranging from easy/simple (1) to difficult/confusing (7). Similarly, they rated their enjoyment from fun/enjoyable (1) to boring/frustrating (7). Overall, the respondents reported that the tasks were easy, see Fig. 13a, and neither fun nor boring, see Fig. 13b. Unexpectedly, the respondents did not report the Descriptive Table presentation style to be more difficult than the other tasks. Comparing the ordinal results from the Likert questions for the Colour-Coded Values presentation style versus Descriptive Table, and the

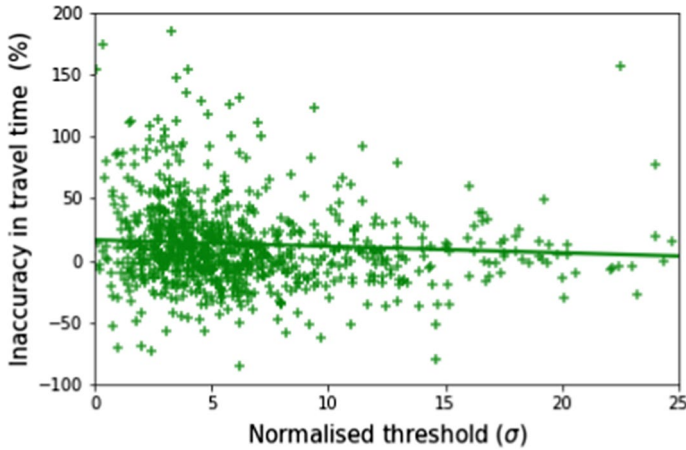


Fig. 12 Inaccuracy in reported travel time compared to normalised lateness threshold. The negative relationship is weak but statistically significant (95% confidence bands on the coefficient are plotted but not visible)

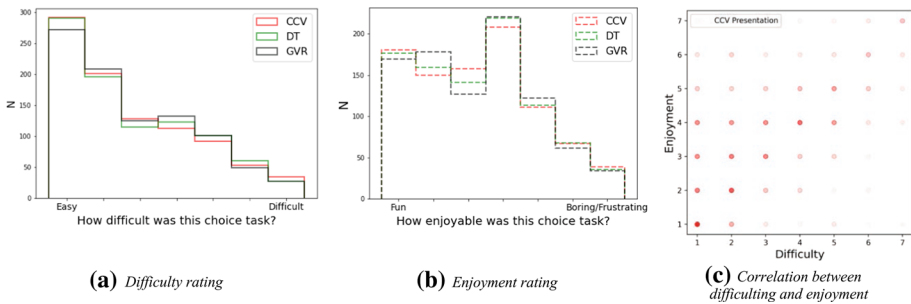


Fig. 13 Respondents view of the difficulty and enjoyability of the three presentations. CCV = Colour-Coded Values, DT=Description Table, GVR=Graphical Values and Risk

Graphical Values and Risk presentation style versus Descriptive Table with a Mann–Whitney U test (Mann and Whitney 1947) yields p values of 0.498 and 0.783, respectively for difficulty; and 0.945 and 0.921 respectively for enjoyment—in all cases we cannot reject the null hypothesis that the scores for the Descriptive Table presentation style are drawn from the same distribution as the Colour-Coded Values or Graphical Values and Risk presentation styles. There is correlation between easiness and enjoyment as shown in Fig. 13c for the Colour-Coded Values presentation style (the same correlation was present for all three presentations), meaning that respondents who found the tasks easy were more likely to enjoy them.

Choice consistency across presentation styles

Since each respondent completed the same five choice tasks 3 times, their responses can be tested for consistency across the presentation styles. On the one hand, self-consistency is evidence that the participants are not selecting alternatives at random. On the other hand,

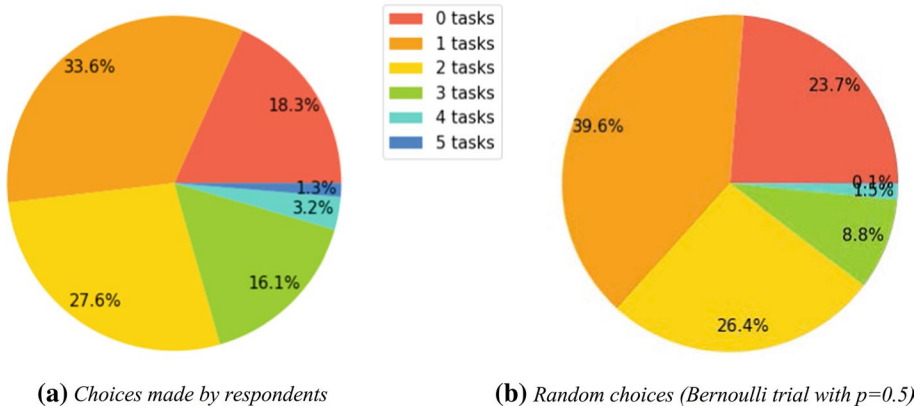


Fig. 14 Number of choice situations with consistency across presentation styles

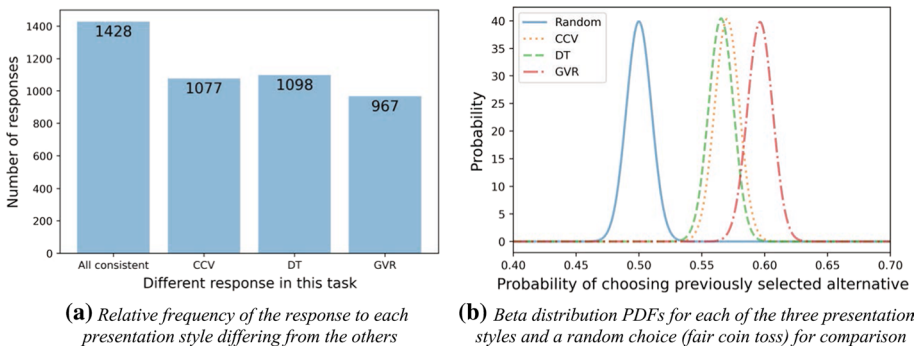


Fig. 15 Comparison of consistency across presentation styles. CCV = Colour-Coded Values, DT = Description Table, GVR = Graphical Values and Risk

lack of self-consistency can indicate that differences in the presentation styles result in participants exhibiting different behaviours. The following analysis explores the presence of these traits in the sample.

Only 12 of the 914 respondents (1.3%) chose the same alternative in all three presentation styles for all five choice tasks as shown in Fig. 14a, and nearly 82% of the sample chose consistently for at least one of the five tasks. A comparison with random choices shows that the responses overrepresent consistency between 2, 3, 4 or 5 tasks and underrepresent consistency occurring in 1 or 0 tasks. This means that the frequency of consistency is higher than what would be expected from random choices.

Inconsistency or randomness indicates that the information needed to make a non-random choice was either missing or unable to be understood from that presentation. The breakdown of choice patterns between the three presentation styles is shown in Fig. 15a showing that consistency between the presentation styles is the most common pattern. Each respondent completes 5 choice tasks, and the pattern is evaluated for the individual task. If the respondents choose randomly for each task, we expect the four bars to have the same height. If the information used to make the choice is available in all three presentations, all the responses would be in the ‘All consistent’ category.

Consistency between any two presentations can be derived from Fig. 15a. For example, the number of choices with consistency between the Descriptive Table and Graphical Values and Risk presentation styles is the sum of the height of the column for Colour-Coded Values (i.e., the number of choice tasks where the respondent chose the same alternative in the Descriptive Table and the Graphical Values and Risk but a different alternative in the Colour-Coded Values) and the height of the column labelled All consistent (i.e., the number of choice tasks where the respondent chose the same alternative for all three presentations). The plot shows that Colour-Coded Values and Graphical Values and Risk presentation styles are the most consistent, followed closely by Descriptive Table and Graphical Values and Risk, while Colour-Coded Values and Descriptive Table presentation styles are the least consistent (although not much less than the other pairings).

To understand the statistical significance of the difference between the column heights in Fig. 15a, we conceptualise the survey as a randomised trial of three presentation styles. For each respondent, we identify the two presentations that yielded the same choice (since there are two alternatives and three presentations, there must be at least two that are consistent with each other), and the choice for the last presentation is between the alternative that is selected in the other two presentations or the other alternative. Analogous to a (possibly biased) coin flip, this choice is then a Bernoulli trial with the possible outcomes of *same* or *different*. The Beta distribution describes the probability distribution of the Bernoulli process parameter which varies between 0 and 1 based on the observed outcomes. A random choice (a fair coin flip) has an expected parameter value of 0.5.

In this context, the Bernoulli process is the outcome of the 3rd presentation layout for any task where the respondent made a consistent choice in the other two layouts. The success rates are calculated as the fraction of the time that the consistent choice is made—the height of the all-consistent bar in Fig. 15a divided by that value plus the height of the relevant bar. The resulting probability distributions are shown in Fig. 15b where all the presentations produce outcomes that differ from the random (fair coin) choice as evidenced by the lack of overlap with the blue curve. The Graphical Values and Risk presentation style is the least random with a probability of 0.596 of choosing the alternative selected in the other presentations, but the probabilities for the Colour-Coded Values and Graphical Values and Risk presentation styles (0.570 and 0.565 respectively) suggest that these presentations convey similar information to the respondents. Given that all choices, experimental or real-life, are made with imperfect information, it is expected that a successful presentation offers a less-than-perfect improvement over randomness. All three presentations are distinct from the random-choice outcome, which means that each conveys some information to the respondent that is used to make the choice. The Graphical Values and Risk presentation style offers a small improvement over either the Colour-Coded Values or the Descriptive Table.

Model estimation results

Parameters estimates and standard errors (s.e.) for the three random-utility models (i.e., traditional, underlying, and underlying+) are presented in Table 1, where we consider models with either mean travel time or median travel time as central tendency.

Parameter estimates for the mean, median, and standard deviation of travel time are statistically significant and have the expected sign in all models. Resulting reliability ratios for each presentation style, including their 95% confidence interval, are shown in Fig. 16. Reliability ratios reported in the literature vary between 0.1 (Batley and Ibanez 2009) and

Table 1 Estimated coefficients and their standard errors for each of the model specifications and each presentation style

	Colour-coded values			Description table			Graphical values and risk					
	β	s.e		β	s.e		β	s.e				
<i>Traditional</i>												
Median ^a	-	-	-0.112	0.005	-	-	-	-	-0.069	0.004		
Mean ^a	-0.116	0.005	-	-	-0.180	0.008	-	-0.073	0.004	-		
Std dev ^a	-0.063	0.010	-0.084	0.010	-0.097	0.013	-0.153	0.016	-0.049	0.009		
Reliab. ratio	0.542		0.747		0.538		0.837		0.670	0.870		
AIC	5557		5587		5107		5113		5970	5998		
<i>Underlying</i>												
Median ^b	-	-	-0.149	0.007	-	-	-0.183	0.008	-	-0.098	0.006	
Mean ^b	-0.144	0.007	-	-	-0.180	0.008	-	-	-0.096	0.006	-	
Std dev ^b	-0.117	0.014	-0.165	0.015	-0.097	0.013	-0.153	0.016	-0.098	0.013	0.014	
Reliab. ratio	0.812		1.106		0.538		0.837		1.020	1.289		
AIC	5563		5546		5107		5113		5964	5965		
<i>Underlying +</i>												
Best	-0.014	0.006	-0.016	0.006	-0.033	0.006	-0.034	0.006	-0.017	0.005	0.005	
Median ^b	-	-	-0.144	0.009	-	-	-0.152	0.010	-	-	-0.089	0.008
Mean ^b	-0.146	0.009	-	-	-0.150	0.010	-	-	-0.090	0.008	-	
risk	-0.803	0.672	-0.812	0.671	-0.139	0.008	-0.140	0.008	-0.192	0.009	0.009	
Std dev ^b	-0.039	0.017	-0.079	0.019	-0.117	0.016	-0.166	0.019	-0.071	0.015	0.016	
Worst	-0.036	0.002	-0.035	0.002	-0.016	0.005	-0.016	0.005	-0.019	0.002	0.002	
Risk ratio	5.510		5.639		0.925		0.917		2.138	2.161		
Reliab. ratio	0.266		0.552		0.782		1.087		0.790	1.111		
AIC	5044		5053		4715		4717		5141	5149		

^aOf presented values, ^bOf underlying distribution

2.5 (Small et al. 1999), as summarised by Carrion and Levinson (2012a,b). The reliability ratios from the 16 models estimated in this work sit within the range reported in the literature and also indicated in Fig. 16 (de Jong et al. 2007; Asensio and Matas 2008; Tilahun and Levinson 2010; Li et al. 2010; Carrion and Levinson 2012b; Kouwenhoven et al. 2014; Leahy et al. 2016). Previous studies tend to show fewer travel time values in their choice tasks (for example, five in Small et al. (1999) compared to 10 here), which gives less information about the travel time distribution.

The reliability ratios in the models using the underlying distributions yield higher reliability ratios than those estimated from traditional models based on the presented values. A possible explanation is that respondents overestimate the standard deviation from the presented values, resulting in a lower coefficient in the traditional model (approximately half as big when comparing the top and middle sections of Table 1) and a smaller reliability ratio.

Models with additional information (underlying+) models, shown in the bottom section of Table 1, produce reliability ratios that are similar to those found in the simpler models, see also Fig. 16. However, for those presentations that provided information about the extreme delays (Descriptive Table, Graphical Values and Risk), risk is a statistically significant factor in route choice *in addition to* the standard deviation of travel time, and these models have better goodness of fit based on Akaike's Information Criteria (AIC).

The reported risk ratios at the end of Table 1 and in Fig. 17 suggest that respondents would exchange a 0.9–2.0% reduction in risk of extreme delay for one minute reduction in typical travel time. For the Colour-Coded Values presentation style, where no explicit information was provided about the risk of extreme delay and only 10 travel times were presented, the coefficient on risk is not significant as expected, so the large error bars in Fig. 17 illustrate a valuable null result. This suggests that respondents value lower risks of extreme delay where that information is available, although such information may not be available or perceived in practice and therefore explicitly showing this information in a stated choice experiment may make it more salient than it is in reality.

With respect to the expected utility model, results show statistically significant travel time coefficients of 0.20 and 0.13 for the Colour-Coded Values and Graphical Values and Risk presentation styles, respectively, which are similar to the negative of the coefficients for mean and median travel time found in all three sections of Table 1. The associated values for risk attitude, α , are respectively 0.12 and 0.21, indicating slight risk-seeking attitudes on average. This is consistent with the high thresholds of delay, but it contradicts the negative coefficient on standard deviation of travel time and risk of extreme delay estimated in the choice models. Past authors (Wijayarathna and Dixit 2016; Li and Hensher 2020 and references therein) have also found risk seeking behaviour. This finding may reflect a distribution of risk attitudes within the population or the attitude that variable travel times are an opportunity to be faster-than-typical more than a chance to be delayed. We see evidence of this asymmetry in the difference between the coefficients for the best and worst travel times for the Colour-Coded Values presentation style in the underlying+ model presented at the bottom of Table 1.

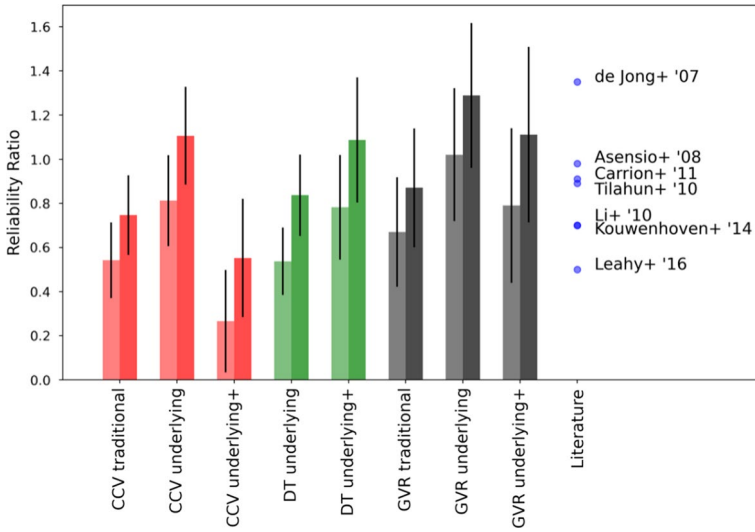
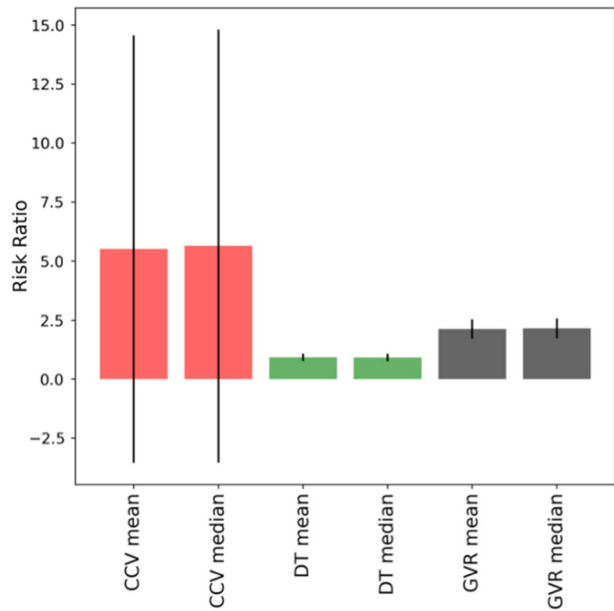


Fig. 16 Reliability ratios and their 95% confidence intervals with the central tendency represented by mean (lighter shaded bar on the left) and median (darker shaded bar on the right) for each presentation style and specification shown. CCV = Colour-Coded Values, DT = Description Table, GVR = Graphical Values and Risk

Fig. 17 Risk ratios calculated from the 3 layouts each using the mean (left) and median (right) as the central tendency. The error bars show the 95% confidence interval ($1.96\sigma_{risk}$ ratio) on the estimated coefficient



Discussion

Value of time studies often assign different values by trip purpose, mode and journey segment as summarised by Wardman (2004). Those models recognise that time spent in waiting for a bus is different to time spent driving, yet the commonly used reliability framework simplifies the value on variability to be simply one aspect of the travel time distribution. The work above follows the path of value of time studies to explore a more nuanced valuation of reliability. Specifically, the model results demonstrate that travellers value day-to-day variation independently from extreme delays in route choice decisions. The results show that explicitly valuing rare events gives a better representation of the full benefit of an alternative. The models that include the risk parameter result in better goodness of fit, reliability ratios consistent with the literature *as well as* a statistically significant risk ratio.

The results presented above also explore the importance of presentation style in a stated-choice experiment. One outcome from this analysis is that using mean or median travel time will provide consistent estimates for the reliability ratio—in every model shown in Fig. 16, 95% confidence level from the model using mean (light shaded on left) overlaps with the same confidence level for the model using median (darker shaded on right). Furthermore, the three presentation styles result in reliability ratios within the 95% confidence interval for each specification (traditional, underlying, underlying+). These findings de-emphasise the importance of presentation style for the reliability ratio. However, the results shown in Fig. 17 show that rare-event reliability is valued only when the presentation style makes it explicitly possible—this explains why the risk coefficient is not significant in the Colour-Coded Values presentation style. Moreover, risk is valued less in the Descriptive Table presentation style (where it is one of 5 attributes in a table) compared to the Graphical Values and Risk presentation style (where it is one of two graphics). Together, these findings indicate that the selection of *what* is included in the presentation is more important than *how* it is displayed.

In the transport context, these outcomes have implications for the application of value-of-reliability measures to hypothetical scenarios. One might argue that the current framework captures the likelihood of extreme delays through its typical relationship to the standard deviation of travel time, but in new contexts, this relationship might change. For example, the adoption of autonomous vehicles may result in changes in the shape of the travel time distribution such that existing relationships between the prevalence of rare events and variance of travel time change. An autonomous vehicle system may be extremely reliable day-to-day, but the system will perform extremely poorly in the rare cases that it does fail because it is operating closer to capacity. Alternatively, the improved safety of autonomous vehicles may reduce the occurrence of delays due to crashes, which are non-recurrent or possibly rare events. In this scenario, recurrent congestion might get worse due to increased car travel because drivers can do other tasks while driving. In either scenario, these changes may motivate multiple measures for the value of reliability that account for different attributes of the travel time distribution.

Although assisting the literature to put further focus on rare events, one limitation of this study is a narrow definition of rare events. The analysis focuses only on extreme travel times without considering other, sometimes-concomitant rare occurrences such as unavailability of parking, crowding on public transport, cancellation of transit services or being directly involved in a vehicle crash or breakdown. These broader definitions should be explored in future work.

The first section of this paper references literature on how episodic and semantic memory contribute to decision making. It remains ambiguous which category is appropriate for a regular commute. If someone makes the same commute for years and most recently yesterday that may be stored in either long or short interval retention. Specifically referring to rare events, if an extreme delay occurred within a few weeks of the respondent participating in the survey, that trip is episodic and the peak-end rule is applicable following the findings of Geng et al. (2013). If the rare event was more than a few weeks before the survey participation, this would contribute to semantic memory and its importance in the decision could be different. Time elapsed since the last extreme delay could contribute to the mild risk-seeking attitude measured with the Expected Utility Theory models. The frequency of the trip and the socio-demographic attributes of the respondents, which are both acquired in the survey, may play a role in determining which memory mechanism is dominant. This will be explored in future work.

Previous work has highlighted the potential bias introduced eliciting choice information about rare events (De Palma et al. 2014). The results presented above reinforce this finding where the influence of rare event reliability is most significant where it is presented most explicitly (Descriptive Table) even if that presentation is expected to have the highest cognitive load. We can better understand the true value on rare events by comparing the decision-from-description choices presented here with decision-from-experience choices from a matching revealed preference survey. Building on the discussion of the time since the extreme delay, the revealed preference design will ascertain if the travellers' semantic knowledge of the tail of the travel time distribution contributes significantly to their route choice. This will be explored in future work.

Author contribution All authors contributed to the study conception and design. Multinomial logit model estimations were led by EM and THR. The estimation of the expected utility models was led by MB and EM. The first draft of the manuscript was written by EM and all authors commented on subsequent versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Asensio, J., Matas, A.: Commuters' valuation of travel time variability. *Trans Res Part E: Logist Transport Rev* **44**(6), 1074–1085 (2008). <https://doi.org/10.1016/j.tre.2007.12.002>
- Australian Bureau of Statistics. 2016. "Census QuickStats: Greater Sydney." 2016. https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/1GSYD?opendocument.
- Barsky, R.B., Thomas Juster, F., Kimball, M.S., Shapiro, M.D.: Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement study. *Q. J. Econ.* **112**(2), 537–579 (1997). <https://doi.org/10.1162/003355397555280>
- Batley, Richard, and J.N. Ibanez. 2009. "Randomness in Preferences, Outcomes and Tastes; an Application to Journey Time Risk." In *Proceedings of the International Choice Modelling Conference*. Harrogate.
- Beaud, M., Blayac, T., Stéphan, M.: The impact of travel time variability and travelers' risk attitudes on the values of time and reliability. *Transport Res Part B: Methodol* **93**, 207–224 (2016). <https://doi.org/10.1016/j.trb.2016.07.007>
- de Bekker-Grob, E.W., Bliemer, M.C.J., Donkers, B., Essink-Bot, M.-L., Korfage, I.J., Roobol, M.J., Bangma, C.H., Steyerberg, E.W.: Patients' and urologists' preferences for prostate cancer treatment:

- a discrete choice experiment. *Br. J. Cancer* **109**(3), 633–640 (2013). <https://doi.org/10.1038/bjc.2013.370>
- de Bekker-Grob, E.W., Rose, J.M., Bliemer, M.C.J.: A closer look at decision and analyst error by including nonlinearities in discrete choice models: implications on willingness-to-pay estimates derived from discrete choice data in healthcare. *Pharmacoeconomics* **31**(12), 1169–1183 (2013). <https://doi.org/10.1007/s40273-013-0100-3>
- Bliemer, M.C.J., Rose, J.M., Chorus, C.G.: Detecting dominance in stated choice data and accounting for dominance-based scale differences in logit models. *Trans Res Part B: Methodological* **102**, 83–104 (2017). <https://doi.org/10.1016/j.trb.2017.05.005>
- Carrion, C., Levinson, D.: Value of travel time reliability: a review of current evidence. *Trans Res Part A: Policy Practice* **46**(4), 720–741 (2012a). <https://doi.org/10.1016/j.tra.2012.01.003>
- Carrion, C., Levinson, D.M.: A Model of Bridge Choice Across the Mississippi River in Minneapolis. In: Levinson, D.M., Liu, H.X., Bell, M. (eds.) *Network Reliability in Practice*, Transportation Research, Economics and Policy. Springer, New York (2012b)
- Chen, A., Ji, Z., Recker, W.: Travel time reliability with risk-sensitive travelers. *Trans Res Record: J Transport Res Board* **1783**, 27–33 (2002). <https://doi.org/10.3141/1783-04>
- Chen, A., Zhou, Z.: The α -reliable mean-excess traffic equilibrium model with stochastic travel times. *Transport Res Part B: Methodological* **44**(4), 493–513 (2010). <https://doi.org/10.1016/j.trb.2009.11.003>
- Palma, De., André, M.A., Attanasi, G., Ben-Akiva, M., Erev, I., Fehr-Duda, H., Fok, D., et al.: Beware of black swans: taking stock of the description-experience gap in decision under uncertainty. *Mark. Lett.* **25**(3), 269–280 (2014). <https://doi.org/10.1007/s11002-014-9316-z>
- Determann, D., Korfage, I.J., Lambooi, M.S., Bliemer, M., Richardus, J.H., Steyerberg, E.W., de Bekker-Grob, E.W.: Acceptance of vaccinations in pandemic outbreaks: a discrete choice experiment. *PLoS ONE* **9**(7), e102505 (2014). <https://doi.org/10.1371/journal.pone.0102505>
- Drottz-Sjöberg, Britt-Marie. 1991. *Perception of Risk: Studies of Risk Attitudes, Perceptions and Definitions*. Perception of Risk: Studies of Risk Attitudes, Perceptions and Definitions. Stockholm, Sweden: Stockholm School of Economics.
- Geng, X., Chen, Z., Lam, W., Zheng, Q.: Hedonic evaluation over short and long retention intervals: the mechanism of the peak-end rule. *J. Behav. Decis. Mak.* **26**(3), 225–236 (2013). <https://doi.org/10.1002/bdm.1755>
- Hensher, D.A., Li, Z., Rose, J.M.: Accommodating risk in the valuation of expected travel time savings. *J. Adv. Transp.* **47**(2), 206–224 (2013). <https://doi.org/10.1002/atr.160>
- Ida, T., Goto, R.: Simultaneous measurement of time and risk preferences: stated preference discrete choice modeling analysis depending on smoking behavior*. *Int. Econ. Rev.* **50**(4), 1169–1182 (2009). <https://doi.org/10.1111/j.1468-2354.2009.00564.x>
- de Jong, G.C., Bliemer, M.C.J.: On including travel time reliability of road traffic in appraisal. *Transport Res Part A: Policy and Practice* **73**, 80–95 (2015). <https://doi.org/10.1016/j.tra.2015.01.006>
- G De Jong, Y Tseng, M Kouwenhoven, E Verhoef, J Bates 2007 The value of travel time and travel time reliability Public Works and Water Management Ministry of Transport, The Netherlands
- Kahneman, D., Fredrickson, B.L., Schreiber, C.A., Redelmeier, D.A.: When more pain is preferred to less: adding a better end. *Psychol. Sci.* **4**(6), 401–405 (1993). <https://doi.org/10.1111/j.1467-9280.1993.tb00589.x>
- Kato, T., Uchida, K., Lam, W.H.K., Sumalee, A.: Estimation of the value of travel time and of travel time reliability for heterogeneous drivers in a road network. *Transportation*, April. (2020). <https://doi.org/10.1007/s11116-020-10107-x>
- Knuth, Donald E. 1998. *The Art of Computer Programming: Volume 3: Sorting and Searching*. Addison-Wesley Professional.
- Kouwenhoven, M., de Jong, G.C., Koster, P., van den Berg, V.A.C., Verhoef, E.T., Bates, J., Warffemius, P.M.J.: New values of time and reliability in passenger transport in the Netherlands. *Res Transport Eco, App Trans* **47**, 37–49 (2014). <https://doi.org/10.1016/j.retrec.2014.09.017>
- Leahy, C., Batley, R., Chen, H.: Toward an automated methodology for the valuation of reliability. *J Intell Transport Sys* **20**(4), 334–344 (2016). <https://doi.org/10.1080/15472450.2015.1091736>
- Li, Z., Hensher, D.: Understanding risky choice behaviour with travel time variability: a review of recent empirical contributions of alternative behavioural theories. *Transport Lett* **12**(8), 580–590 (2020). <https://doi.org/10.1080/19427867.2019.1662562>
- Li, Z., Hensher, D.A., Rose, J.M.: Willingness to pay for travel time reliability in passenger transport: a review and some new empirical evidence. *Transport Res Part E: Logist Transport Rev* **46**(3), 384–403 (2010). <https://doi.org/10.1016/j.tre.2009.12.005>

- van Lint, J.W.C., van Zuylen, H.J., Tu, H.: Travel time unreliability on freeways: why measures based on variance tell only half the story. *Transport Res Part A: Policy Practice* **42**(1), 258–277 (2008). <https://doi.org/10.1016/j.tra.2007.08.008>
- Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **18**(1), 50–60 (1947)
- Moylan, Emily, Michiel Bliemer, and Taha Hossein Rashidi. 2019. “The Impact of Rare Events on Valuations of Reliability.” In *International Choice Modelling Conference 2019*. <http://www.icmconference.org.uk/index.php/icmc/ICMC2019/paper/view/1706>.
- Small, Kenneth A., Robert B. Noland, Xuehao Chu, and David Lewis. 1999. *Valuation of Travel-Time Savings and Predictability in Congested Conditions for Highway User-Cost Estimation*. National Cooperative Highway Research Program, Report 431. Transportation Research Board.
- Soza-Parra, J., Raveau, S., Muñoz, J.C.: Public transport reliability across preferences, modes, and space. *Transportation* (2021). <https://doi.org/10.1007/s11116-021-10188-2>
- Soza-Parra, J., Raveau, S., Muñoz, J.C., Cats, O.: The underlying effect of public transport reliability on users’ satisfaction. *Transport Res Part A: Policy Practice* **126**, 83–93 (2019). <https://doi.org/10.1016/j.tra.2019.06.004>
- Taylor, M.A.P.: Travel through time: the story of research on travel time reliability. *Transportmetrica B: Trans Dyn* **1**(3), 174–194 (2013). <https://doi.org/10.1080/21680566.2013.859107>
- Tilahun, N.Y., Levinson, D.M.: A Moment of time: reliability in route choice using stated preference. *J Intell Transport Sys* **14**(3), 179–187 (2010). <https://doi.org/10.1080/15472450.2010.484751>
- Tseng, YinYen, Verhoef, E., de Jong, G., Kouwenhoven, M., van der Hoorn, T.: A pilot study into the perception of unreliability of travel times using in-depth interviews. *J Choice Modell* **2**(1), 8–28 (2008)
- Tulving, E.: *Episodic and semantic memory*. Academic Press, Cambridge (1972)
- Wardman, M.: Public transport values of time. *Transp. Policy* **11**(4), 363–377 (2004). <https://doi.org/10.1016/j.tranpol.2004.05.001>
- Wijayaratna, K.P., Dixit, V.V.: Impact of information on risk attitudes: implications on valuation of reliability and information. *J Choice Modell* **20**, 16–34 (2016). <https://doi.org/10.1016/j.jocm.2016.09.004>
- Zhang, W., Jenelius, E., Ma, X.: Freight transport platoon coordination and departure time scheduling under travel time uncertainty. *Transport Res Part E: Logist Transport Rev* **98**, 1–23 (2017). <https://doi.org/10.1016/j.tre.2016.11.008>
- Zhang, Yu, Jiafu Tang, and Lily Chen. 2014. “Itinerary Planning with Deadline in Headway-Based Bus Networks Minimizing Lateness Level.” In *Proceeding of the 11th World Congress on Intelligent Control and Automation*, 443–48. <https://doi.org/10.1109/WCICA.2014.7052754>.
- Zhu, Z., Mardan, A., Zhu, S., Yang, H.: Capturing the interaction between travel time reliability and route choice behavior based on the generalized bayesian traffic model. *Transport Res Part b: Methodol* **143**, 48–64 (2021). <https://doi.org/10.1016/j.trb.2020.11.005>

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.