Transportation Research Part C 118 (2020) 102712

Contents lists available at ScienceDirect

Transportation Research Part C

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

A methodological framework to incorporate psychophysiological indicators into transportation modeling

Marisol Castro[a], C. Angelo Guevara[b,*], Angel Jimenez-Molina[c]

[a] CIS Asociados Consultores en Transporte S.A., Santiago, Chile
[b] Department of Civil Engineering, Instituto Sistemas Complejos de Ingeniería (ISCI), Universidad de Chile, Santiago, Chile
[c] Department of Industrial Engineering, Instituto Sistemas Complejos de Ingeniería (ISCI), Universidad de Chile, Santiago, Chile

A R T I C L E I N F O

Keywords:
Public transportation
Discrete choice
Travel behavior
Psychophysiological indicators
Tracking technologies
Biosensors
Moods and emotions

A B S T R A C T

Reporting and hypothetical biases are inherent to canonical methods of transportation data collection and had implied that analysis in this field has often neglected aspects that are strong behavioral drivers, such as uncertainty, physical effort or stress. Granular information on these aspects would allow measuring their valuation and/or addressing a pervasive source of endogeneity. Recent advances in miniaturization and data processing, as well as evidence that indicators from biosensors correlate with psychophysiological states and emotions, suggest that there is an opportunity to close this gap by collecting a new type of data from transportation users. This research works on leveraging this opportunity by putting forward, illustrating and testing a methodological framework to incorporate psychophysiological indicators gathered from biosensors into transportation choice behavioral modeling. The proposed framework adapts the integrated choice and latent variable approach by incorporating the psychophysiological responses as additional indicators of a latent psychophysiological state that covariates with utility. For the practical implementation of the proposed framework we also consider a specific form of aggregation of the indicators across time to avoid the curse of dimensionality arising from the unmanageably large number of folds for integration. The proposed framework is illustrated and validated using Monte Carlo simulations. Besides, a prototype field experiment was designed and performed to confirm the validity of three crucial components of the proposed framework: (i) the relation between transportation markers and emotions; (ii) the possibility of measuring those emotions through biosensors installed on travelers, (iii) and the validity of the proposed aggregation needed for practicality. In the experiment, a public transportation user travelled wearing a Printed Circuit Board that integrated tiny biosensors to capture electrodermal activity, heart rate variation, temperature and acceleration. Results provide positive evidence for the research questions, suggesting the convenience of developing larger data collection efforts in the future to take full advantage of the new framework.

1. Introduction

The planning, evaluation and management of transportation services in practice have been predominantly based on measures of travel time and cost (see e.g. Sectra, 2013; Lundqvist and Mattsson, 2002; Florian, 2008). Despite various incipient efforts reported in scientific literature that try to broaden the attributes considered, which we review in Section 2, applications in this area commonly

* Corresponding author.
E-mail addresses: mcastro@cistrans.cl (M. Castro), crguevar@ing.uchile.cl (C.A. Guevara), ajimenez@dii.uchile.cl (A. Jimenez-Molina).

https://doi.org/10.1016/j.trc.2020.102712
Received 25 October 2019; Received in revised form 2 July 2020; Accepted 2 July 2020
Available online 29 July 2020
0968-090X/

ignore aspects that are widely recognized as relevant drivers of transportation choice behavior. Among others, models commonly neglect the degree of: (i) uncertainty and emotional stress associated with waiting for public transport; (ii) comfort, cognitive load and physical effort associated with driving or making transfers; (iii) exhaustion, frustration and insecurity at congested stops or crowded vehicles; (iv) comfort and physical effort associated with having to travel standing; (v) and pleasantness with the scenic view and/or the possibility to read while seated in a bus.

The habitual neglection of various crucial drivers of transportation choice behavior in practical planning applications, and their rather incipient consideration in research efforts, is in part the result of their granularity (in time and space) and their subjectivity, which make them difficult to measure with traditional methods for data collection. The problem is that canonical data sources, based on Revealed Preferences (RP) or Stated Preferences (SP) data, are not suited to collect granular information and are also limited by reporting and hypothetical bias. Granularity is a problem because, for example, it would be highly impractical and experimentally inappropriate to inquire about subjects' mood or perceptions at every instant. Report bias may result because these data are often based on the statement of memories of past choices or experiences. Hypothetical bias may arise from a problem of ecological validity, because SP data are based on responses to hypothetical scenarios instead of real behaviors (Dhami et al., 2004).

However, recent advances in miniaturization and data processing, as well as evidence that indicators from certain biosensors correlate with psychophysiological states and emotions, suggest that there is an opportunity to close this gap by collecting a new type of data from transportation users. Yet, this advancement faces various challenges that our research is aimed to contribute addressing.

The first challenge corresponds to the use of portable sensor devices that can collect such signals and a software architecture to store, clean, and process them in transportation environments. The WesstLab at Universidad de Chile has developed a methodology that can achieve this goal, which is implemented in a wristband that makes use of tiny sensors to capture electro dermal activity (EDA), heart rate (HR), heart rate variation (HRV), skin temperature (ST) and acceleration (Jimenez-Molina et al., 2018). Similar efforts have been developed by other laboratories worldwide, from which we can highlight the development of the Empatica wristband (Garbarino et al., 2014). In this article we design and perform a prototype experiment in which we use WesstLab's wristband to validate the proposed framework.

A second challenge consists on obtaining data that are clean of noise and artifacts. This is accomplished by preprocessing each psychophysiological indicator with appropriate filters. Jimenez-Molina et al. (2018) discuss various issues to this respect, from which it worth remarking the importance of the normalization of all signals through their respective baseline, which allows standardizing the different measures within the subjects due to the variance of biological and physical conditions. We use this approach for the analysis of the prototype field experiment described in Section 5.

The third challenge, and the main area of contribution of our article, corresponds to the development of a methodological framework for the incorporation of the information gathered from these biosensors into the modeling and analysis of transportation choices. The principal hypothesis of the present research is that, if it were possible to collect granular and objective psychophysiological indicators on the evolution of aspects traditionally omitted in transportation, indicators of the utility function may be buildable from them. Then, the integration of these indicators into a common framework for transportation choice behavior would allow estimating the valuation of (or at least controlling for) transportation attributes that have been often neglected, such as overcrowding or the availability of seats, addressing the pervasive problem of endogeneity due to omitted attributes in this field. Besides, it is expected that the incorporation of these types of data would help in building models that are less prone to reporting and hypothetical bias, because they are based on psychophysiological indicators, which are directly taken from the field and cannot be falsified.

The proposed methodological framework for modeling choice behavior incorporating psychophysiological data is constructed as an extension of the conceptual framework for modeling integrated choices and latent variables (ICLV) (Walker, 2001; Vij et al., 2013). Under this setting, the psychophysiological responses are incorporated in the model as indicators of a latent psychophysiological state that covariates with utility. In addition, because the formal solution quickly becomes impractical due to the need of innumerous folds of integration, an approximation of the method that can be fully applied in practice is proposed.

The proposed framework is illustrated and validated using Monte Carlo simulations. Besides, a prototype field experiment that uses WesstLab's wristband (Jimenez-Molina et al., 2018), was designed and performed to confirm the validity of three crucial components of the proposed framework: (i) the relation between transportation markers and emotions; (ii) the possibility of measuring those emotions through biosensors installed on travelers, (iii) and the validity of the proposed aggregation needed for practicality. In the experiment, a public transportation user travelled in Santiago de Chile wearing a Printed Circuit Board that integrated tiny biosensors to capture electrodermal activity, heart rate variation, temperature and acceleration. Results show that the proposed method can incorporate psychophysiological indicators to enrich the modeling of transportation choices, improving our understanding of travelers' behavior, enriching the analysis and enhancing the modeling capabilities in this field.

The article is structured as follows. After this introduction, the next section presents a review of the scientific literature on the topic. Then, Section 3 presents the methodological framework proposed to incorporate psychophysiological indicators into discrete choice models. Section 4 presents details of a Monte Carlo simulation effort to examine the ability of the proposed method to recover the population parameters under various circumstances. Section 5 depicts a prototypical application to classify emotions while traveling on public transportation, using data collected form a prototype Printed Circuit Board. The final section offers concluding remarks and directions for further research.

2. Literature

Despite the habitual neglection of attributes beyond travel time and cost in transportation planning practice, incipient research efforts have been made in this line. For example, Raveau et al. (2014) developed a subway route choice model that includes, as model variables, indicators of occupancy and comfort, transfer experience and network topology. The choices and measurements were obtained from RP data from mobility surveys of the London and Santiago de Chile metro systems. Results show that all these additional variables were significant covariates for route choice, stressing the importance of these aspects that are often neglected. Also using RP data, but now from smart-cards, Tirachini et al. (2016) observed that some individuals at the Singapore metro were willing to increase their travel time by moving backwards, to the beginning of the line, to find a seat. This observation was used to estimate a “standing multiplier”, a penalty on the valuation of travel time when travel standing. Li and Hensher (2011), Tirachini et al. (2013, 2017) have developed and assessed various approaches to measure the value of crowding and comfort in public transportation by means of SP surveys. In these studies, the level of crowding was depicted by means of pictures, diagrams or text, with which they were able to consistently find a significant “penalty” on the valuation of travel time under crowding or standing conditions.

Other research studies have examined how the use of the travel mode affects moods, emotional states and subjective well-being, which can be understood as a broader type of quality that, to a certain extent, can capture all aspects potentially omitted that are relevant to the individual. For example, Carrel et al. (2016) connected information from satisfaction surveys with experienced public transportation level-of-service gathered from smartphone tracking and automatic vehicle location data. Indicators on feelings and satisfaction with public transportation attributes were collected through likert-scales, from which ordinal logit models where estimated. Results suggested, among other things, a strong sensitivity of satisfaction toward in-vehicle delays and the convenience of considering a baseline and a variable satisfaction level in the analysis. Gao et al. (2017) related trip satisfaction with mood and personality using path choice models at the trip-stage level. As the authors indicate, collecting mood-related data after the trip ends can distort the answers, as respondents may not correctly recall instantaneous emotions or trip-related events. To overpass this problem, they asked respondents to re-enact their trip, to reconstruct each trip stage and report the corresponding mood. Moods were classified in nine categories: enjoyment, relaxed, worried, sad, happy, depressed, anger, stressed and tired, and were measured on a scale ranging from 0 to 10. Ettema et al. (2011) also studied the relationship between trip satisfaction and emotions. They designed a survey where individuals could rate their moods in a three-level scale for pleasantness (happy–sad, satisfied–dissatisfied, joyful–depressed) and activation (active–passive, alert–sleepy, awake–dull). The results showed that mood is correlated with trip satisfaction. Morris and Guerra (2015) used the American Time Use Survey (ATUS) to study the connection between moods and travel mode. Their results showed a weak link between moods and modes, although bus and train users seem to be the ones that experience the most negative emotions.

Plenty empirical evidence has shown that psychophysiological indicators covariate among each other, and correlate with emotions and psychophysiological states such as stress, cognitive load, various emotions, and fatigue (Cacioppo, 2007). For example, it has been found that Electrodermal Activity (EDA) correlates linearly with arousal (Ganglbauer, 2009), and is also a measure of stress (Park, 2009) and frustration (Bethel, 2007). The heart rate (HR), controlling by a person’s mobility, is sensitive to cognitive demands, time constraints and uncertainty (Allanson, 2004). HR has been also be found to be sensitive to attention and correlates with arousal (Park, 2009). Heart rate variation (HRV) is used as a measure of mental work (Park, 2009), provides variables to extract a measure of fatigue in sustained attention tasks, and works as an indicator to evaluate the positive or negative emotional valence of an experience (Ganglbauer, 2009). Electroencephalography (EEG) permits detecting different states of the cognitive and emotional processes (Ganglbauer, 2009) and it also has a high precision in the detection of phasic and tonic changes in the skin that complement the detection of arousal and valence of the EDA (Kivikangas, 2011). Under controlled conditions of luminosity, pupillary dilation correlates with cognitive load (Ganglbauer, 2009) and measures overall changes on information processing (Kramer, 1991). It also responds to emotional valence (Allanson, 2004). Also, breathing frequency is a measure of the demand for tasks (Allanson, 2004) and in turn of negative valence and arousal (Park, 2009).

There is also some evidence in the same line that is specific for the area of transportation. Fitch et al. (2017) postulates that psychophysiological sensors that collect HRV can be used to assess the level of self-reported stress caused by a change in cycling conditions. Preliminary results suggest that this relationship appears to be valid only on road segments with constant speed, possibly due to the link between stress and the road environment. Sharma and Gedeon (2012) survey methods that can be used to classify stress on drivers and on with stress related diseases, and Li and Chung (2013) study a similar problem using specifically HRV. Hogertz (2010) explores the relationship between the urban environment and arousal (level of excitement), using a bracelet that captures EDA to study urban spaces influence on emotional responses. In the experiment, the participants were asked to walk on a predefined route designed to have variations of the level of excitation. Afterwards, the participants had to declare four types of emotions according to their experience: pleasant, comfortable, nervous, and uncomfortable. Most people reported that the walk on a pedestrian walkway was comfortable, while walking on a lonely, dark road was uncomfortable. The main result was to confirm the feasibility to perform an emotional evaluation through the psychophysiological signal. Besides, Wener et al. (2003) focused on the effect of a new transit service on stress levels. They linked self-reported stress with the presence of salivary cortisol collected in samples at home on a non-workday and after a trip on a workday. The results show a reduction in stress levels for the new transit service users. Finally, a recent study by Shoval et al. (2018) mapped the emotional characteristics of 68 tourists over the city of Jerusalem, using both real-time self-reported emotions and EDA signals. Emotions were measured in a 7-point likert-style scale of arousal while tourist visited the city.

3. Modeling psychophysiological indicators in discrete choices

This section details the modeling framework proposed for the incorporation of psychophysiological indicators into transportation choice behavioral modeling. We begin describing a data generation process in which these indicators are included, which may be

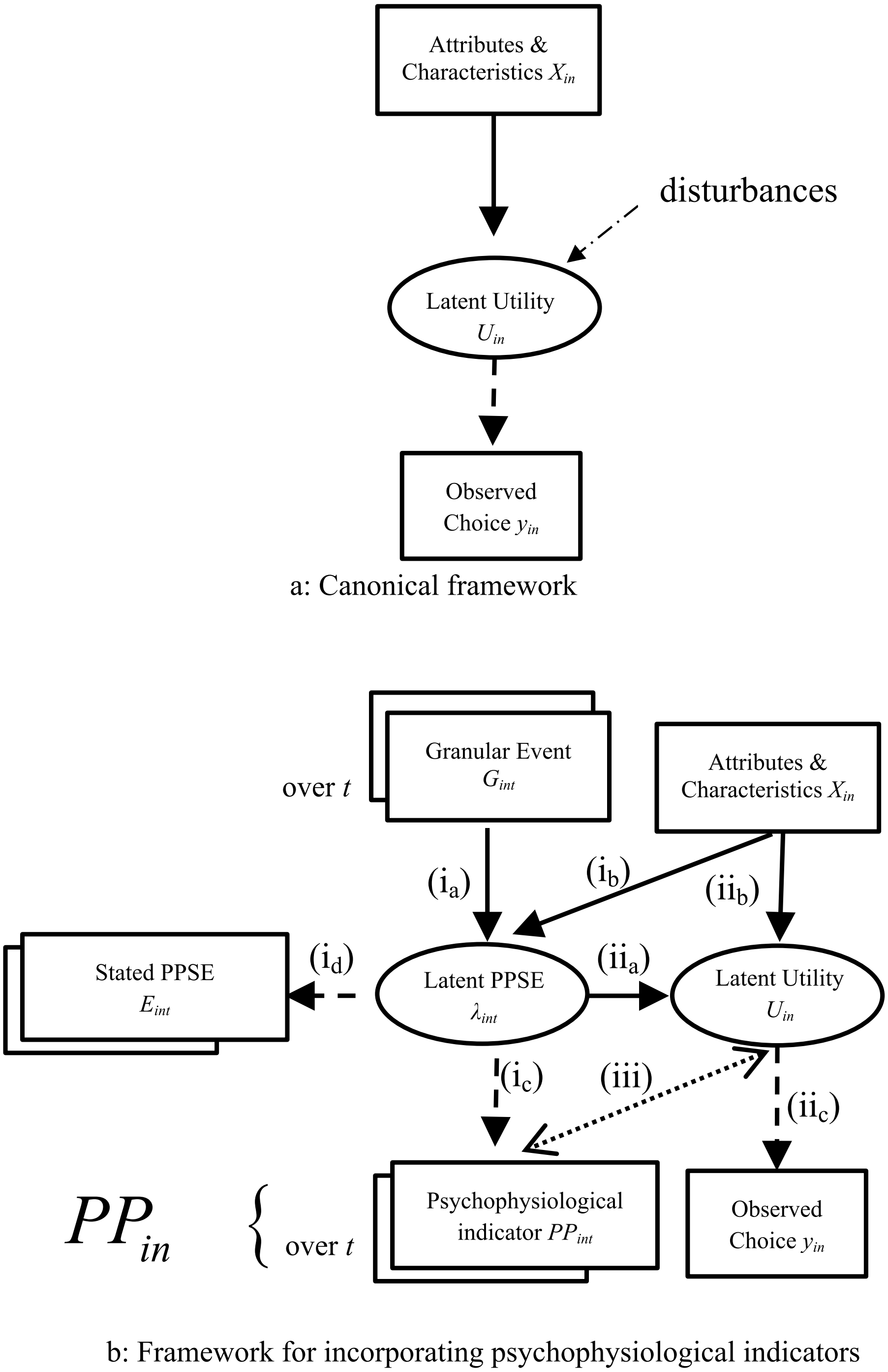


Fig. 1. Discrete Choice Modeling Framework.

understood as an extension of the integrated choice and latent variables (ICLV) framework proposed by Walker (2001). Then, we propose an approach to perform the estimation of the model parameters under this framework in practice, circumventing the curse of dimensionality (see e.g. Cherchi and Guevara, 2012) problem arising from needing to calculate many integral folds.

3.1. Conceptual framework

Consider first the canonical conceptual framework for discrete choice modeling depicted in Fig. 1a, where rectangles correspond to observed data, ovals are latent variables, full arrows correspond to model relations and dashed arrows correspond to indicators (see e.g. Walker and Ben-Akiva, 2002).

Under this canonical framework it is assumed that individual n chooses alternative i considering a mix of attributes and characteristics, synthetized by the vector X_{in}, plus some exogenous disturbances ε_{in} that conform the indirect utilities U_{in}, which are latent to the researcher. Under the Random Utility Maximization (RUM) model, which allows the identification of the model parameters, the indirect utility is divided into a systematic part V_{in} that depends on the observed variables X_{in} and the random part, formed by the disturbances ε_{in}. Then, assuming rationality, it is considered that the individual chooses the alternative with the largest utility among those in the choice-set C_n, election that is observed by the researcher through the indicator y_{in}, which takes value 1 if alternative i is chosen by individual n and zero otherwise.

In the canonical model X_{in} typically consists of travel time and cost attributes of the alternatives, potentially interacted with individual's characteristics, like income or age. Besides, the disturbances account for all omitted attributes, *i.e.* unobserved by the researcher, which must be assumed to be independent of X_{in} to guarantee the consistent estimation of the model parameters (see Guevara, 2015).

The omitted attributes embedded in the disturbances in Fig. 1a may be, among many others, the fact that individual n is travelling standing or that another passenger is playing annoying "music" without headphones in public transportation; or maybe that the road is uneven or the driver in front behaves erratically, when driving a car. The fact that these attributes are omitted by the modeler not only makes it impossible to measure their valuation, but also may preclude the consistent estimation of other parameters of the model when they are not independent of X_{in}. This may happen, e.g., if the bumpier road tends to be the cheapest or the most crowded public transportation route tends to be the shortest.

The potential endogeneity that arises from the omission of dependent attributes may be addressed, among other methods, using the latent variables (Walker, 2001; Palma et al., 2016; Rossetti et al., 2018) or the multiple indicator solution (Fernández-Antolín et al., 2016; Guevara et al., 2020; Guevara and Polanco, 2016; Mariel et al., 2018) approaches. For both cases, the researcher would need to collect indicators of the omitted attributes and/or build a proper structural equation for them, something that may prove difficult to achieve in practice, causing further problems (Guevara, 2015). The indicators may be gathered, e.g., from post-trip questions about the "overall comfort" perceived on the recent trip, and the structural equations may include characteristics of the individual, to try to capture differences in perception, together with some recorded events, such as a measure of the level of crowding.

However, post-trip questions are inevitably feeble to report bias and limited granularity. Likewise, the building of the proper structural equations is always questionable due to incomplete information. Those limitations may be circumvented with the proper incorporation of psychophysiological indicators into the choice modeling framework.

Consider now that, using biosensors, it is possible to collect psychophysiological indicators that are directly related to psychophysiological states and emotions (PPSE) perceived by the transportation user. These devices have been significantly miniaturized and cheapened in recent years, opening the possibility for using them to collect transportation information in the field, with the great advantage of collecting granular and objective information in time and space in a very non-intrusive way. Arguably, report bias may not be an issue in this case because, whatever the individual truly felt at any moment could, in theory, be objectively recorded with these devices.

The question that remains is how to incorporate in practice the plethora of information that would be obtainable from the biosensors into the choice modeling framework. This is the main contribution of this research.

The first step toward that objective is to note that, if indicators gathered from biosensors are related to individuals' PPSE, they could also be understood as an *incidental* measure of the indirect utility that explains the choice. Although PPSE are arguably caused by some aggregated or granular events occurring during the trip, they are the individual's perception to those events and, as such, to some degree, they must be related with the indirect utility that defines the choice.

Thus, it can be considered that individuals choose their alternatives somehow combining both "budgetary" attributes and "granular" events that can occur during the trip. The former could correspond, e.g., to the traditional cost, travel time and the later correspond to, e.g., noise, crowding, insecurity, sense of danger, comfort or a pleasant sight, which are somehow incorporated in the indirect utility U_{in} through the PPSE.

This conceptual framework is depicted in Fig. 1b, where G_{int} corresponds to a vector of granular events g_{inst} experienced by individual n at time t on alternative i, and PP_{int} corresponds to the respective vector of psychophysiological indicators. PP_{int} is assumed to be observed by the researcher, but G_{int} may or may not be fully observed and it is thus represented by a rectangle with dashed borders.

Consider first relations (i_a) and (i_b) in Fig. 1b. These links represent the fact that psychophysiological states and emotions PPSE λ_{int}, perceived by an individual n while traveling on alternative i at time t, depend both on the vector G_{int} of granular events g_{inst} occurring at time t and on the attributes and characteristics X_{in}. All the disturbances in Fig. 1b are assumed to be exogenous and are not the same of Fig. 1a.

Besides, relation (i_c) corresponds to a measurement of λ_{int} through the psychophysiological indicators PP_{int}. Relations (i_a), (i_b) and (i_c) can be synthetized as shown in Eq. (1), where f is a function, which we assume approximately linear, δ_{int} represents a measurement error of the psychophysiological indicator, which is assumed to be exogenous, and $\alpha$and β are parameters.

$$PP_{int} = f(\lambda_{int}(G_{int}, X_{in})) \approx \alpha_{in} + \beta_n \lambda_{int}(G_{int}) + \delta_{int} \tag{1}$$

Note that, despite PP_{int} depends both on G_{int} and X_{in}, since the latter does not depend on time, it can be argued that only G_{int} will truly remain as an independent variable in Eq. (1), while the effect of X_{in} is being reduced to a constant α_{in} that depends on the individual and the alternative, as shown at the right of Eq. (1). Furthermore, since psychophysiological indicators are often analyzed after subtracting their mean over time to account for a baseline (Burt and Obradović, 2013), even the effect of the constant may vanish from Eq. (1). This effect will potentially wipe out the impact of relations (i_b) shown in Fig. 1b, or at least reduce them to a constant by alternative i, α_i.

Relation (i_d) in Fig. 1b correspond to a measurement of λ_{int} through the stated PPSE E_{int}. This could include questions about mood, like in Carrel et al. (2016) or Gao et al. (2017), or directly about emotions, like the ones we use in the prototypical field experiment based on the circumplex framework that is described in Section 5.

Similarly to Eq. (1), relations (i_a), (i_b) and (i_d) can be synthetized as shown in Eq. (2), where f is a function, which we assume approximately linear, $\tilde{\delta}_{\text{int}}$ represents a measurement error of the stated PPSE indicator, which is assumed to be exogenous, and $\tilde{\alpha}$ and $\tilde{\beta}$ are parameters. If E_{int} is recorded as a discrete variable, Eq. (2) could be formulated instead as a discrete choice model.

$$E_{int} = f(\lambda_{int}(G_{int}, X_{in})) \approx \tilde{\alpha}_{in} + \tilde{\beta}_n \lambda_{int}(G_{int}) + \tilde{\delta}_{int} \tag{2}$$

Consider now relations (ii_a) and (ii_b) in Fig. 1b. Without loss of generality, it can be considered that effect of the granular and budgetary attributes in the indirect utility U_{in} can be respectively separated in the systematic utility V_{in} into two terms, as shown in Eq. (3).

$$U_{in} = V_{in}\{V_{in}^G(G_{in}), V_{in}^B(X_{in})\} + \varepsilon_{in} \tag{3}$$

V_{in}^B corresponds to the traditional systematic part of the utility, which is often considered to be linear in the budgetary attributes X_{in}, capturing relation (ii_b) in Fig. 1b. In turn, V_{in}^Gcorresponds to some function of the granular events G_{int}, transformed through the psychophysiological states and emotions PPSE and integrated over t across the whole trip, capturing relation (ii_a) shown in Fig. 1b.

Besides, relation (ii_c) corresponds to a measurement of U_{in} through the observed choice y_{in}. Assuming rationality, relations (ii_a), (ii_b) and (ii_c) can be synthesized as shown in Eq. (4).

$$y_{in} = 1[U_{in} \geqslant U_{jn}\ \forall\ j \in C_n] \tag{4}$$

Finally, consider relation (iii) in Fig. 1b, which represents a correlation between the psychophysiological indicators PP_{int} and the utility U_{in}. Different from relations of types (i) and (ii), relation type (iii) is not causal, is not a link that comes from the behavior of the individual, but is a mere correlation implied by the way in which the data is generated. To remark its incidental nature, relation (iii) is depicted as a dashed two-headed arrow in Fig. 1b. It worth noting that a similar incidental relation could be stablished between PP_{int} and E_{int}.

Despite PP_{int} may in principle correlate with the whole utility U_{in}, since budgetary attributes X_{in} can be assumed to remain unchanged during the trip, PP_{int} would only be accounting for the granular attributes or events G_{int}. Under this setting, any mean effect ignored by such a treatment of the psychophysiological indicators would be captured by a unique additive alternative specific constant of the systematic utility, shared with V_{in}^B. In general, this constant would then have to be heterogeneous in the sample, but that may not be needed if the baseline is subtracted from PP_{int}. Furthermore, under this setting, the vector of psychophysiological indicators PP_{in}, somehow integrated over t, can be a function f of V_{in}^G, as shown at the left bottom of Fig. 1b, and in Eq. (5), where ξ_{in} is an exogenous error term.

$$PP_{in} \approx f(V_{in}^G(G_{in})) + \xi_{in} \tag{5}$$

3.2. Toward a practical implementation of the conceptual framework

The problem that remains to be solved is to propose a comprehensive, and yet practical, implementation of the conceptual model depicted in Fig. 1b. In this subsection we introduce further functional assumptions and simplifications that allow achieving such a goal.

Consider first that the psychophysiological indicators are measured, for each individual n, on each alternative i, during the entire duration of the trip. As the psychophysiological states and emotions perceived by the individual could change during a trip due to different events (for example, with increased noise or unsafe situations), the psychophysiological indicators are measured in different instants t ($t = 1, \ldots, T_{in}$) for each individual and alternative.

Then, let PP_{int} be a psychophysiological indicator measured on instant t for individual n during his/her trip on alternative i. The link between the psychophysiological indicator with actual events occurring during trips could be defined as

$$PP_{int} = \alpha_{PP_{in}} + \delta_{PP}\lambda_{int}^1 + \delta_{int} = \alpha_{PP_{in}} + \frac{\theta_{PP}}{\mu_1}\ln\sum_s e^{\mu_1\alpha_s g_{inst}} + \delta_{\text{int}}, \tag{6}$$

where g_{inst} is a variable that registers the occurrence (or the intensity) of the of a granular event s at instant t during the trip undertaken by individual n in alternative i; α_s is the parameter associated with event s; and δ_{int} is an exogenous error term. Eq. (6) corresponds to a practical version of Eq. (1) and accounts for relations of type (i) in Fig. 1b.

Eq. (6) follows from assuming that, among all events that occurred at instant t, the individual feels an instant emotion or psychophysiological state and emotion PPSE λ_{int}^1 that depends on the worst event that occurred instant t. Thus, assuming that the impact of event s at instant t is distributed Gumbel with location $\alpha_s g_{inst}$ and scale μ_1, the instant PPSE λ_{int} will be Gumbel distributed with location $\frac{1}{\mu_1}\ln\sum_s e^{\mu_1\alpha_s g_{inst}}$ and scale μ_1 (see e.g. Ben-Akiva et al., 1985, pp. 104).

Besides, since the psychophysiological signal is registered with some error and Eq. (6) is a regression to the mean, $\alpha_{PP} = \frac{\gamma}{\mu_1} + \alpha_{in}$ with γ being the Euler constant (see e.g. Ben-Akiva et al., 1985 pp. 104, 2), α_{in} the constant in Eq. (1), and δ_{int} is then related to the measurement error of the biosensor device.

An equivalent argument could be used to build a practical expression to model V_{in}^{G} as a function of G_{in}. First, for each t, one can consider that there is some sort of instant granular PPSE λ_{int}^{2}, which will be Gumbel distributed with location $\frac{1}{\mu_2}\ln\sum_{s} e^{\mu_2\beta_s g_{inst}}$ and scale μ_2, aggregating by this the effect of all potential events s occurring at instant t.

To aggregate the PPSE over t it will again be assumed that the individual n will try to minimize (maximize) the latent variable λ_{int}^{2}, related with the occurrence of unpleasant(pleasurable) events. Therefore, for making a choice she will consider the worst(best)-case scenario, given her knowledge of actual experiences or assumptions. Under this setting, following Ben-Akiva et al. (1985, pp. 105), the granular systematic utility could be assumed to be $V_{in}^{G} = E(\max(\lambda_{in1}^{2},...,\lambda_{\text{int}}^{2},...,\lambda_{inT}^{2}))$, which will distribute Gumbel with location $\frac{1}{\mu_3}\ln\sum_{t} e^{\mu_3\lambda_{int}^{2}}$ and scale μ_3.

Summarizing, under these assumptions, a plausible functional form to represent V_{in}^{G}, i.e, relations type (*ii*) in Fig. 1b, could be the expression shown in Eq. (7).

$$V_{in}^{G}(G_{in}) = \lambda_{in}^{3} = \frac{1}{\mu_3}\ln\sum_{t} e^{\mu_3\lambda_{int}^{2}} = \frac{1}{\mu_3}\ln\sum_{t} e^{\frac{\mu_3}{\mu_2}\ln\sum_{s}\exp(\mu_2\beta_s g_{inst})} \tag{7}$$

The final step corresponds to the operationalization of the function suggested on Eq. (5), to link the aggregated measure of granular systematic utility V_{in}^{G} with an aggregated measure of the psychophysiological indicators PP_{int}. As it was stated before, the relation of type (*iii*) is not causal but instead a byproduct of the fact that both PP_{int} and V_{in}^{G} depend on the same granular events G_{in}.

The final challenge in this case is to find a function that could aggregate the PP_{int} over the instants t, accounting in the best possible way for the correlation with V_{in}^{G}, representing Eq. (5), a practical aggregated version of relation (*iii*) in Fig. 1b. Despite it is unclear how to derive a formal functional form to solve this challenge, a natural candidate corresponds to mimic the form in which the G_{in} is aggregated over t, as shown in Eq. (8).

$$PP_{in} \approx \ln\sum_{t} e^{\gamma\ln\sum_{s}\gamma_s PP_{inst}} \tag{8}$$

The plausibility of using this aggregated function of the psychophysiological indicators is going to be assessed using Monte Carlo analysis and real data. In the Monte Carlo section, it will be shown that such a setting can indeed recover the population parameters under various settings. In the real data section, it will be shown that this expression can explain stated PPSE E_{int} (see Fig. 1b) felt on a public transportation trip, at a level comparable to that of random forest approach that is agnostic to the functional form.

3.3. Model estimation

Three possible approaches to estimate the model parameters appear as possible, depending on whether the detailed information on the granular events is available or not and on the computational capacity available.

The first approach considers a case when a full account of g_{inst} is available. In this case the model could be estimated by traditional maximum likelihood using some practical version of the utility shown in Eq. (3). This means having a detailed record of each and every granular event that may affect the PPSE, which seems difficult in practice.

The second approach considers the more realistic case in which the g_{inst} are not available. In such a case, one option would be to treat λ_{int} in Eqs. (6) and (7) as latent variables, using the psychophysiological indicators PP_{int} to identify the model, resulting in the likelihood function depicted in Eq. (9).

$$P(y_{in}|\ x_{ink};\ \beta_k,\ \gamma_s,\ \sigma_\varepsilon,\ \sigma_\xi) = \int_{PP_{int}} P(y_{in}|\ x_{ink},\ PP_{int};\ \beta_k,\ \sigma_\varepsilon)\cdot f_\xi(PP_{ijt}|\ x_{ink};\ \gamma_s,\ \sigma_\xi)dPP_{int} \tag{9}$$

However, since the number of folds of the integral in the problem depicted in Eq. (9) depends on the number of instants, it could quickly grow to the hundreds, making this approach for estimation impractical due to what is known as the curse of dimensionality (see e.g. Cherchi and Guevara, 2012).

The third approach consists of a practical to way avoid the curse of dimensionality problem. The method consists in treating instead the aggregated granular utility V_{in}^{G}, using Eq. (8) as an approximated measurement equation. Note that, in a practical application, the weights γ shown in Eq. (7) would have to be estimated together with other model parameters.

In next section we will use Monte Carlo analysis to test the feasibility and performance of the third approach proposed for model estimation of the model parameters under the presence and absence of endogeneity.

4. Monte carlo simulation analysis

4.1. Experimental design

The Monte Carlo simulation exercises undertaken in this section have two objectives. First, by generating simulated data sets with known underlying model parameters, examine the ability of the proposed aggregated function of PP_{int} using Eq. (8) to recover parameters from finite samples in a model. Second, to quantify the effect of omitting granular events from the utility function when these events are correlated with the observed attributes and are perceived by individuals and condition their choices. The latter objective also considers the analysis of the feasibility of correcting such an effect with the proposed method.

For the experiment, consider the case of individual n ($n = 1, ..., N$) choosing between two alternatives ($i = 1, 2$) based on the utility function shown in Eq. (10), which can be seen as a practical representation of Eq. (10).

$$U_{in} = V_{in}\{V_{in}^{G}(G_{in}),\ V_{in}^{B}(X_{in})\} + \varepsilon_{in} = \theta_i + \beta_G V_{in}^{G}(G_{in}) + \beta_1 x_{in1} + \beta_2 x_{in2} + \varepsilon_{in} \tag{10}$$

The explanatory variables $\{x_{in1}, x_{in2}\}$ are drawn from univariate normal distributions. The coefficients β_G, β_1 and β_2 are equal to 0.5, while θ_i are set equal to zero. Values for the error terms ε_{in} are drawn from a standard univariate Gumbel distribution. Individuals choose the alternative with the higher utility value ($y_{1n} = 1$ if $U_{1n} > U_{2n}$, and 0 otherwise).

To construct the function $V_{in}^{G}(G_{in})$, in each instant t ($t = 1, ..., T$) let g_{inst} be a random variable measuring the intensity of an event s at instant t for individual n choosing alternative i. For simplicity, we consider the case with a unique event s. As a result, the PPSE λ_{inst}^{1}will be directly g_{int}, which follows a normal distribution.

Two data generation scenarios are defined for g_{int}. In the first scenario, denoted "uncorrelated data", $\text{cov}(x_{in}, g_{\text{int}}) = 0$ and g_{int} is drawn from a standard normal distribution; that is, the explanatory variables are independent of the granular events. In the second scenario the granular events are generated as $g_{\text{int}} = x_{in1} + \widehat{\varepsilon}_t$ with $\widehat{\varepsilon}_t \sim N(0, 1)$. This last scenario is denoted "correlated data" due to the endogeneity between explanatory variables and granular events. In both cases, the function $V_{in}^{G}(G_{in})$is obtained from Eq. (7), with parameters μ_2, μ_3 and β_s equal to one.

The psychophysiological indicators PP_{int} are generated as a function of the events perceived by the decision makers throughout the entire trip following Eq. (6), with $\alpha_{PP} = \frac{\gamma}{\mu_1} + \alpha_{in}$, $\alpha_{in} = 0$, and θ_{PP}, μ_1 and α_s equal to one. The error term δ_{int} is drawn from a standard normal distribution.

Although the psychophysiological indicators were generated directly from g_{int}, the model is estimated under the assumption that the granular events are unknown for the researcher who can only observe PP_{int} over time. Then, the utility function in Eq. (10) is estimated, as an approximation, just replacing the indicator PP_{in} in the utility, to account for$V_{in}^{G}(G_{in})$ in Eq. (8). Despite this approach introduces an additional measurement error, compared to treating $V_{in}^{G}(G_{in})$ as a latent variable using Eq. (5), it has been shown that its effects may be negligible (Guevara, 2015) and it avoids the need for integration. Besides, it should be noted that under this framework, the parameters γ in Eq. (8) and β_G in Eq. (10) cannot be identified and they are jointly estimated.

4.2. Simulation results

The above data generation process is undertaken 100 times with different realizations of g_{int}, ε_{in} and δ_{int} to generate 100 different data sets for both the correlated data and correlated data scenarios. To estimate the coefficient vector (θ_0, β_G, β_1, β_2), we evaluate an approximation of the real generation process, as proposed in Eq. (8).

The results are presented in Table 1 for different values of N, with $T = 100$. The table is divided in two main sections: the upper section provides the results for the uncorrelated datasets, where the granular events are generated independently from the explanatory variables, while the lower section presents the results for endogenous variables. Additionally, Table 1 shows the results both for the proposed model, using the Extreme Value approximation depicted in Eq (8) ("EV model"), and an ordinary Multinomial Logit that does not account for the granular events perceived by the individuals ("MNL model"). For each model, the table displays the true value of the parameters (second column), followed by the parameter estimate results and the standard error estimate results.

To evaluate the performance of the models, we computed three measures of accuracy, all presented in Table 1. The first measure, referred as the absolute percent error (APE), is the absolute perceptual difference between the estimated parameter $\tilde{\beta}$ and its real value β_0. The second measure is the result of a hypothesis test under the null hypothesis that the estimated parameter is equal to the design value. This measure is referred as the "t-test" in Table 1 and is computed as $(\tilde{\beta} - \beta_0)/\tilde{\sigma}_\beta$, where $\tilde{\sigma}_\beta$ is the estimated standard deviation of the parameter. Values of the t-test above 1.96 indicate the estimated parameter is significantly different from the real value with 95% confidence. Finally, to acknowledge the effect of the scale parameter in both models, the ratio between β_1 and β_2 was computed from the estimated parameters, and a t-test was calculated comparing the estimated ratio to the real ratio.

The results in Table 1 indicate that the proposed method to account for psychophysiological states and emotions using an indirect aggregate measure gathered from biosensors can properly recover the model parameters. Results show that the recovery is better achieved as the number of individuals in the sample increases, as can be observed by comparing the estimate of the parameters with the true values (see the column entitled "parameter estimates"). In fact, the APE is not higher than 3% for any parameter, with an overall mean value of 1.1% across all parameters. The standard error estimates of the parameters indicate good empirical efficiency of the estimator. As a result, the corresponding t-test shows no significant difference between the estimated and true values.

The results of the MNL model without psychophysiological indicators show larger estimation errors, which rise up to 63% when the granular events are endogenous with the explanatory variables. The t-tests confirm these differences as in three of the four simulated scenarios, the estimated parameters are significantly different from the true values.

The distribution of the ratio β_1/β_2 is presented in Fig. 2 for both data set generation processes and $N = 5000$. In the uncorrelated data case, both the EV model and the MNL model correctly estimate the parameter's quotient, while in the correlated data case the MNL model provides a large bias, as confirmed by the t-test presented in Table 1. Besides, in both cases the EV model is more efficient, thanks to the use of the psychophysiological indicator.

In summary, the proposed approximation can accurately recover the true parameters of the utility function, while ignoring the psychophysiological indicators in the utility function can lead to biased estimates, particularly in the case with endogenous granular events.

Table 1
Simulation results.

Uncorrelated data(granular events uncorrelated with explanatory variables									
Parameters	Parameters design value	EV Model				MNL Model			
		Estimates		APE	*t*-test	Estimates		APE	*t*-test
		Parameter	Sd. Error			Parameter	Sd. Error		
N **= 1000**									
θ_0	0.00	0.009	0.248	0.9%	−0.04	0.011	0.230	1.1%	−0.05
β_G	0.50	0.489	0.048	2.3%	0.24	–	–		–
β_1	0.50	0.495	0.059	1.0%	0.08	0.431	0.055	13.8%	1.26
β_2	0.50	0.496	0.061	0.9%	0.07	0.431	0.056	13.8%	1.22
Ratio β_1/β_2	1.00	1.001	–	0.1%	0.95	1.000	–	0.0%	1.13
N **= 5000**									
θ_0	0.00	−0.005	0.110	0.5%	0.05	0.003	0.103	0.3%	−0.03
β_G	0.50	0.485	0.021	3.0%	0.72	–	–		–
β_1	0.50	0.489	0.026	2.2%	0.43	0.426	0.024	14.8%	3.05
β_2	0.50	0.491	0.027	1.7%	0.31	0.425	0.025	14.9%	2.97
Ratio β_1/β_2	1.00	1.005	–	0.5%	−0.41	1.002	–	0.2%	0.59
Correlated data (granular events correlated with explanatory variables									
Parameters	**Parameters design value**	**EV Model**				**MNL Model**			
		Estimates		**APE**	***t*-test**	**Estimates**		**APE**	***t*-test**
		Parameter	**Sd. Error**			**Parameter**	**Sd. Error**		
N **= 1000**									
θ_0	0.00	0.009	0.284	0.9%	−0.03	0.391	0.255	39.1%	−1.53
β_G	0.50	0.503	0.045	0.6%	−0.06	–	–		–
β_1	0.50	0.504	0.077	0.7%	−0.05	0.820	0.069	63.9%	−4.63
β_2	0.50	0.505	0.070	1.0%	−0.07	0.415	0.062	17.0%	1.37
Ratio β_1/β_2	1.00	1.003	–	0.3%	0.60	1.976	–	97.6%	33.35
N **= 5000**									
θ_0	0.00	0.009	0.126	0.9%	−0.07	0.418	0.113	0.418	−3.69
β_G	0.50	0.498	0.020	0.5%	0.12	–	–		–
β_1	0.50	0.500	0.034	0.1%	0.01	0.814	0.031	0.629	−10.26
β_2	0.50	0.497	0.031	0.6%	0.10	0.407	0.028	0.185	3.36
Ratio β_1/β_2	1.00	0.995	–	0.5%	1.07	1.999	–	99.9%	68.09

5. Prototype experiment

5.1. Experimental goals

Methods to collect psychophysiological data from transportation travelers is in its infancy and thus, despite costs and technical complications are steadily declining, it is still expensive and relatively complicated. This prototypical experiment has the overall goal of assessing the feasibility of some critical components of the proposed approach with real data to decide if a future massive data collection effort of this type worth the effort.

As such, three specific goals are pursued. The first is to confirm a relation between transportation markers and emotions, that is, to empirically confirm the existence of relations (i_a) and (i_d) in Fig. 1b. The second goal corresponds to assess the possibility of measuring those emotions through biosensors installed on travelers, which corresponds to relation (i_c) in Fig. 1b. The third goal corresponds to the validation of the proposed aggregation of PP_{int} needed for practicality, as shown in Eq. (8), by comparing its ability to recover the stated PPSE E_{int}, with that of an agnostic random forest approach, which is incidentally supported by relations (i_c) and (i_d).

To achieve the experiment goals, we conducted a field experiment in which one participant performed a multimodal route on the public transportation system of the city of Santiago, Chile, while his psychophysiological responses were measured by biosensors and recorded in a database. The trip included a stage by bus and another by subway, transfers, waiting at bus stops and underground platforms and walks. This experiment has a limited scope and should be understood as a prototype case to assess whether (i) the biosensors are able to correctly measure psychophysiological variables and (ii) the proposed framework can be used with real data.

To elicit the participant's emotions during the journey, in our experiment a simplified version of the circumplex model was used. The circumplex model of affect, proposed by Russell (1980), is one of various models for the measurement of emotions have been developed, and it states that emotions are generated primarily as a combination of two dimensions: valence and arousal. As shown in Fig. 3a, the vertical axis corresponds to the state of arousal where a high arousal is associated with activation and a low one to deactivation. In the same way, on the horizontal axis a positive valence is associated with pleasant states and the contrary with a

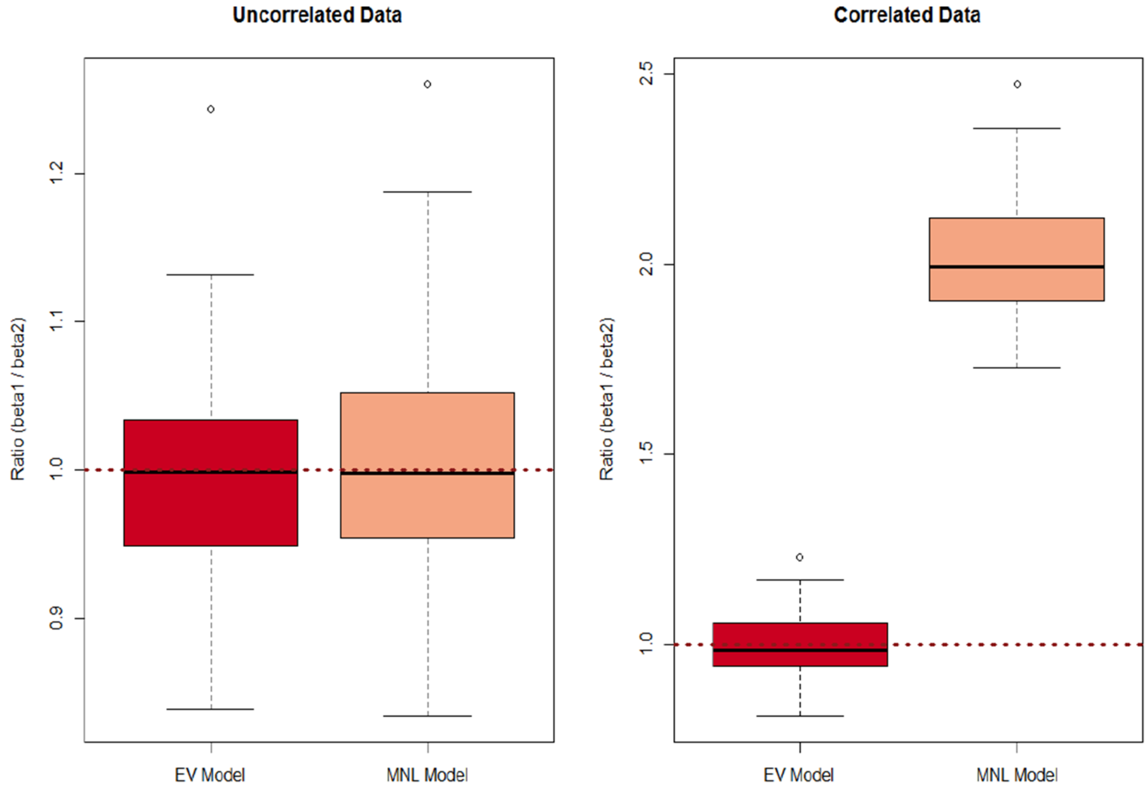
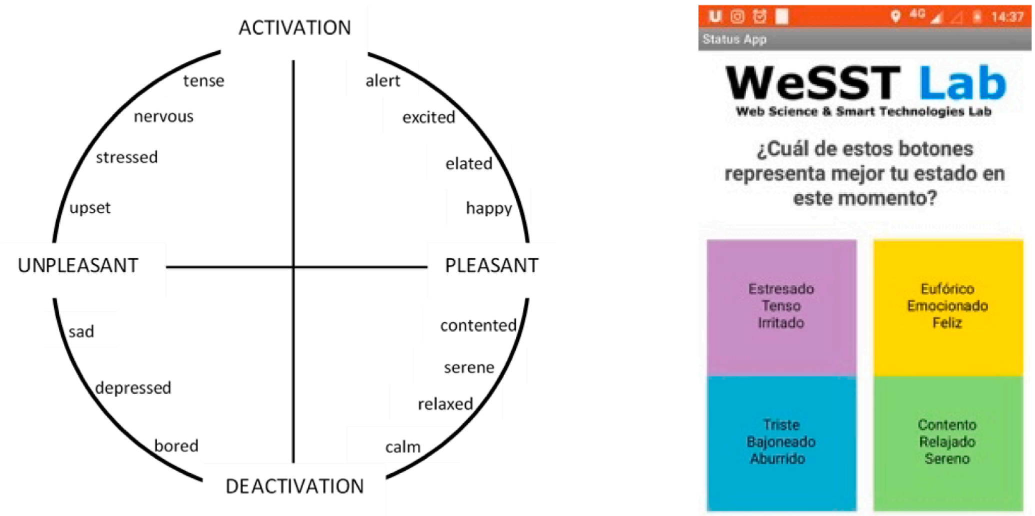


Fig. 2. Boxplot of Ratio β_1/β_2 for Simulated Data (N = 5000).



a: Original Circumplex Model (Adapted from Posner et al., 2005)

b: Mobile Application with Simplified Circumplex Model (In Spanish)

Fig. 3. Graphical Representation of the Circumplex Model.

negative one (Posner et al., 2005). For example, emotions of happiness and excitement are found in the quadrant with positive valence and high arousal, euphoria being a greater value of arousal. On the other hand, in the quadrant of negative valence and high arousal are emotions such as stress or nervousness. In the quadrant of negative valence and low arousal are the emotions of sadness and boredom, negative and more passive emotions. Finally, in the positive valence and low arousal quadrant we find emotions such as calm and serene.

The simplified circumplex model was implemented through a mobile application developed over the open source platform MIT

App Inventor (Pokress and Veiga, 2013). The application captured the self-reported participant's quadrant of emotions in random timestamps of an average of 5.1 min. As shown in Fig. 3b, a preliminary selection of only three representative emotions by quadrant (in Spanish) were considered. A comprehensive development of a simplified circumplex model for transportation users, is left for further research.

5.2. Experimental methods

5.2.1. Participant characteristics

One participant (23 years old, male), an engineering student from Universidad de Chile recruited through a message in the classroom, participated in the experiment. He declared do not suffer cardiovascular diseases or take medications or drugs that could have affected his normal behavior. In addition, the participant is a daily user of the public transportation in Santiago.

He was presented an informed consent about the procedure, the purpose of the experiment, voluntary participation, right to decline to participate at any moment, how to access the research results and researcher's information.

5.2.2. Psychophysiological sensors

For data acquisition, the following signals were used: EDA, photoplethysmography (PPG), and ST. To measure the EDA and HR signals, the Shimmer GSR+ unit sensor was used with a sampling frequency of 120 Hz. The position of the electrodes for measuring the EDA was the palm area of the proximal phalanx of the index and ring fingers of the left hand (Villarejo et al., 2012). The optical sensor that functions as a photoplethysmograph was attached to the lobe of the right ear (Ye et al., 2017). The Shimmer Bridge Amplifier+ unit sensor with a sampling frequency of 50 Hz was used to measure ST. The sensor was applied under the right armpit. This sensor was synchronized with the EDA and pulse sensors using a base provided by Shimmer together with the Consensys software.

5.2.3. Experimental procedure

As soon the participant arrived in the experimental room, the experiment was explained to him, and he was asked to read and sign the informed consent, as well as a questionnaire to obtain his basic anonymous information. The participant was seated in front of a screen, and the sensors were connected in the following order: ST, EDA, and PPG.

Prior to the experiment in the outdoor environment, the user underwent a relaxation session consisting of the visualization of three four-minute videos of landscapes with background instrumental music. Then, the participant was asked to take deep breaths for one minute with his eyes closed and with soft background instrumental music. This procedure aimed to eliminate the Hawthorne effect—modification in the behavior of the subjects due to their awareness of being studied—and physiological effects similar to the "white coat" effect in measured signals (Parsons, 1974).

The chosen trip route includes a segment by bus and another by metro, both supervised by the experimenter. As soon as the baseline condition signals were captured, the participant walks two hundred meters towards the bus stop, where he waits for around five minutes. The participant rides a bus route with 13 bus stops. Then the passenger transfers to the nearest metro station, a walk that takes about 6 min. This stage includes a wait for the metro, and a trip through four stations, at which point the participant undertakes the same trip in the opposite direction. To measure different conditions, the participant is instructed to travel both seated and standing on both modes. During the journey, the participant received automatic, random notifications a mobile application especially crafted to record his emotional state, choosing one among five options. Each notification was randomly generated within 1–10 min from the previous notification. In addition, the experimenter recorded notable events that happened during the trip, such as the appearance of vendors or public transportation inspectors, among others. Finally, the experimenter recorded the timestamps of the participant's actions, such as getting on or off the bus, sitting or standing on the seat and, walking to the bus stop or metro station. The entire experiment lasts almost 2 h.

5.2.4. Data preprocessing

EDA raw data provides the values of electric resistance of the skin in kilohms [kΩ]. To reduce noise and eliminate motion artifacts, two procedures are performed: first, a strict instruction is given to the participant not to move the hand or fingers where the electrodes are attached, and second, the signal is filtered with a low-pass cut-off frequency of 5 Hz. Furthermore, on the recommendation of the literature (iMotions Biometric Research Platform, 2016), capture resolution is reduced without risk of data loss. The EDA signal measured with a sampling frequency of 120 Hz is reduced to 10 samples per second. The phasic component is extracted by applying a median filter with a window width of ± 4 and subtracting the average of the current sample. This component allows the detection of peaks of the EDA signal. With slow transitions, the phasic component does not show major variations.

The raw data of the PPG yield signal values in millivolts [mV]. From this signal, it is possible to obtain the HR. Previously, the PPG signal is processed using a low-pass filter with a cut-off frequency of 16 Hz with a Blackman window, obtaining a cleaner signal. Then, HR is obtained via the following steps: first, the peaks must be found; second, the time between them is substracted (Δt in [miliseconds/pulse]); third, they are converted from hundredths to seconds and from [seconds/pulse] to [pulses/second], which is then multiplied by 60 to convert to [beats/minute].

The raw data yield ST values in Celsius degrees. The processing of this signal consists of using a low-pass filter with a cut-off frequency of 1 Hz, as concluded based on the data collection in Haapalainen et al. (2010).

A time window is chosen between the four and six minutes of the measurement in the relaxation stage to calculate the baseline of each psychophysiological indicator. This choice avoids considering the moment between the initial stage of adaptation to the sensors and the beginning of the walk to the whereabouts. The baseline average is subtracted from all the preprocessed signals. Then, the z-score is applied to the resulting signal in order to normalize the data.

5.3. Data generation

The data used in the analysis is a synthetic database generated from the empirical experiment previously described. As explained before, the person had to choose an emotion among the five available emotions, including a neutral state, defined in the simplified Circumplex depicted in Fig. 3b. Each choice occasion was triggered by a question shown automatically displayed in a smartphone app and, as a result, 23 choice occasions were available for analysis. On the other hand, the psychophysiological indicators were measured at different frequencies and then filtered and normalized, as explained in the previous subsection.

To increase the number of available choice scenarios, we constructed pseudo-choices considering time windows of 5 and 10 s, such that between two different choice occasions, all time windows are assigned the same emotion as the first chosen emotion. Using this procedure, we constructed 1378 pseudo-choices for the 5-seconds windows and 689 pseudo-choices for the 10-seconds windows, which account for the model's dependent variable.

As explanatory variables, the four psychophysiological indicators (HR, ACC, EDA and ST) were initially considered. However, it has been shown that the EDA measurements are associated with emotions (Ganglbauer, 2009; Park, 2009; Bethel, 2007) and, therefore, it was the only one finally considered. Formally, we model E_{inst} as a discrete choice that depends on P_{in}, the aggregated version of P_{inst} proposed in Eq. (8). This incidental relation rests upon relations (i_c) and (i_d) and works an indirect way to assess the possibility of measuring emotions through biosensors installed on travelers and for validating the proposed aggregation of P_{inst} needed for practicality. In addition, 15 different stages of the trip were identified over time, such as waiting for the bus or boarding the metro. Given that EDA can be influenced by these trip stages (for example, EDA could increase with movement, such as walking to the bus stop), these trip stages were also included in the model as control variables and also as evidence of relations (i_a) and (i_d), shown in Figure 1b, that account for the relation between transportation markers and emotions.

It should be noted that even if these relationships exist, emotions could be associated with events and attributes related (travel time or cost) or unrelated to the trip (such as an annoying phone call or a pleasant conversation), though relationships (i_b) or (i_a) in Fig. 1b, respectively. Then, it should be acknowledged that the results of this prototype experiment are limited in linking trip characteristics with emotions.

The sample characteristics are presented in Table 2. For reporting purposes, a short name was defined for the categorical variables presented at the top of the table. It should be noted that the sample shares do not vary across the 5 and 10-seconds windows dataset, as they are constructed from the same original data. Regarding the Circumplex emotions, stressed, sad and relaxed have basically the same sample share, while neutral and happy are more uncommon emotions during the trip. The trip stages were grouped in four unique stages: walking to the bus stop/metro station, waiting for the bus/metro, riding the bus and at the metro (riding the metro or walking within the underground station). The statistics on the table show that in most pseudo-choice occasions, the individual is riding the bus. The main difference among datasets is the standardized EDA, which is higher for the 10-seconds window (EDA was measured, on average, every 0.12 s). Fig. 4 summarizes the estimation sample.

Table 2
Sample Characteristics.

Variable	Variable name	5-seconds window		10-seconds window	
		Observations	Share	Observations	Share
Circumplex Emotions					
Neutral	Neutral	184	13.4%	92	13.4%
Alert, excited, elated, happy	Happy	86	6.2%	43	6.2%
Tense, nervous, stressed, upset	Stressed	382	27.7%	190	27.6%
Sad, depressed, bored	Sad	387	28.1%	194	28.2%
Contented, serene, relaxed, calm	Relaxed	339	24.6%	170	24.7%
Trip Stages					
Walking to the bus stop/metro station	Walk	294	21.3%	148	21.5%
Waiting for the bus/metro	Wait	242	17.6%	121	17.6%
Riding the bus	Bus	633	45.9%	316	45.9%
At the metro	Metro	209	15.2%	104	15.1%
Descriptive statistics					
Variable	**Statistic**	**5-seconds window**		**10-seconds window**	
Standarized Electrodermal Activity (EDA)	Mean		3.69		4.39
	Std. Dev.		1.00		1.00
	Minimum		0.49		1.19
	Maximum		5.58		6.19

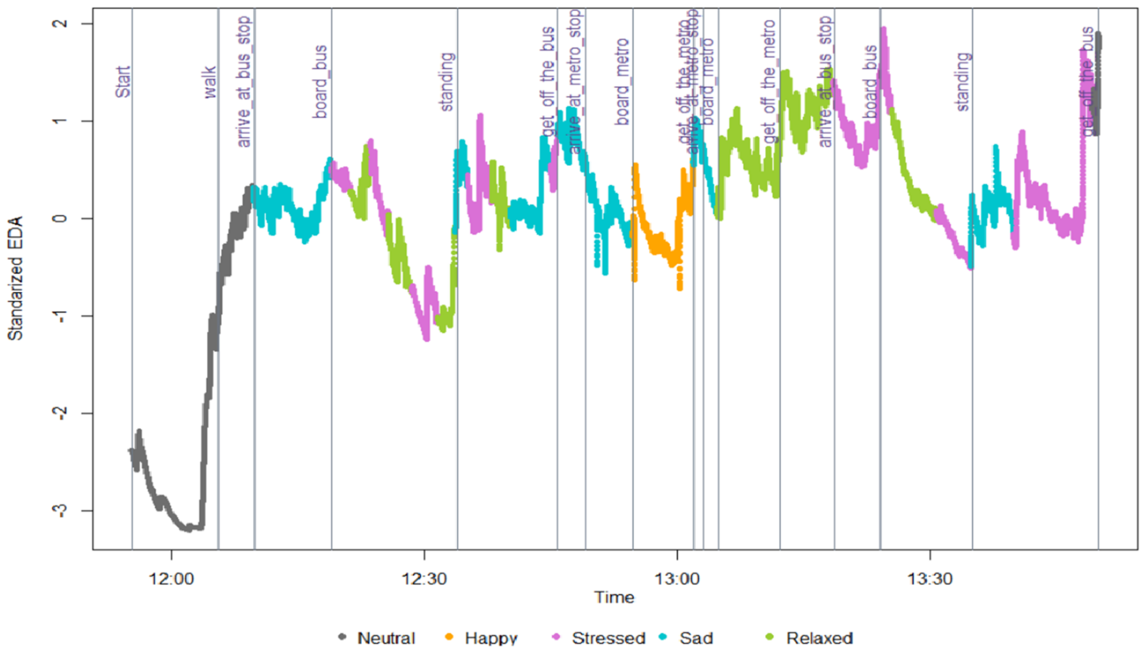


Fig. 4. Relationship between Emotions, EDA and Trip Stages over Time.

5.4. *Empirical results*

Table 3 presents the model estimation results, for both synthetic data samples. Both models include the emotion-specific constants and the dummy variables associated with trip stages. Only statistically significant variables ($\alpha < 5\%$) are presented. The second model also incorporates the psychophysiological indicator EDA, as described in Section 3.

Table 3 shows that walking to the bus stop reduces the probability of feeling stressed, sad and relaxed, while waiting for the bus increases the likelihood of feeling sad. This last result was also found by Gao et al. (2017), while the relationship between sadness and walking to the bus stop was identified by Ettema et al. (2011). Finally, it was found that EDA has a significant effect on all the emotions considered in the analysis.

To classify the emotional states during the journey based on the psychophysiological signals, a random forest model was used. The characteristics extracted for its execution correspond to the mean and the standard deviation of EDA, EDA phasic component, PPG,

Table 3
Model Estimation Results.

Model Variable	Neutral		Happy		Stressed		Sad		Relaxed	
	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
5-seconds window sample										
Emotion-specific constants	-	-	−3.903	−2.84	−7.159	−8.92	−5.465	−4.24	−7.724	0.80
Trip Stages										
Walk	-	-	-	-	−5.073	−7.17	−1.676	−3.99	−1.556	−3.37
Wait	–	–	–	–	–	–	2.345	10.87	–	–
Bus	–	–	–	–	–	–	–	–	–	–
Metro	–	–	–	–	–	–	–	–	0.787	3.70
Psychophysiological indicators										
EDA	–	–	1.098	3.04	2.515	12.49	1.897	6.01	2.515	12.49
10-seconds window sample										
Emotion-specific constants	–	–	−4.682	−2.02	−8.996	−6.74	−6.813	−3.18	−9.556	−7.16
Trip Stages										
Walk	–	–	–	–	−5.197	−0.04	−1.631	−2.76	−1.550	−2.35
Wait	–	–	–	–	–	–	2.321	7.68	–	–
Bus	–	–	–	–	–	–	–	–	–	–
Metro	–	–	–	–	–	–	–	–	0.792	2.64
Psychophysiological indicators										
EDA	–	–	1.102	2.13	2.535	8.85	1.905	4.24	2.535	8.85

Table 4
Performance Measurements for Out-Of-Sample Validation.

Sample	Model	Precision	Recall	F1 Score
5-seconds window	Random forest	77.4%	74.2%	75.5%
	Only constant	26.2%	20.0%	40.2%
	EV model	**51.5%**	**50.1%**	**60.5%**
10-seconds window	Random forest	71.8%	61.7%	64.1%
	Only constant	25.4%	19.9%	39.8%
	EV model	**53.7%**	**50.0%**	**59.9%**

HR and ST.

To evaluate the model's fit, we conducted a cross validation exercise using the K-fold method, where the original sample is randomly partitioned into k equal sized subsamples. For each k, the model is estimated using $k-1$ subsamples; the last subsample is used for out-of-sample validation.

Table 4 presents three performance measurements (precision, recall and the F1 score) for the K-fold method, with $k = 10$. The performance measurements of the proposed approach are compared to those of the random forest model. It should be noted that, given the nature of the random forest model, the performance measurements obtained for that model are the highest possible measurements attainable. For the proposed model, the results is Table 4 show a lower fit than the random forest model, but the fit is considered relatively high, validating the proposed approach.

6. Conclusions

The planning, evaluation and management of transportation services have been based almost exclusively on measures of travel time and cost, ignoring various relevant aspects because they are hard to measure with traditional methods of transportation data collection. These data collection tools are also limited by reporting and hypothetical bias. This research works toward closing this gap using indicators gathered from psycho-physiological sensors, given that empirical evidence has shown that they covariate among each other, and correlate with psycho-physiological states such as stress, cognitive load, and fatigue, among others.

With that goal in mind, this article conveys the development of a framework for random utility modeling by incorporating psychophysiological data extending the conceptual framework for modeling integrated choices by including the psychophysiological responses as additional indicators.

The proposed approach was assessed, enhanced and validated using first Monte Carlo simulations. Results show that the proposed methodological framework is feasible for the incorporation of psychophysiological indicators to the modeling of transportation choices, potentially enriching the understanding of the phenomena and the forecasting capabilities.

A prototype field experiment was designed that makes use of a Printed Circuit Board that integrates tiny psychophysiological sensors to capture electrodermal activity, heart rate, heart rate variation, temperature and acceleration. Although the purpose of this experiment is limited and the results cannot be generalized, it opens new possibilities in terms of data collection and modeling, as it serves to validate some critical components of the proposed framework. Future research can build upon the prototype experiment increasing sample size and/or studying other discrete outcomes, such as modal choice or route choice.

This paper aims to trigger a new research roadmap to create new models of behavior of public transportation users that allow to represent in greater degree aspects that influence satisfaction and choice, and to predict demand in different scenarios with greater precision the than the existing models.

This paper also aims to open new opportunities for incorporating new indicators of the satisfaction of public transportation users, and of causal factors on the physical, emotional and cognitive state of users. We envision a new methodology to measure, analyze and disseminate periodically indicators of user experiences of the entire transit system. This methodology could be integrated into the big data of public transport, at lower cost and more frequency than traditional methods. In addition, it could contribute to design and evaluate changes in infrastructure and public transportation services that increase social welfare.

As future work, the experiment will be replicated in a larger number of subjects to assess and enhance the proposed method for incorporating psychophysiological indicators into transportation analysis. Among other things, this will require the comprehensive development of a simplified circumplex model for transportation users, as well as new devices and methods for data collection and analysis.

CRediT authorship contribution statement

Marisol Castro: Methodology, Formal analysis, Software, Validation, Writing - review & editing, Visualization. **Angelo Guevara:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Supervision, Project administration, Funding acquisition. **Angel Jimenez-Molina:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - review & editing, Supervision, Resources, Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - review & editing, Supervision, Resources.

Acknowledgments

This research was partially funded by ANID, FONDECYT 1191104 and ANID, PIA/BASAL AFB180003. All models and data analysis were performed using the open source software R (R Development Core Team, 2008). The mobile application was developed over the open-source platform MIT App Inventor (Pokress and Veiga, 2013). Research assistantship from Cristian Retamal and Felipe Miranda are also greatly acknowledged.

References

Allanson, J., 2004. A research agenda for physiological computing. Interact. Comput. 16 (5), 857–878.

Ben-Akiva, M.E., Lerman, S.R., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand, vol. 9 MIT Press.

Bethel, C.L., 2007. Survey of psychophysiology measurements applied to human-robot interaction. In: Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on. IEEE, pp. 732–737.

Burt, K.B., Obradović, J., 2013. The construct of psychophysiological reactivity: Statistical and psychometric issues. Dev. Rev. 33 (1), 29–57.

Cacioppo J., L. T., 2007. Handbook of Psychophysiology. Cambridge University Press.

Carrel, A., Mishalani, R.G., Sengupta, R., Walker, J.L., 2016. In pursuit of the happy transit rider: dissecting satisfaction using daily surveys and tracking data. J. Intell. Transp. Syst. 20 (4), 345–362.

Cherchi, E., Guevara, C.A., 2012. A Monte Carlo experiment to analyze the curse of dimensionality in estimating random coefficients models with a full variance–-covariance matrix. Transp. Res. Part B: Methodol. 46 (2), 321–332.

Dhami, M., Hertwig, R., Hoffrage, U., 2004. The role of representative design in an ecological approach to cognition. Psychol. Bull. 130 (6), 959–988.

Ettema, D., Gärling, T., Eriksson, L., Friman, M., Olsson, L.E., Fujii, S., 2011. Satisfaction with travel and subjective well-being: Development and test of a measurement tool. Transp. Res. Part F: Traff. Psychol. Behav. 14 (3), 167–175.

Fernández-Antolín, A., Guevara, C.A., De Lapparent, M., Bierlaire, M., 2016. Correcting for endogeneity due to omitted attitudes: Empirical assessment of a modified MIS method using RP mode choice data. J. Choice Modell. 20, 1–15.

Fitch, D., Sharpnack, J., Handy, S., 2017. The road environment and bicyclists' psychophysiological stress. In: 6th Annual International Cycling Safety Conference.

Florian, M., 2008. Models and software for urban and regional transportation planning: the contributions of the center for research on transportation. INFOR: Inform. Syst. Operat. Res. 46 (1), 29–49.

Ganglbauer, E.S., 2009. Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. In: Workshop on user experience evaluation methods in product development. August 2009: User Experience Evaluation Methods in Product Development.

Gao, Y., Rasouli, S., Timmermans, H., Wang, Y., 2017. Effects of traveller's mood and personality on ratings of satisfaction with daily trip stages. Travel Behav. Soc. 7, 1–11.

Garbarino, M., Lai, M., Bender, D., Picard, R. W., Tognetti, S., 2014. Empatica E3—A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In: 2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH). IEEE, pp. 39–42.

Guevara, C.A., 2015. Critical assessment of five methods to correct for endogeneity in discrete-choice models. Transp. Res. Part A: Policy Pract. 82, 240–254.

Guevara, C.A., Polanco, D., 2016. Correcting for endogeneity due to omitted attributes in discrete-choice models: the multiple indicator solution. Transportmetr. A: Transp. Sci. 12 (5), 458–478.

Guevara, C.A., Tirachini, A., Hurtubia, R., Dekker, T., 2020. Correcting for endogeneity due to omitted crowding in public transport choice using the Multiple Indicator Solution (MIS) method. Transp. Res. Part A: Policy Pract. 137, 472–484.

Haapalainen, E., Kim, S., Forlizzi, J.F., Dey, A.K., 2010. Psychophysiological measures for assessing cognitive load. In: Proceedings of the 12th ACM International Conference on Ubiquitous Computing, Copenhagen, Denmark, 26–29 September, pp. 301–310.

Hogertz, C., 2010. Emotions of the urban pedestrian: sensory mapping. Pedestrians' Qual. Needs 31, 31–52.

iMotions Biometric Research Platform, 2016. GSR Pocket Guide; iMotions Biometric Research Platform: Boston, MA, USA.

Jimenez-Molina, A., Retamal, C., Lira, H., 2018. Using Psychophysiological Sensors to Assess Mental Workload During Web Browsing. Sensors 18, 458.

Kivikangas, J.M., 2011. Review on psychophysiological methods in game research. J. Gaming Virtual Worlds 3 (3), 181–199.

Kramer, A.F., 1991. Physiological metrics of mental workload: a review of recent progress. Multiple-task Perform. 279–328.

Li, W., Chung, W., 2013. Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier. Sensors 13, 16494–16511.

Li, Z., Hensher, D.A., 2011. Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. Transp. Policy 18 (6), 880–887.

Lundqvist, L., Mattsson, L.G., 2002. National transport models: recent developments and prospects. Springer Science & Business Media.

Mariel, P., Hoyos, D., Artabe, A., Guevara, C.A., 2018. A multiple indicator solution approach to endogeneity in discrete-choice models for environmental valuation. Sci. Total Environ. 633, 967–980.

Morris, E.A., Guerra, E., 2015. Mood and mode: does how we travel affect how we feel? Transportation 42 (1), 25–43.

Palma, D., de Dios Ortúzar, J., Rizzi, L.I., Guevara, C.A., Casaubon, G., Ma, H., 2016. Modelling choice when price is a cue for quality: a case study with Chinese consumers. J. Choice Modell. 19, 24–39.

Park, B., 2009. Psychophysiology as a tool for HCI research: promises and pitfalls. In: International Conference on Human-Computer Interaction. Springer, Berlin Heidelberg, San Diego, CA, pp. 141–148.

Parsons, H.M., 1974. What Happened at Hawthorne? New evidence suggests the Hawthorne effect resulted from operant reinforcement contingencies. Science 183, 922–932.

Posner, J., Russell, J.A., Peterson, B.S., 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev. Psychopathol. 17 (3), 715–734.

Pokress, S.C., Veiga, J.J.D., 2013. MIT App Inventor: Enabling personal mobile computing. arXiv preprint arXiv:1310.2830.

R Development Core Team, 2008. R: A language and environment for statistical computing. Vienna, Austria. Downloaded from http://www.R-project.org (ISBN 3-900051-07-0).

Raveau, S., Guo, Z., Muñoz, J.C., Wilson, N.H., 2014. A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. Transp. Res. Part A: Policy Pract. 66, 185–195.

Rossetti, T., Guevara, C.A., Galilea, P., Hurtubia, R., 2018. Modeling safety as a perceptual latent variable to assess cycling infrastructure. Transp. Res. Part A: Policy Pract. 111, 252–265.

Russell, J.A., 1980. A circumplex model of affect. J. Pers. Soc. Psychol. 39 (6), 1161.

Sectra, 2013. Manual de diseño y evaluación social de proyectos de vialidad urbana (MESPIVU). Ministerio de Desarrollo Social, Chile. < http://www.sectra.gob.cl/metodologias/mespivu.htm > .

Sharma, N., Gedeon, T., 2012. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. Comput. Methods Programs Biomed. 108 (3), 1287–1301.

Shoval, N., Schvimer, Y., Tamir, M., 2018. Tracking technologies and urban analysis: Adding the emotional dimension. Cities 72, 34–42.

Tirachini, A., Hensher, D.A., Rose, J.M., 2013. Crowding in public transport systems: effects on users, operation and implications for the estimation of demand. Transp.

Res. Part A: Policy Pract. 53, 36–52.
Tirachini, A., Hurtubia, R., Dekker, T., Daziano, R.A., 2017. Estimation of crowding discomfort in public transport: results from Santiago de Chile. Transp. Res. Part A: Policy Pract. 103, 311–326.
Tirachini, A., Sun, L., Erath, A., Chakirov, A., 2016. Valuation of sitting and standing in metro trains using revealed preferences. Transp. Policy 47, 94–104.
Vij, A., Carrel, A., Walker, J.L., 2013. Incorporating the influence of latent modal preferences on travel mode choice behavior. Transp. Res. Part A: Policy Pract. 54, 164–178.
Villarejo, M.V., Zapirain, B.G., Zorrilla, A.M., 2012. A stress sensor based on Galvanic Skin Response (GSR) controlled by ZigBee. Sensors 12, 6075–6101.
Walker, J., 2001. Extended discrete choice models: integrated framework, flexible error structures, and latent variables. Ph.D. Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.
Walker, J., Ben-Akiva, M., 2002. Generalized random utility model. Math. Soc. Sci. 43 (3), 303–343.
Wener, R.E., Evans, G.W., Phillips, D., Nadler, N., 2003. Running for the 7: 45: The effects of public transit improvements on commuter stress. Transportation 30 (2), 203–220.
Ye, Y., He, W., Cheng, Y., Huang, W., Zhang, Z., 2017. A robust random forest-based approach for heart rate monitoring using photoplethysmography signal contaminated by intense motion artifacts. Sensors 17, 385.