An artificial neural network based approach to investigate travellers' decision rules

van Cranenburgh, Sander; Alwosheel, Ahmad

DOI
10.1016/j.trc.2018.11.014

Publication date
2019

Document Version
Final published version

Published in
Transportation Research. Part C: Emerging Technologies

Citation (APA)
van Cranenburgh, S., & Alwosheel, A. (2019). An artificial neural network based approach to investigate travellers' decision rules. *Transportation Research. Part C: Emerging Technologies*, *98*, 152-166. https://doi.org/10.1016/j.trc.2018.11.014

Transportation Research Part C 98 (2019) 152–166

Contents lists available at ScienceDirect

Transportation Research Part C

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

An artificial neural network based approach to investigate travellers' decision rules

Sander van Cranenburgh*, Ahmad Alwosheel

Transport and Logistics Group, Delft University of Technology, the Netherlands

ARTICLE INFO

Keywords:
Decision rule
Artificial neural network
Latent class
Discrete choice

ABSTRACT

This study develops a novel Artificial Neural Network (ANN) based approach to investigate decision rule heterogeneity amongst travellers. This complements earlier work on decision rule heterogeneity based on Latent Class discrete choice models. We train our ANN to recognise the choice patterns of four distinct decision rules: Random Utility Maximisation, Random Regret Minimisation, Lexicographic, and Random. Next, we apply our trained ANN to classify the respondents from a recent Value-of-Time Stated Choice experiment in terms of their most likely employed decision rule. We cross-validate our findings by comparing our results with those from: (1) single class discrete choice models estimated on subsets of the data, and (2) latent class discrete choice models. The cross-validations provide strong support for the notion that ANNs can be used to identify underlying decision rules in choice data. As such, we believe that ANNs provide a valuable addition to the toolbox of analysts who wish to investigate decision rule heterogeneity. The substantive contribution of this study is that we provide strong empirical evidence for the presence of decision rule heterogeneity amongst travellers.

1. Introduction

Artificial Neural Networks (ANNs) are gaining increasing popularity in many research fields, including in transportation (e.g. Hensher and Ton, 2000; Mohammadian and Miller, 2002; van Lint et al., 2005; Omrani et al., 2013; Borysov et al., 2016; Alwosheel et al., 2017; Wong et al., 2017). ANNs are mathematical models which are loosely inspired by the structure and functional aspects of biological neural systems. Their recent uptake can be explained by major breakthroughs in ANN research, affecting the daily lives of many people (such as e.g. in the context of natural language processing or facial recognition), in combination with the rise of emerging data (Vlahogianni et al., 2015; Chen et al., 2016). Particularly, the versatile architecture of ANNs makes them well-equipped to deal with large volumes of (unstructured) emerging data (Maren et al., 2014).

However, despite the general excitement in the field of transportation about the potential of ANN (and other data-oriented techniques), the number of areas of application of ANNs in transport, and in particular in the travel behaviour research subfield, is still fairly limited. Currently, ANNs are predominantly used to analyse observed movement patterns and to make short-term travel demand predictions. However, a recent paper in this journal (Chen et al., 2016) advocates to move beyond this state, and use these data-oriented techniques to improve understanding of the travel behaviour underlying human mobility patterns. In particular, Chen et al., suggest using data-oriented methods to identify factors explaining travel decisions and to uncover underlying decision rules.

This paper answers to this call: it develops an ANN based approach to investigate travellers' decision rules. Decision rules are the decision mechanisms humans use when making choices (Payne et al., 1993). Decision rules are widespread in transportation research

* Corresponding author.
E-mail address: s.vancranenburgh@tudelft.nl (S. van Cranenburgh).

https://doi.org/10.1016/j.trc.2018.11.014
Received 19 September 2017; Received in revised form 26 September 2018; Accepted 26 November 2018
Available online 03 December 2018
0968-090X/

Notation

C	Set of alternatives	β_m	Taste parameter associated with attribute m
i, j	Choice alternatives in C	x_{mi}	Attribute level of the m'th attribute of alternative i
y_i	Indicator which denotes whether alternative i is chosen	P_i	Choice probability of alternative i
U_i	Total utility of alternative i	RR_i	Total regret of alternative i
V_i	Observed part of utility of alternative i	R_i	Observed part of regret of alternative i
ε_i	Unobserved part of utility or regret of alternative i	$\overline{m}$	The most important of the M attributes
		J	Cardinality of the choice set
		$x_{j\overline{m}}$	Attribute level of alternative j for the attribute $\overline{m}$

as they are embedded in discrete choice models. Although the vast majority of discrete choice models are built on a single decision rule (random utility maximisation), there is a growing recognition amongst transport researchers that travellers are heterogeneous in terms of their decision rules. Also, it is increasingly acknowledged that insights on decision rule heterogeneity are crucial for understanding and predicting travel behaviour. In this context, a number of recent studies have explored decision rule heterogeneity amongst travellers (using traditional discrete choice models) (Hess et al., 2012; Leong and Hensher, 2012; Hess and Chorus, 2015; Balbontin et al., 2017; Boeri and Longo, 2017; Gonzalez-Valdes and Raveau, 2018).

Specifically, in this study we develop a novel pattern recognition Artificial Neural Network (ANN) to classify travellers in terms of their most likely employed decision rules based on observed sequence of choices. By doing so, this study aims to shed new light on decision rule heterogeneity amongst travellers. In order to detect patterns in the data ANNs need to be trained based on so-called training data, which includes correct classifications. However, in the context of decision rules the 'true' decision rule is inherently unknown. In fact, decision rules should rather be perceived as quintessential models explaining choice behaviour in a parsimonious way than as models that accurately reflect the complex decision-making processes (Chorus, 2014). Therefore, in the absence of real-world training data consisting of the 'correct' classifications, we train our ANN using synthetic data. The synthetic decision-makers which we created for training are heterogeneous not only in terms of their employed decision rules, but also in terms of their preferences. By doing so, the ANN is trained to classify travellers in terms of their decision rules in the realistic condition in which besides decision rule heterogeneity also taste heterogeneity and heteroscedasticity are present. Finally, we apply our trained ANN to classify the respondents from a recent Value-of-Time (VoT) Stated Choice (SC) experiment, and cross-validate its classifications using traditional discrete choice analysis.

The methodological contribution of this paper is that it is the first to show how ANNs can be employed to investigate travellers' decision rule heterogeneity. We present a novel ANN topology that is particularly suited to deal with sequences of choice observations, and show how to train it using synthetic data. As such, this research complements earlier work on decision rule heterogeneity based on Latent Class discrete choice models. The substantive contribution of this study is that we provide new, strong, empirical evidence for the presence of decision rule heterogeneity amongst travellers.

The remainder of this paper is organised as follows. Section 2 first presents the empirical data that we aim to analyse using the ANN based approach. Section 3 develops the ANN and applies it to the empirical data set. In Section 4 we cross-validate our results obtained from the ANN by comparing them with results obtained using traditional LC discrete choice models. Finally, Section 5 draws conclusions and provides a discussion.

2. Data

For this study we use data from a relatively small VoT SC experiment.[1] We choose this data set because of its simplicity. Its simplicity provides a degree of tractability on the underlying decision-making mechanisms used by the respondents. In these data we can, for instance, straightforwardly identify the compromise alternative[2] and we can easily detect non-trading behaviour. This allows us latter on to cross-validate the results of our ANN.

Fig. 1 shows a screenshot of the first choice tasks presented to respondents in the SC experiment. Choice tasks in this experiment consist of three unlabelled route alternatives, each consisting of two generic attributes: Travel Cost (TC) and Travel Time (TT). Attribute levels are selected as follows: the range of the travel times was chosen such that they are in consonance with the range of the travel times presented in previous European VoT SC-experiments–. The minimum travel time is set at 23 min, and the maximum at 35 min, with equally spaced 4 min intervals. In this experiment respondents were presented $T = 10$ choice tasks. To optimise the statistical efficiency of the experiment, a so-called D-efficient design is used. The statistical efficiency of the experimental design was optimised for a combination of RUM and P-RRM models, see Van Cranenburgh et al. (2018) for more details. The complete experimental design can be found in Appendix A.

[1] To avoid any misunderstanding, we note upfront that these data are not used for the training ANNs. These data would be too small to do so (Van Cranenburgh and Chorus, 2018).

[2] Compromise alternatives have an intermediate performance on each or most attributes (relative to other alternatives in the choice set) rather than having a poor performance on some attributes and a strong performance.

	Route A	Route B	Route C
Travel time (one-way)	23 minutes	27 minutes	35 minutes
Travel cost (one-way)	€ 6	€ 4	€ 3

Fig. 1. Screenshot of 1st choice task (translated to English).

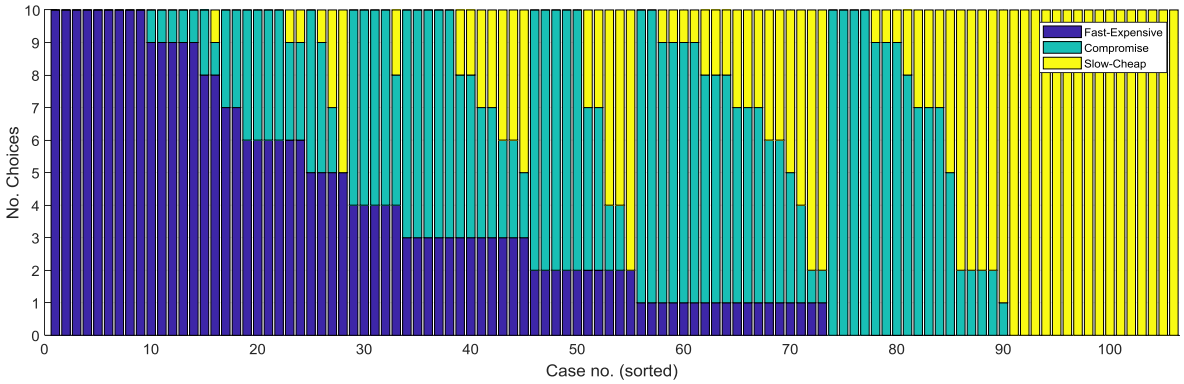

Fig. 2. Choices of respondents.

2.1. Data collection

The data collection took place in The Netherlands in May 2016. To recruit respondents a panel company was used (TNS-NIPO). Only car commuters were admitted to conduct the choice experiment. In total 106 respondents completed the full survey. A relatively balanced sample has been obtained in terms of gender, age, education and income. The sample statistics can be found in Appendix B. See Van Cranenburgh et al. (2018) for more details on the data collection.[3]

2.2. Decision rules

Close inspection of the choice data allows us to derive indications on the decision rules that may have been used by the respondents in this route choice experiment. In each choice task respondents could choose between a 'fast and expensive', a compromise, and a 'slow and cheap' alternative, although they were not explicitly labelled as such in the SC experiment. Across all choice observations, in 30 per cent of the cases the 'fast and expensive' alternative is chosen; in 38 per cent of the cases the compromise alternative is chosen; and in 32 per cent of the cases the 'slow and cheap' alternative is chosen. The observation that the compromise alternative acquires the highest market share provides an indication that an RRM decision rule may have been used by a considerable share of the respondents. After all, RRM models predict a market share bonus for the compromise alternative when compared to Random Utility Maximisation (RUM) (Chorus and Bierlaire, 2013; Guevara and Fukushi, 2016).

The stacked bar graph in Fig. 2 provides more insights on the distribution of respondents' choices. It visualises for each of the 106 respondents in the data set the number of times the 'fast and expensive', the compromise and the 'slow and cheap' alternative has been chosen (note that we sorted the respondents on the x-axis for the sake of clarity). The bar graph reveals that 9 respondents consistently choose for the 'fast and expensive' alternative (full blue bars on the left) and 16 respondents who consistently chose for the 'slow and cheap' alternative (full yellow bars on the right). This suggests that a substantial share of the respondents opted for a lexicographic decision rule.[4]

A substantial share of the respondents (77) switched between alternatives across the ten choice tasks. Of these 77 respondents, 29 respondents consistently chose either the 'fast and expensive' alternative or the compromise alternative (blue-green bars) and 13 respondents consistently chose either the 'slow and cheap' alternative or the compromise alternative (yellow-green bars). This signals a degree of rationality underlying the trade-offs, which seems consistent with RUM. Thirty-three respondents chose all three alternatives at least once (tri-coloured bars). Close inspection of the choices of these respondents using the half-space method (Rouwendal et al., 2010) reveals that their choices do not suggest a stable underlying VoT – at least not from a RUM modelling perspective. In addition, two respondents even chose either the 'fast and expensive' alternative or the 'slow and cheap' alternative. This observation suggests that a substantial share of respondents made (seemingly) random choices.

[3] The data can be retrieved from http://doi.org/10.4121/uuid:1ccca375-68ca-4cb6-8fc0-926712f50404.

[4] Under a lexicographic decision rule a decision-maker first evaluates the alternatives, then identifies the most important attribute and subsequently chooses the alternative that performs best in terms of this particular attribute.

Table 1
Decision rules.

Decision rule	Choice rule	Total utility/regret	Value function	MNL choice probability
Utility maximisation	$y_i = 1 \Leftrightarrow U_i \geqslant U_j \ \forall j \in C$	$U_i = V_i + \varepsilon_i$	$V_i = \sum_m \beta_m x_{im}$	$P_i = \frac{e^{V_i}}{\sum_{j \in C} e^{V_j}}$
P-RRM	$y_i = 1 \Leftrightarrow RR_i \leqslant RR_j \ \forall j \in C$	$RR_i = R_i + \varepsilon_i$	$R_i = \sum_{j \neq i} \sum_m \max(0, \beta_m [x_{jm} - x_{im}])$	$P_i = \frac{e^{-R_i}}{\sum_{j \in C} e^{-R_j}}$
Lexicographic	$y_i = 1 \Leftrightarrow x_{i\overline{m}} \geqslant x_{j\overline{m}} \ \forall j \in C$	N/A		N/A
Random	$y_i = 1 \Leftrightarrow U_i \geqslant U_j \ \forall j \in C$	$U_i = V_i + \varepsilon_i$	$V_i = 0 \ \forall i \in C$	$P_i = \frac{1}{J}$

Based on these descriptive analyses we obtain first indications that the following four decision rules may be present in our data: RUM, RRM, Lexicographic, and Random. In the remaining part of this paper we will focus on these four decision rules. The latter decision rule 'Random' may seem a bit odd, in the sense that random behaviour is typically not considered a decision rule and is therefore typically not explicitly accounted for in discrete choice studies. However, seasoned SC researchers know that in SC data typically a considerable share of respondents makes (seemingly) random choices. Therefore, we treat random choice behaviour as a separate decision rule in the context of this study. Table 1 shows the mathematical formulations of these decision rules as well as their implementation in a discrete choice modelling framework. Note that the RRM model we use in this study is the so-called P-RRM model (Van Cranenburgh et al., 2015a). This model is increasingly used in the RRM literature as yields the strongest regret minimization behaviour, i.e., the highest level of regret aversion which is possible, within the RRM modelling framework. As such, this RRM model postulates choice behaviour which is strongly different from RUM – which intuitively should make it easier to distinguish between these two decision rules.

3. An artificial neural network based approach

Section 3.1 and Section 3.2 discuss the development of the ANN for decision rule classification. Next, Sections 3.3 and 3.4 present the training data and the performance of the trained network. Finally, in Section 3.5 we elaborate on how to employ the trained network to classify travellers in empirical data, and apply it to our empirical VoT data to classify the respondents.

3.1. Artificial neural networks

Artificial neural networks are inspired by the structure and functional aspects of biological neural systems. ANNs originate from the field of neuro and computer sciences, but are currently rapidly spreading out to other research disciplines (Maren et al., 2014). Underlying this rapid expansion is the emergence of so-called Big Data. The combination of large volumes of (unstructured) data on the one hand and the versatile architecture of ANNs on the other hand has led to numerous ground-breaking results in a variety of disciplines, such as speech recognition, gene detection of autism and natural language processing.

Computations in ANNs are structured in terms of interconnected groups of artificial neurons, processing information using a so-called connectionist approach (Bishop 1995). ANNs are composed of nodes. Three types of nodes are commonly distinguished: input nodes, hidden nodes and output nodes. The input nodes contain the explanatory variables. In the context of choice models, these typically concern the attribute levels of the alternatives. The output nodes contain the dependent variables. In the context of choice models, e.g. i.e. when predicting choices, the output nodes consist of the choice probabilities. The signals propagate in forward direction, through the links which connect the nodes. The links have a numeric weight w, which needs to be learned from the data. At each node the weights are multiplied with the input value from the previous nodes and summed. Then the signal is propagated to the next layer using an activation function. Commonly used activation functions are tan-sigmoid, softmax and purelin. See Bishop (1995) for an extensive overview of ANNs and their characteristics.

Despite the fact that an extensive variety of ANNs have been developed (Maren et al., 2014) to tackle all sorts of (classification) problems – each of which with strengths particular to their application –, to the best of the authors' knowledge no type of ANN has been put forward that is particularly suited to investigate decision rule heterogeneity. Our classification problem is somewhat unconventional as we aim to make a classification based on an *unordered* sequence of correlated (choice) observations. Note that this is conceptually different from the more commonly encountered ordered sequence-to-sequence classification problem. To classify a traveller in terms of his or her (most likely) employed decision rule we need to assess a sequence of choice observations made by a traveller. After all, based on a single choice observation of a traveller it is virtually impossible to make such a classification, as any single choice could be driven by any decision rule (considering some randomness is present). In contrast, a sequence of choice observations may provide a 'fingerprint' of what is the most likely employed decision rule. However, for our classification problem the order of the observations within the sequence is irrelevant. Essentially, what we want is a network that classifies decision-makers to decision rules, having seen the full sequence of observations of that a decision-maker, regardless of the order in which the observations are presented to the network. Therefore, we cannot use ANN types that are specifically designed to capture serial

correlations, such as the Recurrent ANN. In the absence of a suitable 'off-the-shelf' ANN, the next subsection proposes a new ANN topology that is particularly designed for classification of travellers' decision rules (and the data set presented in Section 2).

3.2. An artificial neural network for decision rule classification

ANNs are often pitched as a generic method that can be used to model any problem. However, this is a misconception: much of the art of machine learning is determining how to incorporate problem specific knowledge into the ANN via its topology, performance function, activation functions, etc. (Maren et al., 2014).

To develop an ANN capable to classify decision-makers to decision rules, we have tested different types of ANNs (Multi-layer Perceptron and LSTM) and experimented with different topologies, number of hidden layers, activation functions as well as the degree of connectivity between hidden layers. We compared the networks in terms of their classification performance, while considering their complexity (in terms of e.g. number of layers and number of nodes at each layer).

Fig. 3 shows the topology of our best performing ANN.[5] The ANN is a Multi-Layer Perceptron (MLP). A key characteristic of the ANN topology is that it processes sequences of choice tasks made by a decision-maker as one chunk of input data (note that the input layer contains all ten choice tasks including the observed choice). Hence, the model treats a sequence of choices of a traveller as one independent observation. This is crucial as the choice observations belonging to the same individual are correlated and therefore cannot be treated as independent observations.

The ANN's topology is behaviourally informed in the sense that behavioural intuition is added to the network to help it grasp the data structure and classification problem. In particular, the ANN is deliberately sparsely connected, i.e., many nodes are *not* connected to one another. For instance, the input nodes of one choice task are not connected with the hidden nodes of other choice tasks. Behaviourally, this makes sense because the attributes (input nodes) presented to a decision-maker in e.g. choice task $t = 10$ simply cannot affect the choice in choice task $t = 1$ (i.e., unless the respondent is allowed to revise his or her choices, which is usually not the case in SC experiments). On the other hand, the choice in choice task $t = 1$ could affect the choice in e.g. the next choice task. However, our aim in this paper is to uncover decision rules based on a sequence of choice observations of an individual, regardless of the order in which they are presented to the network. Therefore, we actually want to prevent the network to pick up learning and inertia effects (if present). Furthermore, in a sense, the first hidden layer takes care of assigning 'values' to alternatives and combining these with the observed choice, which are then passed on to the second hidden layer where the sequence as a whole is processed to make the decision rule classification. From this perspective, the full connection between the second hidden layer and the output layer is intuitive as it is the sequence of choice observations that is informative on the employed decision rule.

On a more technical level, our network uses two hidden layers. We find that adding more hidden layers does not improve the performance, while removing one layer deteriorates the performance drastically. At the first hidden layer (for each choice task) as well as at the second hidden layer we use four nodes. This is found to result in the best performance. Decreasing the number of nodes (either at the first or second hidden layer) decreases the classification performance noticeably, while increasing the number of nodes beyond four does not seem to improve performance (while consuming considerably larger number of weights than needed). Also note that no bias nodes are present in the network (see Fig. 3). During training the order of the alternatives and the order of the choice sets are fully shuffled (see Section 3.3), meaning that bias nodes become superfluous. The total number of weights of the network is 496. Furthermore, our network uses tan-sigmoid activation functions (LeCun et al., 2012) at the nodes of the hidden layers. Other types of activation function are found to perform worse, or result in longer training times. At the nodes of the output layer we however use softmax activation functions. This ensures that the sum of the decision rule classification probabilities adds up to 1.

3.3. Training data

Training the ANN involves exposing it to data containing the correct classifications. However, as noted before, in the context of human choice behaviour the 'true' decision rule is inherently unknown. In fact, decision rules should rather be seen as quintessential mathematical models representing highly complex decision processes. To deal with the fact that we do not have real data containing the 'true' decision rule, we train our ANN using synthetic data. These data are created using the decision rules shown in Table 1.

The training data set consists of 40,000 synthetic decision-makers; 10,000 decision-makers for each decision rule. Hence, the training data set is fully balanced. In consonance with the VoT data that we aim to analyse (see Section 2) each synthetic decision-maker is confronted with $T = 10$ choice tasks. The choice tasks are the same as those used in the empirical data collection (see Appendix A).

Given that synthetic data are in unlimited supply, the number of synthetic decision-makers is deliberately set high. A commonly used rule-of-thumb in machine learning is that the sample size needs to be (at least) 10 times larger than the number of estimable weights in the network (Haykin et al., 2009). A recent study specifically dealing with sample size requirements for using ANNs in the context of choice models is more conservative, and recommends to use a sample size of (at least) 50 times the number of estimable weights (Alwosheel et al., 2018). By using 40,000 observations in our study, our sample size is a comfortable 80 times larger than the number of estimable weights. Thereby, we safely avoid overfitting issues.

[5] Note that the topology of the ANN in Fig. 3 is specifically tweaked in terms of the number of input nodes and the number of decision-rules (output nodes) to match the structure of the empirical data that we aim to analyse (see Section 2). However, it can easily be adjusted to fit other data sets.

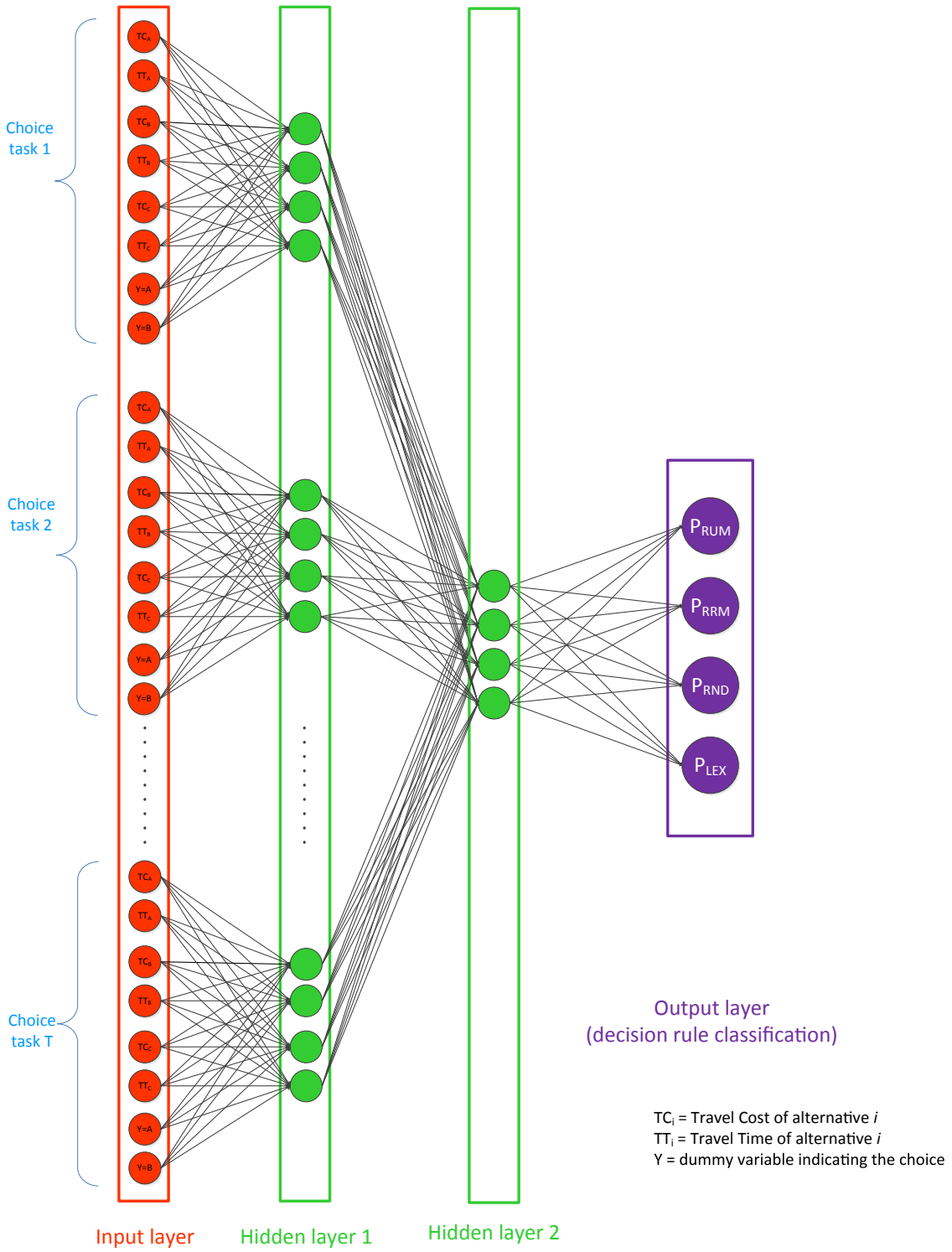

Fig. 3. ANN topology.

Importantly, besides decision rule heterogeneity, in the empirical data that we aim to analyse there are two other potential sources of correlation across the sequence of choice observations of individuals. Firstly, decision-makers can be heterogeneous in their preferences. Such taste heterogeneity creates correlation in the sequence of choices of an individual because they are generated using the same set of individual specific parameters. Secondly, decision-makers may learn from their earlier made choices, creating

serial correlation across the observations. For instance, the respondent could choose the fast and expensive alternative in, say, choice task 5 because he/she also chose the fast and expensive alternative in the choice tasks before, and he/she would like to stick to that choice.

During training we need to account for these sources of correlation, such that the trained network is able to detect the correlation specifically caused by the decision rule (and becomes capable to accurately classify decision-makers in terms of their employed decision rule). To account for learning effects, during the training stage we shuffle the order of the choice tasks and alternatives in the training data. Thereby, we prevent the network from learning (1) the (fixed) structure of the data and (2) ordering effects caused by learning or inertia (if present). Our synthetic decision-makers do not have the ability to learn, meaning that there are no learning effects to capture during training. Still, shuffling the order of alternatives and choice sets at the training stage is important as not doing so would preclude doing so in the application stage (where it is important). Then, when we apply the network to classify respondents in the empirical data set (Section 3.5), we also shuffle – at the level of the individual – the order of the choice tasks and the order of the alternatives. By doing so, serial-correlation in the empirical data is removed (if present).

To account for heterogeneity in preferences, we take a different approach. We created synthetic decision-makers which are heterogeneous in their preferences. By doing so, the ANN is trained to classify travellers in terms of their decision rules in the realistic condition in which both taste heterogeneity and heteroscedasticity are present. In the context of the decision rules considered in this study, taste heterogeneity is only relevant for RUM and RRM decision rules, as these decision rules contain taste parameters governing the decision-making process. Specifically, for both RUM and RRM decision-makers we assume that tastes for travel cost and travel time are symmetrically triangular distributed across decision-makers. Other distributions (normal and uniform) have been tested, but gave by and large similar results.

Accordingly, to generate the RUM and RRM choices every synthetic decision-maker is attributed two independent draws from the associated triangular densities for the marginal utilities/regrets of cost and time. Pseudo-random draws are generated from the Extreme Value Type I distribution for every alternative in the choice task. For RUM decision-makers, the alternative with the highest total utility is chosen; for RRM decision-makers the alternative with the minimum total regret is chosen. For β_{Cost} the lower bound of the distribution is set at $a = -3.2$, the mean at -1.6 and the upper bound is set at $c = 0$; for β_{Time} the lower bound is set at $a = -0.8$, the mean at -0.6 and the upper bound is set at $c = 0$. The means of β_{Cost} and β_{Time} are chosen based results from a LC discrete choice analysis conducted prior to training the ANN. However, we also tested larger and smaller values and found that these do not substantially influence results. For the RRM model a choice set size correction factor of 2/3 is applied to the parameters, see Van Cranenburgh et al. (2015b). Thereby, it is ensured that the using the same parameterisation for RUM and RRM corresponds to approximately the same degree of choice consistency. This avoids that the ANN learns that relatively deterministic choice behaviour is specific for RUM while relatively Random choice behaviour is specific for RRM, or vice versa.

To generate the Lexicographic choices, for 5000 decision-makers we set the choice to be the fastest route, and for the other 5000 decision-makers we set the choice to be the cheapest route. For this decision rule no randomness is added. To generate the Random choices, we simply took a draw z from the standard uniform distribution for each choice task. After all, random choice behaviour implies equal choice probabilities across all alternatives. In case $z < 1/3$ then the choice is set to alternative A, in case$1/3 < z < 2/3$the choice is set to alternative B, and in case $z > 2/3$the choice is set to alternative C.

3.4. Performance and cross validation

The ANN is implemented in MATLAB2017.[6] Prior to training, the data are normalised to minimise training time and reduce the probability of ending-up with suboptimal solutions. To train the ANN Levenberg-Marquardt backpropagation is used. This training algorithm is built-in and found to work well. Particular advantages of this algorithm over other training algorithms are that it is computationally relatively fast and requires relatively little memory. Training the ANN takes a few minutes using a desktop PC with six CPUs.

To test the capability of the ANN to classify decision-makers we use the so-called k-fold cross-validation method, with $k = 10$. With 40,000 synthetic decision-makers, the data set is split into 10 folds of 4000 randomly selected decision-makers, although we made sure that in each fold the four decision rules are equally represented. In each repetition of the training and testing, the data of 36,000 decision-makers are used for training and the data of 4000 decision-makers are used as a holdout set for testing. Holdout sets are selected such that their union over all repetitions is the entire training set. By doing so, the data of every decision-maker is guaranteed to be part of the training and testing data. The trained network is eventually applied in Section 3.5 to classify the respondents in the empirical data.

Table 2 shows the k-fold confusion matrix. Using the trained network each decision-maker in the test data sets are assigned to a decision rule based on the highest classification probability. This assignment is then compared to the true classification. More specifically, the cells on the diagonal show the mean percentage of the decision-makers that are correctly classified, across the 10 folds. The off-diagonal cells show the mean percentage of decision-makers that are misclassified into a certain decision rule, across the 10 folds. In parenthesis below the means values, the standard deviation is reported.

Based on Table 2 a number of inferences can be made. Firstly, the percentage correctly classified decision-makers is fairly good. It ranges between 54 (for RUM) to 99.8 per cent (for Lexicographic). Altogether, across the 10 folds between 73 and 76 per cent of the

[6] Code is available upon request from the first author.

Table 2
Classification of based on highest likelihood.

True DGP	ANN classification [%]				
	RUM	RRM	Lexicographic	Random	∑
RUM	54.6(3.2)	16.1(1.9)	9.7(1.1)	19.7(3.4)	100%
RRM	18.7(2.3)	67(1.9)	1.1(0.3)	13.2(1.7)	100%
Lexicographic	0.2(0.2)	0.0(0.0)	99.8(0.2)	0.1(0.1)	100%
Random	12.5(1.4)	9.9(1.4)	0.9(0.5)	76.8(2.5)	100%

Table 3
Classification based on highest likelihood $N = 106$.

Decision rule	RUM	P-RRM	Lexicographic	Random
No. respondents	22	51	27	6

decision-makers are correctly classified by the trained ANN (with an average of 75 per cent). Finally, the small standard deviations show that the classification is relatively stable across the folds.

The fact that not all decision-makers are correctly classified may, at first sight, seem not very promising. However, this was actually to be expected for two reasons. One reason is that when generating choices[7] we explicitly add randomness to account for unobserved factors on the side of the analyst. Therefore, in principle, any generated sequence of choices may just by coincidence appear to be generated by a certain decision rule, while it was generated by another. Another reason is that for certain parameterisations of the decision rules the implied choice behaviour becomes indistinguishable. Specifically, the taste parameters for RUM and RRM are drawn from symmetric triangular distributions. These distributions have mass very close to zero, meaning that individuals can have taste parameters that will generate choice behaviour that is virtually indistinguishable from Random choice behaviour. Likewise, the taste parameters may have been drawn such that one attribute is relatively far more important than the other, for instance, in case $\beta_{\text{Cost}} = -3$ and $\beta_{\text{Time}} = -0.01$. The resulting choice behaviour is then virtually indistinguishable from lexicographic choice behaviour. After all, with extreme ratios of parameters a decision-maker will always choose for either the cheapest or the fastest alternative. On a more general note, the fact that not all decision-makers are correctly classified highlights that classification of decision-makers to decision rules based on small finite sequence of choices – in the presence of randomness and taste heterogeneity – is inherently a hard task.

Close inspection of Table 2 shows that the ANN has most difficulty with distinguishing between RUM and RRM decision rules. 16.1 per cent of the RUM decision-makers are misclassified as RRM and 18.7 per cent of the RRM decision-makers are misclassified as RUM. This highlights that RUM and RRM differ from one another in rather subtle ways, even though we used the P-RRM model – which imposes very strong regret minimization behaviour.

3.5. Application to empirical data

Next, we use our trained ANN to classify the 106 respondents in the data set presented in Section 2. To do so, we use a resampling approach. That is, for each respondent we shuffle the order of the sequence of choice observations and the order of the alternatives (left, middle, right) and apply the trained network. The respondent is then classified to the decision rule that attains the highest probability. But, rather than classifying each respondent just once, we classify each respondent 1000 times based on 1000 reshuffles of the order of the sequence of choice observations. After that, each respondent is classified to the decision rule that attains the highest number of 'votes' across the 1000 trials. We find that some respondents are very consistently classified by the network to one particular decision rule – regardless of how the data are shuffled. For other respondents the particular manifestation of the data does affect the decision rule classification. However, by classifying each respondent 1000 times, we obtain stable results in terms of what is the most likely decision rule employed for each respondent in our data.

Table 3 shows the final classifications. Based on Table 3 the following observations can be made. Firstly, and we consider this a noticeable substantive finding, we see that only 22 out of the 106 respondents are classified as random utility maximisers. This finding may spark further debate on the dominance of RUM models in discrete choice modelling practices – although we certainly do not want to claim that these market shares are one-to-one transferable to other choice contexts. Secondly, we see that the largest number of respondents (51) is classified as random regret minimisers. Thirdly, respectively 27 and 6 respondents are classified as Lexicographic and Random decision-makers.

[7] Except for lexicographic choice behaviour.

The design of the SC experiment itself may partly explain the obtained markets shares, for a number of reasons. Firstly, the relatively high market share of the RRM decision rule may (in part) be attributed to the fact that the compromise alternative is very easy to identify for respondents. Therefore, it seems possible that the design of the SC experiment may actually have *triggered* an RRM like decision rule on the side of the respondent. Secondly, the relatively high market share of the Lexicographic decision rule may be due to the fact that the experimental design consists of just two attributes (possibly in combination with the ranges of the attribute levels). This makes it fairly easy for respondents to use a lexicographic decision rule; a respondent may first decide on the most important attribute (either travel cost or travel time) and then choose consistently the best alternative based on that attribute.

The proposed method is rather flexible in application. It can be applied to SC as well as to Revealed Preference (RP) data. Applying the method to SC data seems however most natural since SC data are generally less noisy, which is an important asset since differences between decision rules can be subtle. We have applied the method in the context of a straightforward three alternative, two attribute SC data set. However, the method can essentially be applied to data sets which involve any number of attributes or alternatives. Finally, it is worthwhile to note that since the actual training is conducted on synthetic data, the size of the empirical data set is never a limiting factor to apply the method (this in contrast to discrete choice based methods). This makes the method also particularly suited to investigate decision rule heterogeneity in small data sets.

4. Cross-validation using discrete choice models

This section aims to cross-validate the classification results of Section 3.5. Although the performance of the ANN on the test data set is encouraging, it is good to keep in mind that we used synthetic data to train our ANN (Section 3.3) while in Section 3.5 we ultimately apply the ANN to real empirical data (Section 3.5). Due to this discrepancy between training and application additional analyses are needed to further examine the capability of the ANN to classify travellers to decision rules. The next two subsections present these analyses.

4.1. Model fit based on subsets

One way to cross-validate the capability of the ANN to classify respondents to decision rules is by estimating discrete choice models based on carefully created subsamples of the data. In particular, we split the 106 respondents in the data set into four subsets based on the highest classification probability as predicted by the ANN. That is, subset 1 consists of the 22 respondents classified by the ANN as random utility maximisers; subset 2 consists of the 51 respondents classified by the ANN as random regret minimisers, and so on. For these subsets we formulate the following conjectures:

If the ANN has accurately classified respondents to decision rules, then

1. The RUM model outperforms the P-RRM model in subset 1
2. The P-RRM model outperforms the RUM model in subset 2
3. Both the RUM and the P-RRM model have a very poor model fit in subset 4.
4. The respondents allocated to subset 3 include those 25 respondents which in Section 2.2 are identified to have consistently chosen for the fast and expensive or the slow and cheap alternative across all choice tasks.

On these subsets we estimate three types of discrete choice models: a linear-additive RUM model, a P-RRM model and a μRRM model (Van Cranenburgh et al., 2015a).[8] The latter model is a very flexible type of RRM model, which holds both the linear-additive RUM and the P-RRM model as special cases. The μRRM model consists of one additional parameter (as compared to linear-additive RUM and P-RRM). This 'regret aversion parameter' μ captures the degree of regret minimisation behaviour, where $\mu \rightarrow 0$ implies a very strong degree of regret aversion (i.e. a P–RRM decision rule) and $\mu \rightarrow \infty$ implies no regret aversion (i.e. a linear-additive RUM decision rule). As such, this model allows to empirically establish the underlying decision rule (RUM or RRM with varying degrees of regret aversion). All models are estimated in Multinomial Logit (MNL) form.

Table 4 shows estimation results.[9] First we look at the results for subset 1. We see that the RUM model very strongly outperforms the P-RRM model on this subset. The difference in final log-likelihood when put to the Ben-Akiva and Swait (1986) test for nonnested models, is very significant at $p < 0.000$. In fact, the very low ρ^2 for the P-RRM model indicates this model is hardly able to describe the data generating process in any meaningful manner. Despite this, we see that the signs are recovered. Looking at the results for the μRRM model, we see that the regret aversion parameter μ hits the upper bound, which is set at $\mu = 100$ in the estimation. Hence, the μRRM model collapses to the linear-additive RUM model, implying that imposing any degree of regret minimisation behaviour would worsen the model fit. Therefore, conjecture (1) is supported by these discrete choice analyses. As such, we are rather confident that the ANN has been successful in identifying those travellers whose choice behaviour is best described by RUM.

Looking at the results for subset 2 we see that the P-RRM model very strongly outperforms the RUM model. The difference in final log-likelihood – which is over 120 LL points – is highly significant. Furthermore, the signs of the parameters are all in the expected

[8] A choice set size correction factor is applied to the RRM models to be fully consistent with Section 3. This has no impact on model fit; it merely scales the estimates by a fix factor.

[9] Estimation results based on the full data set can be found in Appendix C.

Table 4
Estimation results.

Model	RUM MNL			P-RRM MNL			μRRM MNL		
Subset 1: Respondents identified by the ANN as random utility maximisers									
Number of observations	220			220			220		
Number of individuals	22			22			22		
Null Log-likelihood	−241.7			−241.7			−241.7		
Final Log-likelihood	−212.2			−238.0			−212.2		
ρ^2	0.12			0.02			0.12		
Parameters	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*
β_{cost}	−1.26	0.176	−7.13	−0.29	0.115	−2.57	−1.24	0.174	−7.11
β_{time}	−0.30	0.046	−6.61	−0.05	0.027	−1.82	−0.30	0.045	−6.58
μ							100		
Subset 2: Respondents identified by the ANN as random regret minimisers									
Number of observations	510			510			510		
Number of individuals	51			51			51		
Null Log-likelihood	−560.3			−560.3			−560.3		
Final Log-likelihood	−510.0			−383.2			−383.2		
ρ^2	0.09			0.32			0.32		
Parameters	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*
β_{cost}	−0.88	0.110	−8.01	−1.63	0.130	−12.57	−1.63	0.130	−12.57
β_{time}	−0.28	0.030	−9.19	−0.50	0.037	−13.29	−0.50	0.037	−13.29
μ							0.01		
Subset 4: Respondents identifified by the ANN to make choices randomly									
Number of observations	60			60			60		
Number of individuals	6			6			6		
Null Log-likelihood	−65.9			−65.9			−65.9		
Final Log-likelihood	−61.1			−61.2			−61.1		
ρ^2	0.07			0.07			0.07		
Parameters	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*
β_{cost}	0.37	0.309	1.19	0.27	0.210	1.28	0.37	0.309	1.19
β_{time}	0.00	0.080	−0.01	−0.02	0.058	−0.40	0.00	0.080	−0.01
μ							100		

directions. The μRRM model indicates that the choice behaviour of the respondents in this subset is best described by very strong regret minimisation behaviour. The regret aversion parameter μ in the μRRM model hits the lower bound, which is set at $\mu = 0.01$ in the estimation, implying that the μRRM model collapses to a P-RRM model. Therefore, also conjecture (2) is supported by these discrete choice analyses. This gives us confidence that the ANN has also successfully identified those travellers whose choice behaviour is best described by RRM.

Furthermore, it is encouraging to see that the ratios of the parameter estimates are roughly constant across the subsets. In a RUM modelling framework, the ratios of parameters represent the marginal rate of substitution, which – in this context – translate into VoTs. The implied VoTs by the RUM models in subset 1 and 2 are respectively $\beta_{Time}/\beta_{Cost} = 0.24$ €/minute and $\beta_{Time}/\beta_{Cost} = 0.30$ €/minute. These results provide some support for that the ANN has classified travellers based on decision rules, rather than based on taste heterogeneity – as is a major methodological problem when using a LC discrete choice modelling approach to investigate decision rule heterogeneity (Hess et al., 2012).

Looking at the results for subset 4 we see that the RUM, P-RRM and the μRRM model all fit the data very poorly ($\rho^2 \leq 0.08$). This means that none of these models is able to describe the data generating process meaningfully. The small (and insignificant) taste parameters in all three models confirm this inference. These tell us that these models basically predict rather random choices. Therefore, conjecture (3) is supported.

Finally, we analyse subset 3. The ANN classified 27 respondents to the Lexicographic decision rule. To examine this classification we do not need to estimate discrete choice models. After all, lexicographic behaviour can fairly easily be detected by inspection of the data.[10] As expected, we find that all 25 respondents who consistently chose for the 'fast and expensive' alternative (9) and the 'slow and cheap' alternative (16) are classified as lexicographic by the ANN. In addition, two respondents which chose twice the 'fast and expensive' route and 8 times the 'slow and cheap' route are 'falsely' allocated to this class. Nonetheless, despite these misclassifications, it is very reasonable to say that also the final conjecture (4) is supported. Taken together, these results encourage us to believe that the ANN is well capable to classify travellers in terms of their underlying decision rules.

[10] Although we acknowledge that extreme preferences may also lead to seemingly lexicographic choice behaviour.

Table 5
LC estimation results.

Subset 1, 2 and 4																		
Model	3-class discrete mixture LC									3-class mixed-mixed logit								
Number of observations	790									790								
Number of individuals	79									79								
Number of draws	N/A									500								
Null Log-likelihood	−867.9									−867.9								
Final Log-likelihood	−675.8									−617.8								
ρ^2	0.221									0.288								
BIC	1391.6									1289.0								
Decision-rule	*Class 1 RUM [19%]*			*Class 2 P-RRM [61%]*			*Class 3 RND [20%]*			*Class 1 RUM [24%]*			*Class 2 P-RRM [72%]*			*Class 3 RND [4%]*		
Model parameters	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*	*Est*	*Std error*	*t-val*
β_{cost}	−1.35	0.488	−2.77	−1.87	0.176	−10.65	0	–fixed		−2.84	0.531	−5.35	−1.73	0.210	−8.25	0	–fixed	
β_{time}	−0.74	0.191	−3.87	−0.49	0.047	−10.24	0	–fixed		−0.55	0.095	−5.79	−0.64	0.061	−10.49	0	–fixed	
σ_{cost}										1.55	0.185	8.36						
σ_{time}										0.21	0.106	1.99						
Class allocation parameters																		
	Est	Std error	t-val							Est	Std error	t-val						
s1	0.00	–fixed								0.00	–fixed							
s2	1.17	0.321	3.63							1.12	0.34	3.31						
s3	0.06	0.409	0.15							−1.73	0.787	−2.19						

Table 6
Cross-table ANN and LC classification based on highest classification probability.

		Mixed-mixed logit classification			
		RUM	P-RRM	Random	∑
ANN classification	RUM	15	7	0	22
	P-RRM	3	43	5	51
	Random	0	2	4	6
	∑	18	52	9	79

4.2. Latent class modelling approach

Another way to learn about the capability of ANNs to classify travellers in terms of their underlying decision rule involves comparing the classification of the ANN with those obtained from a traditional Latent Class discrete choice modelling approach. To do so, we estimate – in consonance with the ANN – LC discrete choice models with three predefined classes: a RUM, a P-RRM and a Random class.[11] Specifically, we estimate a discrete mixture LC model as well as a so-called mixed-mixed logit model (Keane and Wasi, 2013). This latter model is a discrete mixture of mixed logit models. This model allows for taste heterogeneity within decision rule classes. For the random parameters different distributions are tested (normal, triangular, uniform). Uniform distributions are found to give the best performance in terms of model fit. Furthermore, we have estimated 'generic' sigmas, in the sense that they are shared across decision rule classes, as we find this specification to outperform a specification in which class-specific sigmas are estimated (when taking the model parsimony into account).

Latent class models are estimated using Pythonbiogeme (Bierlaire, 2016). To avoid getting stuck in local maxima, estimations are repeated 20 times using randomly drawn starting values between -1 and 1. For these LC analyses we use subsets 1, 2 and 4. Together these subsets comprise of 79 respondents. Subset 3 – consisting of respondents classified as Lexicographic and also identified as such in Section 2.2 – is excluded from this analysis because discrete choice models are not well-equipped to deal with lexicographic choice behaviour (Hess et al., 2010).

Table 5 shows the LC estimation results. First we look at the results for the discrete mixture LC model. Looking at the market shares of the decision rules, we see that these are rather similar to those predicted by the ANN[12] (RUM: 27%, P-RRM: 65%, RND: 8%). Both models find that P-RRM is the most common decision rule in this sample, with between 60% and 70% market share. However, whereas the ANN classified the remaining decision-makers mostly as RUM (27%), in the LC model the Random decision rule attains the second highest market share. Next, we look at the ratios of the parameter estimates, and compare these to those presented in Table 4. The ratios of the time and cost parameter in Table 5 are respectively $\beta_{time}^{RUM}/\beta_{cost}^{RUM} = 0.55$ and $\beta_{time}^{P-RRM}/\beta_{cost}^{P-RRM} = 0.26$, for the RUM and P-RRM class. In Table 4 we however find a ratio of $\beta_{time}^{RUM}/\beta_{cost}^{RUM} = 0.24$ for decision-makers classified as RUM, and a ratio of $\beta_{time}^{P-RRM}/\beta_{cost}^{P-RRM} = 0.30$ for decision-makers classified as P-RRM. Hence, particularly the ratio of the RUM parameters seems 'outside' from what we would expect based on Table 4. This signals that the LC discrete mixture model has also captured taste heterogeneity – aside from decision rule heterogeneity. This exposes the methodological shortcoming of LC models to study decision rule heterogeneity.

Looking at the results for the mixed-mixed logit model we see that accounting for taste heterogeneity within the decision rule classes substantially improves model fit. With regard to the predicted market shares we see a moderate change as compared to the discrete mixture LC model market shares. In particular, the mixed-mixed logit model predicts larger shares for RUM and P-RRM at the expense of the Random decision rule class. Noteworthy, the market shares predicted by the mixed-mixed logit model and those predicted by the ANN are even closer to one another, than are the ones predicted by the discrete mixture LC model and those predicted by the ANN.

Finally, we investigate the consensus between the ANN and mixed-mixed logit model in terms of their classifications of individual respondents. A strong consensus between the two modelling approaches would provide confidence in both methods for their capability to investigate decision rule heterogeneity. To compute the classification probability of individual respondents for the mixed-mixed model, first we simulate the choice probabilities to obtain their expected values. Then, we apply Bayes rule. That is, we compute the likelihood of the model – which in this context is the decision rule – given the observed sequence of choices of each respondent. Respondents are allocated to the decision rule (i.e. the class) with the highest classification probability, just like is done for the ANN classification in Section 3.4.

Table 6 shows the results in a cross-table. The rows contain the ANN classification; the columns contain the LC model classification. The cells on the diagonal show the number of cases in which the same respondents are classified to the same decision rules by both modelling approaches. The off-diagonal cells indicate disagreements between the two modelling approaches in terms of what is the most likely underlying decision rule of a respondent.

[11] For the sake of completeness we also tested LC models with more and less than 3 classes. 3-class models are found to obtain the lowest BIC values.

[12] Market shares excluding the respondents classified by the ANN as lexicographic.

A number of inferences can be made based on Table 6. Firstly, in the majority of the cases the ANN and the LC model agree on the most likely underlying decision rule: almost 80 per cent of the respondents are allocated to the same decision rule by the ANN and the mixed-mixed logit model. We see a strong agreement between the ANN and mixed-mixed logit classification particularly for the RRM decision rule. Strongest disagreement between the two methods is seen for the Random decision rule. These results show that even though the two methods find very similar market shares for the decision rules, their underlying results can be considerably different. This highlights the added value of having a second method available to investigate decision rule heterogeneity.

All together, we believe that the notion that ANNs can be used to investigate decision rule heterogeneity has convincingly been demonstrated. We have cross-validated the ANN outcomes using two approaches: (1) by estimating discrete choice models based on carefully created subsets of the data and (2) by comparing results with those obtained from LC discrete choice models (discrete mixture and mixed-mixed logit). Both approaches provide strong support for the capability of ANNs to identify underlying decision rules. As such, we believe that ANNs provide a promising addition the toolbox of analysts who wish to investigate decision rule heterogeneity. A potential benefit of ANNs over conventional Latent Class models is that ANNs can be trained to recognise decision rules in the presence of taste heterogeneity. Therefore, and given that ANNs are a sort of LC models on 'steroids', they may be better capable to disentangle decision rule heterogeneity from taste heterogeneity than LC models. However, further research is needed to investigate this conjecture more in depth. A disadvantage of our ANN based approach– as compared to traditional LC modelling – is however that in the latter model the membership function can be used to directly provide insights on what type of traveller is best described by what type of decision rule e.g. in terms of socio-demographic variables. With ANNs this can only be done by means of correlation analysis ex post (due to the fact that the ANN is trained on synthetic data). Furthermore, LC discrete choice models do provide confidence intervals, while ANNs do not.

5. Conclusions and discussion

This study is the first to investigate decision rule heterogeneity amongst traveller using a novel artificial neural networks based approach. We have shown how ANNs can be employed to investigate decision rule heterogeneity amongst travellers. In particular, we have proposed a novel ANN topology which is equipped to deal with the panel structure of SC data and we have shown that the ANN can be trained using synthetic data. Based on the encouraging results we have obtained, we believe that ANNs provide a valuable addition to the toolbox of analysts who wish to investigate decision rule heterogeneity. The substantive contribution is that we enriched the growing body of empirical studies providing evidence for the presence of decision rule heterogeneity amongst travellers.

Finally, we would like to point out several limitations to this study, providing avenues for further research. Firstly, to keep track of our results we have trained our ANN to learn a fairly small number of decision rules (4). In future research ANNs can be trained to learn recognise more types of decision rules, such as Contextual concavity (Kivetz et al., 2004), Relative Advantage Maximization (RAM) (Leong and Hensher, 2014), Reference dependent utility maximization (Koszegi and Rabin, 2006), Stochastic Satisficing (Gonzalez-Valdes and Raveau, 2018) and models that capture learning effects, such as e.g. Value Learning (McNair et al., 2012; Balbontin et al., 2017). Furthermore, given that our empirical analyses are based on just one data set, it is advisable to repeat these analyses using other data sets. This will provide a richer view on the extent to which ANNs are a valuable tool to investigate decision-rule heterogeneity. Furthermore, new types of ANNs can be developed, possibly inspired by the rapid developments in computer science. Lastly, a well-known limitation of ANNs relates to its black-box nature. Given their complex internal structure, ANNs provide limited insights on underlying causal relations (e.g. by what mechanism does it actually detect the decision-rules?). Future research may be directed to illuminate the black boxes of ANNs (Castelvecchi, 2016). This may help researcher to better understand human decision-making, and perhaps even may lead to the discovery of new decision-rules.

Acknowledgements

Support from the King Abdulaziz City for Science and Technology (KACST) is gratefully acknowledged by the second author.

Appendix A. Choice tasks in the value-of-time choice experiment

Choice task	Route A		Route B		Route C	
	TT	TC	TT	TC	TT	TC
1	23	6	27	4	35	3
2	27	5	35	4	23	6
3	35	3	23	5	31	4
4	27	4	23	5	35	3
5	35	3	27	6	31	5
6	23	6	27	5	35	3
7	35	3	31	5	23	6
8	27	5	23	6	31	3
9	35	3	31	4	27	6
10	23	6	27	4	35	3

Appendix B. Sample statistics

Variable		Sample frequency	Percentage [%] in sample
Gender			
	Male	44	42%
	Female	45	42%
	Missing	17	16%
Age			
	18–24 yr.	2	2%
	25–34 yr.	24	23%
	35–44 yr.	25	24%
	45–54 yr.	21	20%
	55–64 yr.	15	14%
	65–74 yr.	2	2%
	Missing	17	16%
Completed education			
	No education	0	0%
	Elementary school	7	7%
	Lower education	5	5%
	Middle education	39	37%
	Higher education	34	32%
	University education	4	4%
	Missing	17	16%
Income			
	I < €9,400	2	2%
	€9,400 ≤ I < €14,700	4	4%
	€14,700 ≤ I < €20,600	5	5%
	€20,600 ≤ I < €33,500	14	13%
	€33,500 ≤ I < €67,000	37	35%
	I ≥ €67,000	27	25%
	Missing	17	16%

Appendix C. Estimation results based on full data set

Full data set									
Model	RUM MNL			P-RRM MNL			μRRM MNL		
Number of observations	1060			1060			1060		
Number of individuals	106			106			39		
Null Log-likelihood	−1164.5			−1164.5			−1164.5		
Final Log-likelihood	−1123.0			−1128.4			−1118.4		
ρ^2	0.036			0.031			0.040		
Parameters	Est	Std error	t-val	Est	Std error	t-val	Est	Std error	t-val
β_{cost}	−0.64	0.072	−8.85	−0.43	0.053	−8.07	−0.64	0.073	−8.82
β_{time}	−0.16	0.019	−8.58	−0.10	0.013	−7.69	−0.16	0.019	−8.35
μ							1.19	0.511	2.32

Appendix D. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trc.2018.11.014.

References

Alwosheel, A., Van Cranenburgh, S., Chorus, C.G., 2017. Artificial neural networks as a means to accommodate decision rules in choice models. ICMC2017, Cape Town.

Alwosheel, A., van Cranenburgh, S., Chorus, C.G., 2018. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. J. Choice Modell. 28, 167–182.

Balbontin, C., Hensher, D.A., Collins, A.T., 2017. Integrating attribute non-attendance and value learning with risk attitudes and perceptual conditioning. Transportation Res. Part E: Logistics Transportation Rev. 97, 172–191.

Ben-Akiva, M., Swait, J., 1986. The akaike likelihood ratio index. Transportation Sci. 20 (2), 133–136.

Bierlaire, M., 2016. PythonBiogeme: a short introduction. Technical Report TRANSP-OR 160706.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press).

Boeri, M., Longo, A., 2017. The importance of regret minimization in the choice for renewable energy programmes: Evidence from a discrete choice experiment. Energy Econ. 63, 253–260.

Borysov, S., Lourenço, M., Rodrigues, F., Balatsky, A., Pereira, F., 2016. Using internet search queries to predict human mobility in social events. In: 2016 IEEE 19th

international conference on intelligent transportation systems (ITSC).
Castelvecchi, D., 2016. Can we open the black box of AI? Nat. News 538 (7623), 20.
Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation Res. Part C: Emerging Technol. 68, 285–299.
Chorus, C.G., 2014. Capturing alternative decision rules in travel choice models: a critical discussion. In: Hess, S., Daly, A. (Eds.), Handbook of Choice Modelling. Edward Elgar, pp. 290–310.
Chorus, C.G., Bierlaire, M., 2013. An empirical comparison of travel choice models that capture preferences for compromise alternatives. Transportation 40 (3), 549–562.
Gonzalez-Valdes, F., Raveau, S., 2018. Identifying the presence of heterogeneous discrete choice heuristics at an individual level. J. Choice Modell. 28, 28–40.
Guevara, C.A., Fukushi, M., 2016. Modeling the decoy effect with context-RUM Models: Diagrammatic analysis and empirical evidence from route choice SP and mode choice RP case studies. Transportation Res. Part B: Methodol. 93 (Part A), 318–337.
Haykin, S.S., 2009. Neural networks and learning machines: Pearson Upper Saddle River, NJ, USA:).
Hensher, D.A., Ton, T.T., 2000. A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice. Transportation Res. Part E: Logistics Transportation Rev. 36 (3), 155–172.
Hess, S., Chorus, C.G., 2015. Utility maximisation and regret minimisation: a mixture of a generalisation. RASOULI, S. & TIMMERMANS H., Bounded Rational Choice Behaviour: Applications in Transport. Bingley UK: Emerald, pp. 31–48.
Hess, S., Rose, J.M., Polak, J., 2010. Non-trading, lexicographic and inconsistent behaviour in stated choice data. Transportation Res. Part D: Transport Environ. 15 (7), 405–417.
Hess, S., Stathopoulos, A., Daly, A., 2012. Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. Transportation 39 (3), 565–591.
Keane, M., Wasi, N., 2013. Comparing alternative models of heterogeneity in consumer choice behavior. J. Appl. Econometrics 28 (6), 1018–1045.
Kivetz, R., Netzer, O., Srinivasan, V., 2004. Alternative models for capturing the compromise effect. J. Mark. Res. 41 (3), 237–257.
Koszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. Quart. J. Econ. 121 (4), 1133–1165.
LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.-R., 2012. Efficient BackProp. In: Montavon, G., Orr, G.B., Müller, K.-.R. (Eds.), Neural Networks: Tricks of the Trade, second ed. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 9–48.
Leong, W., Hensher, D.A., 2012. Embedding decision heuristics in discrete choice models: a review. Transport Rev. 32 (3), 313–331.
Leong, W., Hensher, D.A., 2014. Relative advantage maximisation as a model of context dependence for binary choice data. J. Choice Modell. 11, 30–42.
Maren, A.J., Harston, C.T., Pap, R.M., 2014. Handbook of Neural Computing Applications. Academic Press.
McNair, B.J., Hensher, D.A., Bennett, J., 2012. Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: a probabilistic decision process model. Environ. Resour. Econ. 51 (4), 599–616.
Mohammadian, A., Miller, E., 2002. Nested logit models and artificial neural networks for predicting household automobile choices: comparison of performance. Transportation Res. Rec.: J. Transportation Res. Board 1807, 92–100.
Omrani, H., Charif, O., Gerber, P., Awasthi, A., Trigano, P., 2013. Prediction of individual travel mode with evidential neural network model. Transportation Res. Record: J. Transportation Res. Board 2399, 1–8.
Payne, J.W., Bettman, J.R., Johnson, E.J., 1993. The Adaptive Decision Maker. Cambridge University Press.
Rouwendal, J., de Blaeij, A., Rietveld, P., Verhoef, E., 2010. The information content of a stated choice experiment: A new method and its application to the value of a statistical life. Transportation Res. Part B: Methodological 44 (1), 136–151.
[Data set] Van Cranenburgh, S., Chorus, C.G., 2018. Value-of-time experiment, Netherlands. In 4TU.Centre for Research Data. https://doi.org/10.4121/uuid:1ccca375-68ca-4cb6-8fc0-926712f50404.
Van Cranenburgh, S., Guevara, C.A., Chorus, C.G., 2015. New insights on random regret minimization models. Transportation Res. Part A: Policy Practice 74, 91–109.
Van Cranenburgh, S., Prato, C.G., Chorus, C., 2015b. Accounting for variation in choice set size in Random Regret Minimization models. uuid:1de1446b-1feb-4213-bef4-12031fa61579.
Van Cranenburgh, S., Rose, J.M., Chorus, C.G., 2018. On the robustness of efficient experimental designs towards the underlying decision rule. Transportation Res. Part A: Policy Practice 109, 50–64.
van Lint, J.W.C., Hoogendoorn, S.P., van Zuylen, H.J., 2005. Accurate freeway travel time prediction with state-space neural networks under missing data. Transportation Res. Part C: Emerging Technol. 13 (5–6), 347–369.
Vlahogianni, E.I., Park, B.B., van Lint, J.W.C., 2015. Big data in transportation and traffic engineering. Transportation Res. Part C: Emerging Technol. 58 (Part B), 161.
Wong, M., Farooq, B., Bilodeau, G.A., 2017. Latent behaviour modelling using discriminative restricted Boltzmann machines. ICMC 2017, Cape Town.