

How do cyclists make their way? - A GPS-based revealed preference study in Copenhagen

Hans Skov-Petersen^a, Bernhard Barkow^b, Thomas Lundhede^{c,d,e}, Hans Skov-Petersen^a, Bernhard Barkow^b, Thomas Lundhede^{c,d,e}

aDepartment of Geoscience and Nature Management, University of Copenhagen, Frederiksberg C, Denmark; bCreative Eyes, Vienna, Austria; cDepartment of Food and Resource Economics, University of Copenhagen, Frederiksberg C, Denmark; dCentre for Macroecology, Evolution and Climate, University of Copenhagen, Frederiksberg C, Denmark; eCentre for Environmental Economics and Policy in Africa, Department of Agricultural Economics, Extension and Rural Development, University of Pretoria, South Africa

Abstract:

It is the objective of the study to determine the extent to which human navigation is affected by perceptions of our immediate surroundings or by already established knowledge in terms of a cognitive map. The motivation is to contribute to the knowledge about human navigation and to inform planning with estimates of bicyclists' route preferences and 'willingness-to-pay' (in terms of transport distance vs utility/disutility of route characteristics). The core method is choice modelling of observed route data. 1,267 trips performed by 183 cyclists in Copenhagen (Denmark) were recorded by GPS. The trips were map-matched to a digital road and path network, which enabled the generation of choice sets: one for navigation as influenced by perception of immediate surroundings, comprising edges connected to network-nodes (hereafter called the edge dataset), and one for navigation based on a priori knowledge, comprising the trip itself and a number of alternative routes generated by a labelling algorithm (hereafter called the route dataset). The results document that choices based on characteristics of both the route and the edge data can be estimated and provide reasonable and significant parameter estimates regarding cyclists' preferences. Length was significant and negative, illustrating that cyclists – everything else kept equal – prefer to bike shorter distances. Preferences regarding characteristics of bike path, presence of traffic lights and road types show similar results to the two types of data; most importantly that routes with facilities, such as curbed tracks and segregated bikeways, were significantly preferred. The study concludes that cyclists' wayfinding can be modelled as choices based on both an edge dataset and a route dataset and, thus, may be influenced by both perceived information and a priori knowledge. We suggest that future analyses of movement and route preferences take both modes into account as actual movement may be based on a combination of the two and because assessment of the influence of the immediate, perceivable surroundings can provide information not to be considered in a wayfinding approach. In our case differences in preferences along the route is for example found, but this can be expanded in future studies to also include dynamic aspects such as weather and crowding.

Key words: Movement, navigation, choice modelling, cognition, bicycles

1. Introduction

Movement, the pattern of relocation of individuals, can be conceived at different temporal and/or spatial scales. This includes e.g. movement between cities due to a change in life phase, migrating between the North and South according to season, the daily commute from home to work or school, or simply the movement pattern when mowing the lawn. While some types of movement are directed towards a pre-determined destination, others rather aim at 'consuming' the environment along the taken route. Such 'undirected wayfinding' (Weiner et al, 2009) may be in the form of exploring, shopping, hunting or enjoying nature while hiking. Montello (2005) has suggested the term 'navigation' for goal-directed and coordinated movement which can be divided into two components: 'locomotion' and 'wayfinding'. Locomotion is movement coordinated by the local environment, as it is directly perceptible to our sensory and motor systems. Wayfinding is planned and efficient movements based on

spatial knowledge in form of a cognitive map (Montello and Sas, 2006). In actual navigation, neither of the two is in action in isolation. Rather, a combination is constantly applied even though behaviour based on only one of the two can be observed in extreme cases (Montello and Sas, 2006).

It can be hypothesised that navigation in unfamiliar environments primarily will be influenced by locomotion based on perceptible surroundings and legible landmarks (Lynch, 1960), and that navigation in well-known areas will take advantage of existing knowledge which, accordingly, can be characterised as wayfinding.

Navigation as a sub-category of spatial behaviour involves choices between alternatives or, as Rushton (1976, p119) states, that the spatial choices process must be seen '*... as the subjective selection of the most preferred alternative from a subset of alternatives*'. Accordingly studies of navigation must be based not only on what the subject actually does, but also on the alternative options that were apparently considered, but not chosen. In relation to navigation constrained by a transport infrastructure, modelled as a topological network, we suggest that the set of options/choices available during wayfinding is the potential routes leading from start to end of a trip (including the recorded route). Wayfinding therefore consist of only one choice in the very beginning of the trip. Locomotion, on the other hand, is based on immediately perceived information during the trip, and therefore consist of a number of choices along the route. For each node of a recorded route, the set of choices is constituted by the edges connected to the node. Generation of route and edge choice sets can be further constrained, which will be discussed in detail below.

In the present study, our primary aim is to assess the extent to which human navigation related to bicycle movement is affected by perception involved in locomotion in addition to wayfinding. Secondly, we investigate whether the attributes involved influence navigators' utility in a similar or comparable way for the two components of navigation.

2. Knowledge of cyclist behaviour and preferences

Cycling is receiving increasing political attention as an alternative to automobile transport (Parkin and Horton, 2012; Pucher and Buehler, 2012). The reasons for this include a) a desire to reduce congestion in the transport network and parking facilities, b) an environment protection aim from e.g. particle contamination and CO₂ emissions, and c) a wish to increase citizens' physical activity and, thereby, improve their health (see, for instance, Krizek et al., 2009; Ogilvie et al., 2011). Policy initiatives that aim to improve urban design/infrastructure to encourage bicycling have been implemented in many cities around the world (see e.g., Dextre et al., 2013). Improvements to bicycle infrastructure are regarded as the main instrument to achieving such a goal (Parkin and Koorey, 2012). However, studies of cyclists' actual reactions to such improvements, for instance, as willingness to 'pay' in terms of increased distance are still in the developmental phase. This paper addresses this question by studying cyclists' navigational behaviour in Copenhagen, Denmark. From an international perspective Copenhagen is special due to its long-standing tradition of cycling and pro-cycling planning (Gössling, 2013); a total of 35% of all trips made by inhabitants in the Municipality of Copenhagen are made by bicycle (Jensen 2013; Buehler and Pucher 2012).

Several GPS-based studies of cyclists' route preferences and navigation have been conducted and for a recent review we refer to Buehler and Dill (2016). In applications, cyclists' recorded routes are often compared to shortest paths (see e.g. Harvey et al., 2008 and Winters et al., 2010). Yeboah and Alvanides (2014) are extending this approach by including certain constraints. Some studies consider general characteristics of the environment contained in corridors surrounding recorded tracks without explicitly considering routes afforded by the topology of the street network (Yeboah, et al., 2015; Madsen et al., 2014).

Only a limited number of studies have attempted to apply choice models, based on GPS-recordings, to reveal preferences involved in biking. Menghini et al. (2010) presented a route choice model for bicyclists, based on GPS observations. A total of 3,387 cycling trips were revealed from transport trajectories performed by 2,435 citizens of Zürich, Switzerland. No personal characteristics were made available. Alternative routes were generated by searches for the shortest distance between origin and destination and the removal of edges in turn until the required set of unique routes was identified. Only a few explaining environmental parameters were included. The results reveal that length dominates the cyclists' choices, but that the distribution of bicycle paths also has a considerable, albeit substantially lesser impact. In addition, the gradient of slopes has a strong, negative impact too.

Boarch et al. (2012) present a study of 164 utilitarian bicyclists' (i.e. commuters) route preferences in Portland, US, to assess the relative attractiveness of different types of bicycle, other characteristics of the infrastructure, and environmental attributes such as slope gradient. Alternative routes were generated by a constrained labelling algorithm guided by minimising/maximising given attributes of the potential routes. Among the findings of the study, slope had a negative impact as had signs and traffic signals. Effects of gender, age group, or parental status were also tested in Boarch et al. (2012), but no significant differences were found. Other studies have detected gender effects, including Boarch, et al (2016) where the effect on mode choice is mentioned and Vedel et.al. (2017) as an example of a stated preference study where a gender effect is detected.

Hood et al. (2011) collected GPS trajectories with a smart phone application. A total of 3,034 bicycle trips were recorded which had been made by 366 respondents. Alternatives were generated by repeated shortest path searches with progressive randomisation of network impedances. The results reveal that respondents preferred bicycle lanes to other types of bicycle facilities. This is particularly the case for infrequent cyclists. Further, the study showed that cyclists expressed a negative utility for length, as well as for turns.

The study presented by Halldórsdóttir et al., (2015) was conducted in Copenhagen, Denmark like the present study. In total 2,681,108 GPS-points comprising 5,027 bicycle trips were identified. Respondents' age and gender profiles and trip and weather-related attributes were also included. Alternative routes were generated by a doubly stochastic generation function. The study found that cyclists were sensitive to length, cycling the wrong way, turn frequency, and terrain gradient.

In all four studies, a path size parameter was used to cater for collinearity of (partly) overlapping alternatives (see below for details). And more important, they are all entirely based on the assumption of wayfinding, i.e. that navigation is complete *a priori* knowledge

The present study adds to the existing literature on bicycle behaviour and preferences, firstly, by analysing choices based on both route and edge data and, secondly, by investigating preferences for the quality of the bikeway infrastructure and topology, and the character of the environment in terms of shopping streets and green surroundings.

3. Data

3.1 GPS recordings

Respondents were recruited as part of a questionnaire survey conducted in Copenhagen in April/May, 2011 (see Vedel et al., 2017). Respondents to the questionnaire survey were contacted via flyers which were distributed at 16 locations in the Copenhagen area and recruited during working hours on weekdays. The sample is considered to represent commuting cyclists, rather than recreationalists. A total of 1,262 volunteered to participate in the GPS survey, but since only 70 GPS devices were available for the project, 210 respondents were randomly selected, stratified over the 16 locations previously applied to ensure an even distribution over all the case areas. Each respondent had to carry a GPS unit for a week, from Monday morning to Sunday evening. A total of 27 cyclists failed to complete the GPS tracking, which meant that recordings of 183 cyclists performing 1,267 trips were included in the study. The average trip duration was 20 minutes, over a distance of 4.46 km, at a speed of 13.31 km/h. The distribution of the recorded point over the road network of Copenhagen is shown in supplementary online material, figure 1.

Profiles of gender, age, income and educational level of the participating cyclists were collected by the questionnaire survey (Vedel et al. 2017). See in the supplementary materials, table A3, for summary statistics.

3.2 Road network and traffic lights

The road network and its transport-related attributes were extracted from the Open Street Map (2017)¹. A total of 6,186 km of infrastructure, comprising 64,866 edges was applied. Also, from OSM 794, traffic lights were extracted and applied to the study. Edges were classified regarding road types and cycle infrastructure according to the categorisation used in Open Street Map. The distribution of the network over classes is found in supplementary online materials A1 and A2.

The edges of the network were further attributed with their proximity (closer than 50 m) to green environments based on digital maps of forests, parks and protected areas. More than 60 % of the road network was situated further than 50 m from green spaces, whereas 19% of the network was completely within 50 m of green spaces.

For the edge dataset, we created a dummy variable indicating whether the node is in the first 25%, in the middle, or in the last section of the route.

¹ The quality of OSM depends on location but is here expected to be sufficient for the purpose of the study.

Table 1. Variables included in the route and the edge datasets.

	Dataset	Variable name	Description
Effort variables	Route	Distance	The length of the route (km)
	Edge	Angle to destination	An edge's deviation from a direct line towards the destination (degree)
	Route	Right turns	Number of right turns along the route
	Edge	Right turn	Whether a given edge is a right turn from the entrance into a node
	Route	Left turns	Number of left turns along the route
	Edge	Left turn	Whether a given edge is a left turn from the entrance into a node
Bikeway quality	Route	Bike track	The proportion (0–1) of the route that has a bike track alongside a road
	Edge	Bike track	Whether a given edge is a bike track alongside a road
	Route	Bike lane	The proportion (0–1) of the route that has a bike lane alongside a road
	Edge	Bike lane	Whether a given edge has a bike lane
	Route	Segregated bikeway	The proportion (0–1) of the route that is a segregated bikeway
	Edge	Segregated bikeway	Whether a given edge is a segregated bikeway
	Route	Bikeway shared with pedestrians	The proportion (0–1) of a route that is a bikeway shared with pedestrians
Edge	Bikeway shared with pedestrians	Whether a given edge is a bikeway shared with pedestrians	
Road/environment	Route	Traffic road	The proportion (0–1) of a route (0–1) that is along a traffic road
	Edge	Traffic road	Whether a given edge is a traffic road
	Edge	Local road	Whether a given edge is a local road
	Edge	Green environment	The percentage of the edge that is near to nature areas

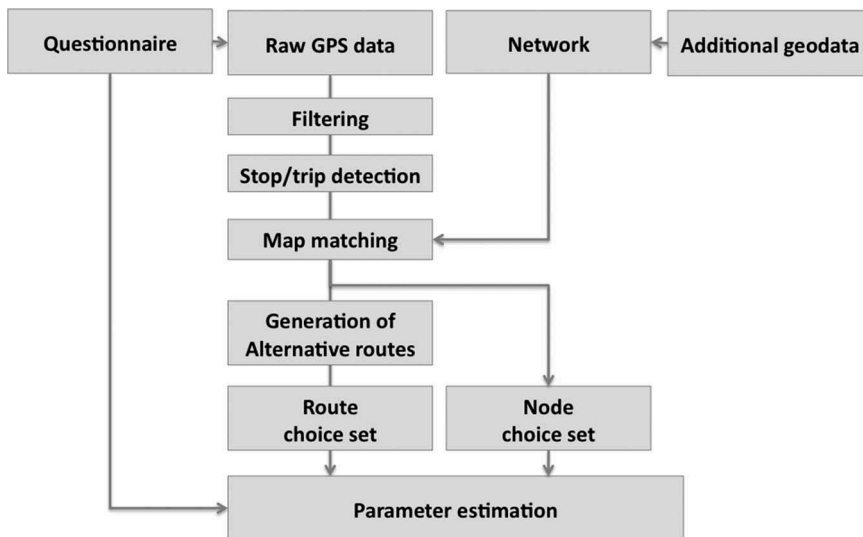


Figure 1. Flow chart of the data processing.

A list of all variables included in the route and the edges datasets is found in Table 1.

4. Method

The steps of the data processing, enrichment and analysis are presented in figure 1. The steps are described in detail in the following sections.

4.1 Filtering and pre-processing

Out of the 2,753,401 originally recorded raw points, 2,162,577 remained for further analysis after a filtering process (Harder et al., 2011).

4.2 Stop detection

A core process in calculating semantic trajectories is identification of points recorded while moving or standing still (Parent et al., 2013). For our study, this was performed prior to their inclusion in the present study, and – for each point - was based on averaging speed, distance, and angular deviation within a spatial and temporal ‘window’ around it (Jensen et al., 2009). Between stops a total of 1,267 trips were identified. On average, 91 points were recorded per trip. Respondents performed an average of 6.3 trips.

4.3 Map-matching

A map-matched route is the series of connected network edges that best represents a set of recorded GPS-points between closest to the first and last point of a trip. In the present case, all points of a trip are linked to their closest edge of the network. Therefore, each edge is marked with the number of points that are related to it. The route with the highest point density, i.e. number of points per unit of length, will be matched to the trip, (for further details, see the on-line material or visit the GitHub page of the projects (Snizek et al., 2011)).

4.4 Generation of choice sets

To generating feasible route alternatives, we applied an approach based on labelling, where the network is traversed from a node of origin, edge by edge, until the destination node is reached or one of a set of constraints has been met (Skov-Petersen et al., 2010). In networks representing real world infrastructure, the number of potential alternatives will often be exhaustive and will not be manageable in terms of computation time. Therefore, a set of heuristic constraints to the search must be applied. At the same time, the proposed alternative needs to be realistic, i.e. a navigating individual would in fact consider it.

The search was carried out in two stages. Stage 2 was only applied when no alternative routes had been found during stage 1. A range of combinations for the two stages was tested. Eventually the following was applied:

- **Sub-network:** In stage 1, a sub-network was used within which the search was performed, including only edges not more than 100 m from the matched route. In stage 2, this distance was set to 1000 m.

- Maximum length of the route: In stage 1, the search was stopped when the length was 50% longer than the matched route. In stage 2, the overhead length was set to 25%.
- Search strategy: In stage 1, a depth first search was applied with priority being given to the edge leading most directly towards the destination. When the destination had been reached, one step edge back was taken and the forward search was repeated (including edges that had not been tested yet). Out of all identified routes 20 were randomly selected. In stage 2, the forward search was performed with no priority given to direction; here, the search was reset to the origin every time a destination had been reached.

The code applied to perform the map matching process and generation of choice alternatives is available in the supplementary online information and can be obtained from the authors.

To address the problems related to the potential overlap of alternative routes in a choice set, a path size (PS) value (Frejinger et al. 2009) was calculated for each route alternative. The PS is 0 in the case where an alternative has no overlap with any of the other alternatives and it is 1 when the alternative completely overlaps all other alternatives.

For each alternative route, the distribution of road types, bicycle facilities, total length and the number of left-, right-, and U-turns was recorded. Turns were counted at nodes and were assessed as the angular deviation between the ingoing (from start- to end node) and the outgoing edge (left: $>45^\circ$, $<135^\circ$, U turn: $>135^\circ$, $<225^\circ$ etc.).

An example of route choices is given in supplementary online material, figure 2.

Compared to the route data set, creating choice sets to analyse locomotion (hereafter referred to as edge data set) is far simpler, only involving selecting all edges connected to a node based on the network topology. In our study, we excluded the ‘entry edge’. The choice set includes edge attributes for each edge connected to the present node and in addition deviations between angle from the start to end node of the edge and the angle to the destination of the trip. Inclusion of the angle towards the destination is not entirely in line with the idea of locomotion as it requires a superficial idea of the direction towards the destination. But it was included to reflect some degree of wayfinding.

4.5 Statistical model

The choice between entire routes or between edges at a given node is estimated as discrete choices within the random utility framework (see, e.g. Train, 2003). Thus, the utility of a given choice (route or edge) j for individual n is modelled as:

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

where V_{nj} is an observable part depending on the characteristics and ε_{nj} is stochastic and unobserved. Assuming that an individual is utility maximising, he or she will choose an alternative over another if the utility derived is greater. Using the conditional logit model, the probability of individual n choosing alternative k over a set of J alternatives can be modelled as:

$$Pr_{nk} = \frac{\exp(\mu\beta'X_{nk})}{\sum_j \exp(\mu\beta'X_{nk})}$$

where β is a vector of parameters to be estimated to the corresponding characteristics given by X , and μ is a scale parameter taking into account the variance of the unobserved part of the utility.²

If an attribute is used as numerator, the scale parameter cancels out and the ratio between the attributes can be compared. In the route dataset, a natural denominator is distance, whereby the ratio between a given characteristic and distance can be interpreted as a willingness to travel further to take or avoid a route with this characteristic. For the edge data set, the denominator used is the angle between a given edge (start to end) and a direct, Euclidian approach to the destination, i.e. how much an individual is willing to deviate from a direct route to avoid or obtain certain characteristics.

It is possible to extend the models and make them more advanced so that they can take heterogeneity of various sorts into account. In the present study we accommodated for heterogeneity by also analysing data with a random parameter models, where a distributions is assumed over the parameters taking the panel structure of the data into account. These models yielded similar results as the conditional logit models and consequently, we use and report results of the simplest approach and show results from the random parameter models in the supplementary online material, table A4 and A5.

The code applied to perform the statistical analysis is available in the supplementary online information and can be obtained from the authors.

5. Results

5.1 Route dataset (navigation based on wayfinding)

First we analyse our data based on the assumption that cyclists use a wayfinding strategy, i.e. decide on their entire route from their point of departure. The results are presented in Table 2 below.

The first set of variables all represents some kind of effort in biking; distance, the number of turns, and traffic lights related to the chosen route. As expected, we see that distance is associated with disutility, which means that cyclists prefer shorter routes to longer ones. We also see that the number of turns on a given route is associated with disutility and, that right turns are preferred to left turns.

The next four variables in Table 2 represent different qualities of a bikeway as opposed to none. As expected, the results show that cyclists are more likely to choose a route when the proportion of any of these bikeways qualities increases. At the same time, there seems to be no significant difference between the different types of

² This does not matter for a single model, but if one wants to compare parameter estimates between models, adjustments need to be made.

Table 2. Parameter estimates for a conditional logit model of the route data set.

	Parameter		st. error	p	WTT	95% Confidence interval	
Distance (km)	-4.16	***	0.209	>0.0001			
Left turns	-0.194	***	0.0356	>0.0001	-0.0466	-0.0646	-0.0287
Right turns	-0.0789	**	0.0350	0.0239	-0.0190	-0.0358	-0.00215
Traffic lights	0.308	***	0.0380	>0.0001	0.0742	0.0573	0.0911
Bike track	5.56	***	0.496	>0.0001	1.34	1.068	1.61
Bike lane	6.35	***	1.11	>0.0001	1.53	0.986	2.07
Segregated biking lane	6.60	***	1.06	>0.0001	1.59	1.06	2.11
Biking lane shared with pedestrians	2.58	***	0.65	0.0001	0.620	0.309	0.931
Traffic road	-2.94	***	0.757	0.0001	-0.707	-1.07	-0.345
Shopping street	-7.09		8.75	0.4178	-1.71	-5.82	2.41
Ln Pathsized correction	1.40	***	0.151	>0.0001	0.337	0.269	0.405
N Individuals	179						
N Observations	1,265						
LL	-883.31						
McFadden's R ²	0.74						

*** and ** = Significance at 1% and 5% levels respectively.

bikeways, except for bikepaths that are shared with pedestrians, which has a significantly lower utility than the other bikeways, although it is still positive.

Analysing different road types and using main road as the reference, a traffic road is less popular than other road types, while local roads are the most popular. All other road types, e.g. shopping streets and the amount of green environment along the chosen route are insignificant in terms of influencing the choice of route.

Finally, we see that the variable ‘path size’ correcting for route overlap is significant and positive (Frejinger et al., 2009), which indicates that not correcting for this would introduce a bias.

In Table 2, we also calculate the Willingness to Travel (WTT) estimates by using the distance, measured in kilometres, as denominator. The WTT estimates clearly demonstrate that the preferences for bikeway types dominate the route types. As the variables indicate the proportion of a route that has the given characteristics, it means that a respondent is, at the margin, willing to travel 13.4 m further per percentage increase in bike track separated from the road instead of none.

5.2 Edge dataset (navigation based on locomotion)

We now turn to analysing our data based on choices made between edge connecting nodes along a route. The results are presented in Table 3 and, once again, they show that the variables related to what we term effort, have a negative effect on the choice made at each node. The parameter estimate of ‘Angle to Destination’ shows that the more direct towards the destination an edge is – in terms of the angle – the higher is the likelihood that the cyclist will choose that edge. As was also the case for the results from the route data analysis, we see that turns are related to disutility, although this model seems to show no significant difference between left, right and U-turns.³

The different types of bikeways are also preferred to the reference in this model, but here the estimates are significantly different from each other, with the segregated biking paths (i.e. bike paths that do not run alongside roads with motorised traffic) being the most preferred. Again, the biking paths shared with pedestrians are the least preferred, significantly different from bike tracks and lanes.

The variable that represents the proportion of green environment, such as parks and forests, is negative and significant.

The type of roads also affects the edge choice. Traffic roads and local roads both affect the choice positively (over the reference). The last type of road in the model is shopping streets and the negative sign of the parameter indicates that these are less popular options compared to other kind of roads.

³ Notice that the actual edge that leads to the node where the choice was made was removed. Accordingly, U-turning edges are other edges leading from the node in an opposite direction of the present ($>135^{\circ}$; $< 225^{\circ}$).

Table 3. Parameter estimates of a conditional logit model of the edge data set.

	Parameter		St. error	p	WTT	95% Confidence interval	
Angle to destination	-1.80	***	0.0150	>0.0001			
left turn	-0.948	***	0.0156	>0.0001	-0.527	-0.547	-0.508
Right turn	-0.920	***	0.0155	>0.0001	-0.512	-0.531	-0.492
U-turn	-0.905	***	0.0568	>0.0001	-0.503	-0.566	-0.440
Bike track	0.571	***	0.0268	>0.0001	0.318	0.288	0.348
Bike lane	0.716	***	0.0465	>0.0001	0.398	0.347	0.449
Segregated bike lane	0.735	***	0.0414	>0.0001	0.409	0.363	0.454
Biking lane shared with pedestrians	0.218	***	0.0280	>0.0001	0.121	0.0908	0.152
Percentage green environment	-0.0680	***	0.0130	>0.0001	-0.0378	-0.0520	-0.0237
Traffic road	0.247	***	0.0420	>0.0001	0.138	0.0916	0.183
Local road	0.292	***	0.0281	>0.0001	0.162	0.131	0.193
Shopping street	-0.144	***	0.0249	>0.0001	-0.0803	-0.107	-0.0531
N Individuals	179						
N Observations	83,487						
LL	-34,132.86						
McFadden's R ²	0.49						

*** = Significance at 1% level.

Table 4. Parameter estimates of a conditional logit model of the edge data set, including significant interaction terms with the start or the end of the trip.

	Parameter		St. error	p	WTT	95% Confidence interval	
Start x Angle to destination	0.303	***	0.0327	>0.0001	0.172	0.138	0.206
Angle to destination	-1.761	***	0.0213	>0.0001	-1		
End x Angle to destination	-0.774	***	0.0437	>0.0001	-0.440	-0.494	-0.385
left turn	-1.01	***	0.0171	>0.0001	-57.0	-59.5	-54.6
End x left turn	0.175	***	0.0377	>0.0001	9.91	5.70	14.1
Right turn	-0.983	***	0.0170	>0.0001	-55.8	-58.2	-53.4
End x right turn	0.260	***	0.0370	>0.0001	14.7	10.6	18.9
U-turn	-0.893	***	0.0565	>0.0001	-50.7	-57.1	-44.2
Bike track	0.735	***	0.0221	>0.0001	41.8	39.1	44.5
Bike lane	0.849	***	0.0453	>0.0001	48.2	43.0	53.4
Segregated bikeway	0.688	***	0.0414	>0.0001	39.1	34.4	43.7
Start x bikeway shared with pedestrians	-0.168	***	0.0547	0.0021	-9.56	-15.6	-3.47
Bikeway shared with pedestrians	0.234	***	0.0320	>0.0001	13.3	9.75	16.9
Percentage green environment	-0.0641	***	0.0131	>0.0001	-3.64	-5.10	-2.18
Start x Traffic road	0.168	**	0.0800	0.0355	9.55	0.653	18.4
Traffic road	-0.0192		0.0421	0.6491	-1.09	-5.78	3.60
Start x Shopping street	0.195	***	0.0457	>0.0001	11.1	5.94	16.1
Shopping street	-0.301	***	0.0276	>0.0001	-17.1	-20.2	-13.9
N Individuals					179		
N Observations					83,487		
LL					-33,852.91		
McFadden's R ²					0.50		

*** and ** = Significance at 1% and 5% levels respectively.

Each choice situation (nodes) has been related to a dummy variable, which states whether the edge choice was made during the start (first 25% of all nodes) or the end (last 25% of all nodes). These two dummies were interacted with all the variables presented in Table 3 to see whether preferences change over the route. The results of a parsimonious model, where all insignificant interactions were left out, are presented in Table 4.

In general, the results show that for a number of the parameters the utility vary along the route. The disutility of choosing an edge with a highly conflicting angle relative to the destination is less pronounced in the beginning of the trip, but much more important at the end. The results also show that the disutility of turning left or right diminishes at the end of the route, whereas the negative utility of U-turns remains constant.

For biking paths shared with pedestrians, we see that the positive preference for this is largely eliminated in the beginning of the route. The positive utility related to traffic roads seems to only exist in the beginning of the trip, whereas the negative utility found for shopping streets is less negative in the beginning of the trip.

6. Discussion

Overall, our results show that analysing preferences based on the edge datasets (reflecting the locomotive component of navigation) is possible and reveals similar results, and sometimes more nuanced results than studies based on route choices (discussed in details below). The two models are based on different types of data, and consequently are not easy to compare in terms of statistical performance. However, we compared the number of correct predictions in each dataset by estimating parameters based on random draws of 80% of the samples and then using the estimated parameters to predict choices on the remaining 20% of the samples. In this way, we were able to predict more than 80% of the choices on the edge data set, i.e. where the actual choice at the edges were identical to the predicted choice, and more than 70% of the route choices, i.e. where the predicted route is identical to the entire route actually biked by the cyclist. In addition to this we also compared the length of the predicted route that overlaps the chosen route. For both datasets, the overlap was about 86%. Still, the indicator is not fully comparable, as the overlap for the route data is potentially higher than reported here, due to the structure of the data.⁴ Yet, they do not serve as a firm indicator as to whether commuter cyclists' navigation is dominated by 'wayfinding' or 'locomotion'. This may reflect that cyclists are combining strategies at different times and places along the route between the starting point and destination. This is indicated by the results shown in Table 4, where we show how preferences differed in the beginning and the end of a route.

With respect to the route data set, cyclist's preferences, the results were as expected and largely in line with the findings from other studies. Biking distance and the number of turns on a given route are associated with disutility. An interesting observation is that right turns are preferred to left turns. As traffic in Denmark is on

⁴ For the edge data, we can estimate the exact length of the overlap between predicted and chosen routes. For the route data, we only calculate the length of the correctly predicted routes in relation to the total length of all routes.

the right, the reason for the lower disutility of right turns compared to left turns is probably because right turns are easier to make as they do not involve having to cross the lane of oncoming traffic. At traffic lights, a right turn also involves less waiting time for cyclists who are willing to make the turn even when the light is red⁵. We also found that cyclists prefer routes with a high number of traffic lights which might represent a safer way of crossing roads compared to crossings that are not regulated by traffic lights. We do acknowledge that traffic lights might be installed more frequently along more popular routes, and thus the parameter is rather indicating a confounding effect.

Turning to the edge data set, in general, we see the same pattern as for the route dataset. The angle to destination and turns were associated with disutility, and various types of bikeways were preferred. The variable that represents the proportion of green environment was negative and significant. This is in contrast to the findings sampled from the same population by Vedel et al (2017) but by the use of stated preference data. A possible explanation is that these areas are sometimes darker and less safe compared to built-up environments and furthermore recreational quality of green areas varies (Panduro & Veie, 2013). Further analysis to take the time of day and age into account might confirm or deny this hypothesis, but we were unable to establish a significant relationship to validate this.

Traffic roads (secondary arterial roads) and local roads both affect the choice of route positively. Although traffic roads may have much more traffic in the form of fellow cyclists and motor vehicles, they are often main arteries that offer an easy connection to other roads and central parts of the city. Local roads, on the other hand, might be preferred to other types of roads as they are, in general, quieter and safer.

We have had focus on movement behaviour for the general public's cycling behaviour. For future policy purposes it could also be relevant to consider different user groups' preferences – gender, age but also for people with specific bike facility needs like cargo bikes or bikes for disabled people.

Regarding external validity testing, our results show similar findings as Vedel et al (2017). The average additional cycling distance in that study was 40% compared to 28% found here. The difference can both be due to a hypothetical bias which can be found in stated preferences studies as Vedel et al (2017) and to be collinear with other characteristics in our study causing pure effects of single characteristics to be less likely.

This study and the four similar studies referred to above (Boarch et al., 2012, Menghini et al., 2010, Halldórsdóttir et al., 2015, and Hood et al., 2011) find that disutility is associated with distance and utility is associated with bicycle facilities along roads and segregated bikeways. The disutility associated with turns found in the present study was also reported by Boarch et al. (2012), Hood et al. (2011), and Halldórsdóttir et al.(2015).

⁵ This is illegal in most places, although it is often practiced

Besides an interest in the cognitive aspects of navigation including edge-based choice set generation, a number of additional reasons for further pursuing this type of analysis in future studies on movement behaviour and navigation can be suggested:

- There is no need to generate alternatives as in wayfinding, which means that potential selection biases can be avoided.
- Locomotion as part of navigation can take into account reactions to emerging or sudden, dynamic events. Reaction to unforeseen congestion, changes in traffic signals, malfunctioning street lights, etc. cannot be analysed with an assumption of *a priori* knowledge for obvious reasons.
- Edge-based analysis can be applied on-the-fly, i.e. while navigating. That is, route preferences can be revealed during navigation and applied to forward route suggestions.
- Edge-based analysis can be conducted on the basis of smaller samples (of routes or respondents) than when based on route datasets.
- It may add more detailed estimates for on-line navigation tools for cyclists (see e.g. Google, 2018; Lovelace et al., 2017) and in relation to agent based simulation of bicyclists behaviour (Snizek, 2015; Yeboah, 2014, Smith, et al. 2017).

On the other hand, the analysis of edge datasets can represent an econometric challenge due to the many observations, each of which contains little information, and this is especially the case if heterogeneity is to be addressed further than we do here. Yet it poses an interesting way ahead, e.g. in terms of linking it to the theory of planned behaviour (Ajzen, 1991) which links behaviour to attitudes, norms and perceived behavioural control.

This is an initial attempt to analyse edge datasets, and it has proven to be useful. We applied it to a commuting case, where one would expect the routes to be well-known and preferences well-established. Analysing data in the same way for recreational cyclists or new-comers to an area, e.g. tourists, may reveal new aspects as we would expect locomotive behaviour to be more dominant (especially in the absence of map-based information).

It may be beneficial for future research to address:

- The conditions and situations that in particular would lead to dominance of one or the other navigation mode (Montello and Sas, 2006).
- The development of methods that can analyse wayfinding and locomotion in conjunction, and thereby revealing clusters of behaviour that are influenced by both modes of navigation.
- Methods that may enable comparison of the utility provided by the two modes of analysis. Ways ahead here may be the use of the recursive logit model as suggested by Fosgerau et al. (2013).

Finally a note should be made on the sample and validity of the results: while representativeness may be reasonable, it is worth noticing that there may be a self-selection bias in that people volunteer to track themselves, and it is possible that they as a consequence change behaviour compared to their daily routines. This latter may be avoided to some extent by tracking people over longer time.

7. Conclusion

This study has analysed bicyclists' navigational behaviour in Copenhagen. The findings indicate that navigation and route preference can be analysed based on the assumptions that subjects have perfect knowledge about their environment (in our study termed 'wayfinding') and that it can also be based on perception of the immediate, perceptible environment (here termed 'locomotion'). Analyses of the two choice sets, generated from cyclists' tracks recorded by GPS, reveal similar tendencies of several of the environmental attributes.

Navigation based on wayfinding is the most commonly applied approach in present day route preference studies in transport modelling. A main concern of studies is the generation of choice sets including alternative, unbiased routes. This obstacle can be avoided by applying approaches that assume that navigation is based on locomotion, where the choice set of potential edges is given with no further simulation required. Further, our study reveals new aspects of navigation behaviour as, for example, different preferences along the route. Therefore, this study suggests that future studies of route preferences should consider conducting analyses that are based on edge selection instead of, or in combination with, analyses that are based entirely on choices between routes as in wayfinding navigation.

Literature

- Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2), 179-211.
- Boarch, J. and Dill, J. 2016. Using Predicted Bicyclist and Pedestrian Route Choice to Enhance Mode Choice Models. *Transportation Research Record: Journal of the Transportation Research Board*. Vol. 2564.
- Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, 46(10), 1730-1740.
- Buehler, R., & Dill, J. (2016). Bikeway Networks: A Review of Effects on Cycling. *Transport Reviews*, 36(1), 9–27.
- Buehler, R. and Pucher, J. (2012). International Overview: Cycling trends in Western Europe, North America and Australia. In Pucher, J and Buehler, R. (eds). *City Cycling*. MIT Press.
- Dextre, J.C., Hughes, M., and Bech, L. (2013). Introduction. In Dextre, J.C., Hughes, M., and Bech, L. 2013. *Cyclists and Cycling around the World*. Fondo editorial.
- Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model with unrestricted choice set. *Transportation Research Part B* 56, 70-80.
- Frejinger, E., Bierlaire, M., & Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, 43(10), 984-994.
- Google, 2017. Google maps. <https://www.google.com/maps/about/>. Last accessed, January 2018.
- Golledge, R. and Rushton, G. (eds). (1976). *Spatial Choice and Spatial Behaviour*. Ohio State University Press: Columbus

- Golledge, R.G. (1999). Human wayfinding and cognitive maps. In Golledge, R.G. (ed) 1999. Human cognitive maps and wayfinding. John Hopkins University Press. Baltimore.
- Gössling, S. (2013). Urban transport transitions: Copenhagen, city of cyclists. *Journal of Transport Geography*, 33, 196-206.
- Harder, H., Suenson, V., Thuesen, N., Jensen, A. S., Knudsen, A. M. S., Tradisauskas, N., Stigsen, T.,K., and Weber, M. (2011). Collecting knowledge of biking behavior in Copenhagen using GPS: The GPS data collection. Department of Architecture and Design, Aalborg Universitet.
- Harvey, F., Krizek, K. J., & Collins, R. (2008). Using GPS data to assess bicycle commuter route choice. In Transportation Research Board 87th Annual Meeting (No. 08-2951).
- Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation letters*, 3(1), 63-75.
- Jensen, A. S., Bro, P., Harder, H., Hansen, J. H., & Tradisauskas, N. (2009). Distinguishing movement from stays during continual GPS tracking. In Kortdage 2009, http://vbn.aau.dk/files/18868221/Distinguising_movement_from_stays_during_continual_GPS_tracking.pdf. Last accessed February 2017.
- Jensen, N. (2013). Planning a cycling infrastructure – Copenhagen – City of cyclists. In Dextre, J.C., Hughes, M., and Bech, L. 2013. Cyclists and Cycling around the World. Fondo editorial.
- Krizek, J.K., Handy, S.L. and Forsyth, A. (2009): Explaining changes in walking and bicycling behavior: challenges for transportation research. *Environment and Planning B: Planning and Design* 2009. volume 36. Olgilv e. D., Bull. F., Powell, J., Cooper.
- Lovelace, R., Goodman, A., Aldred, R., Berkoff, N., Abbas, A., and Woodcock, J. (2017). The Propensity to Cycle Tool: An open source online system for sustainable transport planning. *Journal of transport and land use*. Vol 10, No 1 (2017)
- Lynch, K. (1960). *The image of the city* (Vol. 11). MIT press.
- Madsen, T., Schipperijn, J., Christiansen, L. B., Nielsen, T. S., & Troelsen, J. (2014). Developing suitable buffers to capture transport cycling behavior. *Frontiers in Public Health*, 2, [61]. DOI: 10.3389/fpubh.2014.00061
- Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. *Transportation research part A: policy and practice*, 44(9), 754-765.
- Montello, D. R. (2005). Navigation. In Shah, P. and Miyake, A. (eds). (2005). *The Cambridge handbook of visuospatial thinking*. Pp 257-294. N.Y. Cambridge University Press.
- Open Street Map consortium (2017). Home page of the Open Street Map consortium. www.osm.org, last accessed March 2017.
- Panduro, T.E., Veie, K., 2013. Classification and valuation of urban green spaces: a hedonic house price valuation. *Landscape and Urban planning* 120, 119-128.
- Parkin, J. and Koorey, G. (2012). Network planning and infrastructure design. In Parkin, J. 2012. *Cycling and Sustainability* (pp. 131-160). Emerald Group Publishing Limited.

- Pucher, J., and Buehler, R. (2012). Introduction: Cycling for Sustainable Transport. 2012. In Pucher, J., and Buehler, R. *City cycling* (pp. 1-7). MIT Press, 2012.
- Horton, D., & Parkin, J. (2012). Chapter 12 Conclusion: Towards a Revolution in Cycling. In *Cycling and sustainability* (pp. 303-325). Emerald Group Publishing Limited.
- Rushton, G. (1976). Decomposition of space-preference functions. In Golledge, R. and Rushton, G. (eds). 1976. *Spatial Choice and Spatial Behaviour*. Ohio State University Press: Columbus
- Skov-Petersen, H., Zachariassen, M., & Kefaloukos, P. K. (2010). Have a nice trip: an algorithm for identifying excess routes under satisfaction constraints. *International Journal of Geographical Information Science*, 24(11), 1745-1758.
- Smith, L., Hiscock, R., Chapizanis, d., Karakitsios, S., Sarigiannis, and D. (2017). D10.1 - Report on methodology for properly accounting for SES in exposure assessment. WP10: Taking account of socioeconomic status when modelling. <http://www.heals-eu.eu/wp-content/uploads/2013/08/HEALS-Deliverable-10.1-final.pdf>. Last accessed January 2018.
- Snizek, B. 2015. Mapping cyclists' experiences and agent based, modelling of their wayfinding behaviour. PhD Thesis. Dept. of Geoscience and natural resource management. University of Copenhagen.
- Train, K.E, 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, 334 pp.
- Vedel, S.,E., Jacobsen, J.,B., and Skov-Petersen, H. (2017). Bicyclists' preferences for route characteristics and crowding in Copenhagen - a Choice experiment study of commuters. *Transportation Research Part A: Policy and Practice*, 100, 53-64. <http://dx.doi.org/10.1016/j.tra.2017.04.006> 0965-8564
- Wiener, J. M., Büchner, S. J., & Hölscher, C. (2009). Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Computation*, 9(2), 152-165.
- Winters, M., Teschke, K., Grant, M., Setton, E., & Brauer, M. (2010). How far out of the way will we travel? Built environment influences on route selection for bicycle and car travel. *Transportation Research Record: Journal of the Transportation Research Board*, (2190), 1-10.
- Yeboah, Godwin (2014) *Understanding urban cycling behaviours in space and time*. Doctoral thesis, Northumbria University.
- Yeboah, G., & Alvanides, S. (2015). Route Choice Analysis of Urban Cycling Behaviors Using OpenStreetMap: Evidence from a British Urban Environment. In *OpenStreetMap in GIScience* (pp. 189-210). Springer International Publishing.
- Yeboah, G., Alvanides, S., & Thompson, E. (2015). Everyday Cycling in Urban Environments: Understanding Behaviors and Constraints in Space-Time. In M. Helbich, J. Jokar Arsanjani, & M. Leitner (Eds.), *Computational Approaches for Urban Environments SE - 8* (Vol. 13, pp. 185–210). Springer International Publishing.