The Influence of Familiarity on Route Choice: Edinburgh as a Case Study

Maud van Haeren* and William Mackaness†

The University of Edinburgh School of Geosciences

November 5, 2014

Summary

Automatically generated routing instructions are provided by Satnav and Internet based mapping services in order to assist us in getting to unfamiliar places. Instructions from these devices are based around least cost algorithms, described on a street- by- street basis. Taking no account of what we might already know, the instructions are long, difficult to remember and require effort to interpret. If we could opportunistically route the person via known areas, the recognition process would be easier, the instructions could be fewer, and the users would find greater comfort in travelling through spaces familiar to them. In this paper we model a user's heterogeneous familiarity of the city such that it modifies a cost surface, resulting in directions that route the user via familiar spaces. A familiarity index was created based on historical GPS based trajectories. Participant route choice was found to be closer to outputs from the model than simple shortest path.

KEYWORDS: Familiarity, Shortest Path, Cost Surfaces, Navigation, Pedestrian Wayfinding.

1. Introduction

This research aimed to measure familiarity and its influence on route choice made by pedestrians in the City of Edinburgh. Increasingly, satnav devices and smartphones are used to assist in getting to unfamiliar places (Savage et al., 2011; Schmid, 2008; Zandbergen and Barbeau, 2011). The street by street information given by these devices can appear counterintuitive and contain information that is hard to remember. Yet people do not take the shortest path. Various factors might influence their choice; it is certainly known that people like to take advantage of places they know, as this requires less cognitive effort and feels easier to navigate (Demirbas, 2001; Schmid et al., 2010; Gale et al., 1990; Lovelace and Hegarty, 1999; Li, 2006; Papinski et al., 2009; Pahlavani and Delavar, 2014). Regardless of the complexity or simplicity of the environment people are habitual, preferring to retrace routes they travelled before or know rather than exploring new ones (Golledge, 1999). The aim of this research was to improve navigational solutions for pedestrians in urban environments by taking into account familiarity of their environment. This in turn, required us to: 1) create a quantitative measure of familiarity and 2) optimally incorporate this within a shortest- path algorithm. So how might we measure familiarity and how well can such a model predict such route choices? Here we present the model, and comment on its predictive capacity which was evaluated through quantitative methods and street level experiments.

* s1363550@ed-alumni.net

† william.mackaness@ed.ac.uk

2. Methodology

Spatial familiarity can be quantified as a product of revisit times along routes (Pahlavani and Delavar, 2014; Gale et al., 1990). Additionally other studies have looked at using the duration of staying in a certain place as a qualifier to a familiarity index (Meness and Moreira, 2007; Schmid and Richter, 2006). An index based on revisit times of road segments is relatively easy to implement based on the assumption that the more the individual travels along a route, the stronger the representation of the important elements of the environment within their cognitive map (Golledge, 1999; Montello, 1993).

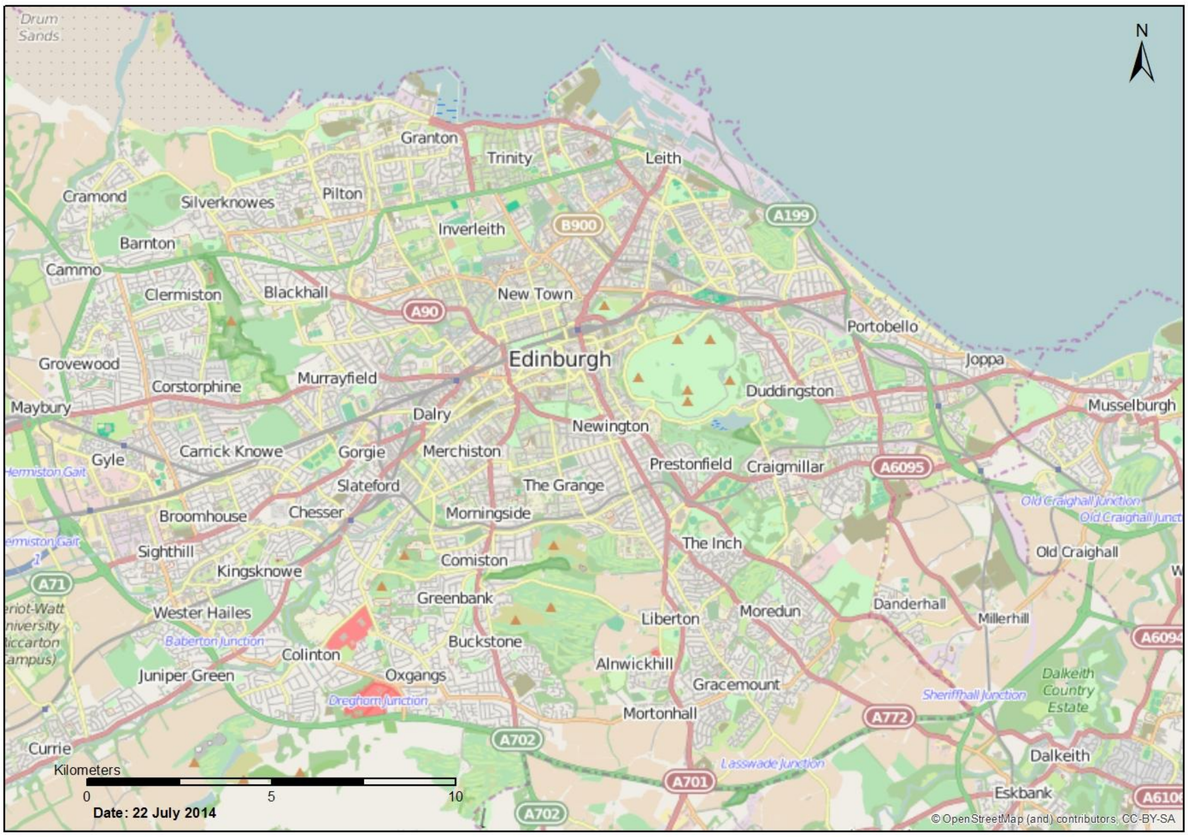

Figure 1 Research study area – City of Edinburgh, UK

In order to quantify familiarity, information about people's whereabouts was collected to create spatial user profiles. The research was focused solely on walking pedestrians. All participants lived and worked in the city of Edinburgh (Figure 1). Ten participants agreed to record their GPS trajectory data and this information was used to identify familiar streets. GPS data was recorded via smartphones which were considered to be sufficiently accurate in capturing location (Zandbergen and Barbeau, 2011; von Watzdorf and Michahelles, 2010). The familiarity index was based on a minimum of seven days' worth of travel. Additionally the participant was asked to participate in lab based experiments following data collection in order to record preferred routes on maps, and for follow up interviews. GPS measurements were extracted using Google's Location History functionality (Google, 2014). Due to changes in sampling rate, WiFi black spots and urban canyoning, some of the familiarity maps were rather patchy in nature (Joshi, 2001; Meness and Moreira, 2007; Ochieng et al., 2003; von Waltzdorf and Michahelles, 2010; Zandbergen and Barbeau, 2011). Figure 2 is an example of a participant's history with its corresponding familiarity network based on revisit times in Figure 3.

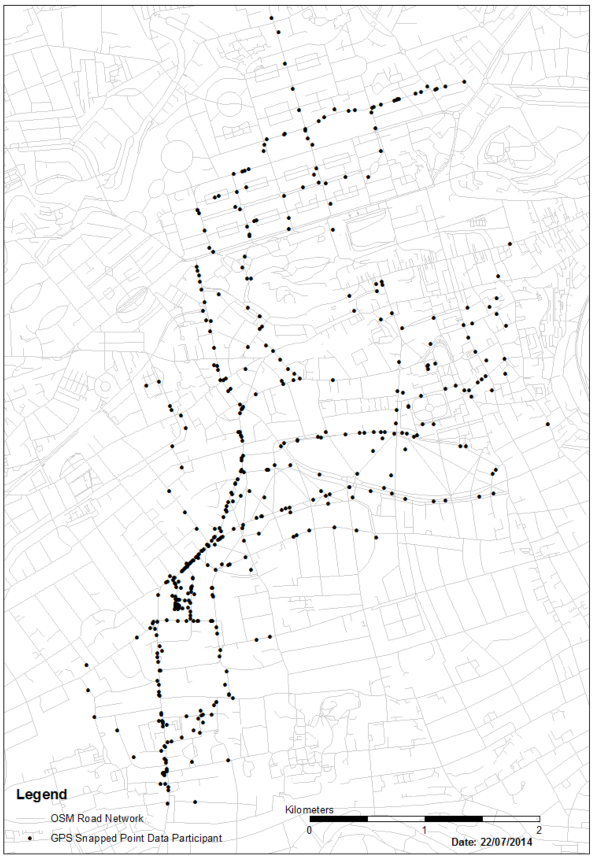

Figure 2 Patchy GPS Coverage Participant 1

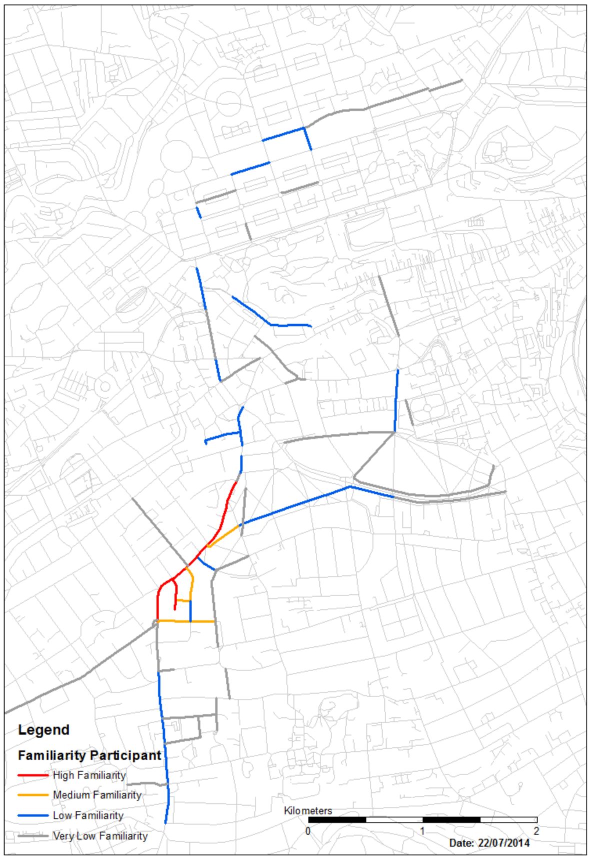

Figure 3 Mapping Familiarity to the Network

When computing just travel distance, Dijkstra's algorithm is a popular choice for calculating the shortest distance between two points in a network (Wise, 2002; Worboys and Duckham, 2004; Sonnier, 2006; Lloyd, 2010). Here we focus on **two** criteria: *travel distance* and *familiarity*. Thus it is a conflicting bi- criterion network problem (Gen and Lin, 2005) – solving for a minimum travel distance and a maximum of familiarity. There are several approaches to solving a multi- criteria path optimization. A traditional way of solving this optimization (and the approach taken here) was first, to linearly weight all independent criteria so that they result in one value for each edge of the network and then second, solve by using Dijkstra's method (Corley and Moon, 1985; Malczewski, 2011). Street level data from OpenStreetMap – OSM (OpenStreetMap, 2014) was used to create the underlying network serving as input in the Network Analyst tool. OSM was chosen because, in urban contexts, it is both topologically accurate and more complete than Ordnance Survey's ITN data (Haklay, 2008).

GPS date stamps were used to count the number of days a participant had revisited a specific street by grouping them based on date of measurement and counting the number of different dates. This count was used to create a Familiarity Index (FI), Equation 1, which is then multiplied by the value of the underlying street network – in this case the length of the road in meters (Corley and Moon, 1985; Malczewski, 2011; Dijkstra, 1959; Ochieng et al., 2003). Thus the cost of the street was reduced, in effect making previously visited edges (streets) more 'attractive' as compared with unvisited/ less visited streets. To the best of knowledge of the authors, defining a Familiarity Index in this manner has not been done before. The street network with these newly calculated total edge weights (Figure 3) served as the basis for network analysis and the application of Dijkstra's Algorithm (Dijkstra, 1959; van Haeren, 2014).

$$FI = 1 - \left(\frac{Revisit\ Days\ Count}{Total\ Number\ Tracked\ Days}\right) \qquad (1)$$

3. Evaluation

A route that was far from familiar regions of the city would not be expected to make large deviations in order to pass through that familiar region; a route falling entirely within a region well known to the pedestrian, might be strongly influenced by their degree of knowledge of it. Therefore to validate the accuracy of the model, origin and destination pairs were chosen that variously crisscrossed familiar parts of the network in order to assess how the paths deviated from shortest path, i.e. taking account of the familiar. Based on visual examination of their GPS- trajectory data, three types of routes were distinguished: 1. Familiar place to familiar place; 2. Familiar place to unfamiliar place; and 3. Unfamiliar place to unfamiliar place.

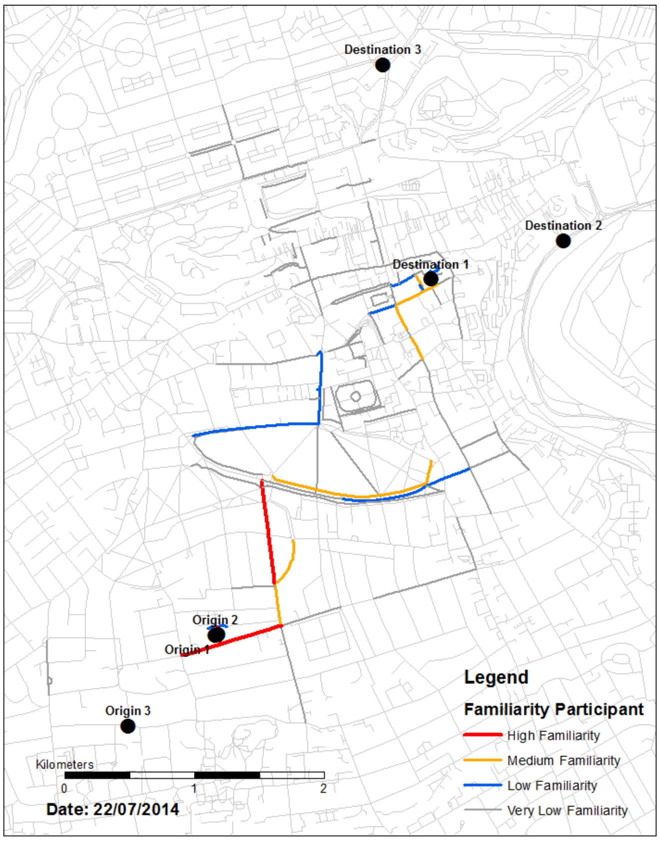

Figure 4 Choosing Origin and Destination Points

To evaluate the veracity of the outputs from the model, participants were asked to draw their own preferred routes on paper. Participants were then asked open ended questions as to their reasons for drawing a specific route (Papinski et al., 2009; Gale et al., 1990). For each of the three different origin – destination situations, three routes per participant were generated: 1) SP (shortest path): using Dijkstra’s algorithm to solve for edge weights of distance only, 2) SPF (shortest path with familiarity): using Dijkstra’s algorithm to solve for a combined edge weight of distance and familiarity and 3) HC (human choice) route as drawn by the participants on paper maps.

In order to quantitatively compare the different outputs, differences in length and differences in the Discrete Fréchet Distance were measured. The Fréchet Distance measures resemblance between lines and takes into account the course of the line (Alt and Godau, 1995; van Haeren, 2014). Figure 5 shows two hypothetical lines and the corresponding Discrete Fréchet Distance. Summation of these distances enables comparison between the three outputs (Danziger, 2011; Eiter and Mannilla, 1994).

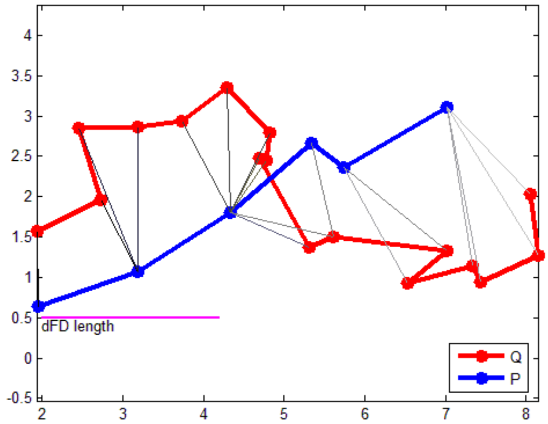

Figure 5 Discrete Fréchet Distance of hypothetical lines P & Q

4. Refinement of the Methodology

Visually comparing the drawn routes against the GPS trajectory data of each individual revealed that people did not always choose routes that took maximum advantage of their familiarity with an area. Figure 6 shows the drawn route takes a detour from what might be expected based on measured familiarity. This could be because 1) there are factors other than familiarity influencing choice, 2) that the model of familiarity was deficient, or 3) that people say one thing, and do another! Golledge (1999) and Montello (1993) both suggest that learning of the environment takes place at the eye level perspective - travel experience gained at the 'environmental scale' (Montello, 1993) whereas learning takes place via the examination of layouts - the 'figural scale' (Montello) using photographs or maps; and that these are different learning processes. Choice patterns and decision processes are not identical at these different perspectives or spatial scales, due to factors such as time and effort (Montello, 1993). This appears to suggest that asking people to draw on paper maps is not the best way of validating the outputs from the model. Therefore a second small sample experiment was set up to improve on evaluation of the outputs. Three of the ten participants were asked to walk from a known origin to a known destination and these were then compared with other outputs.

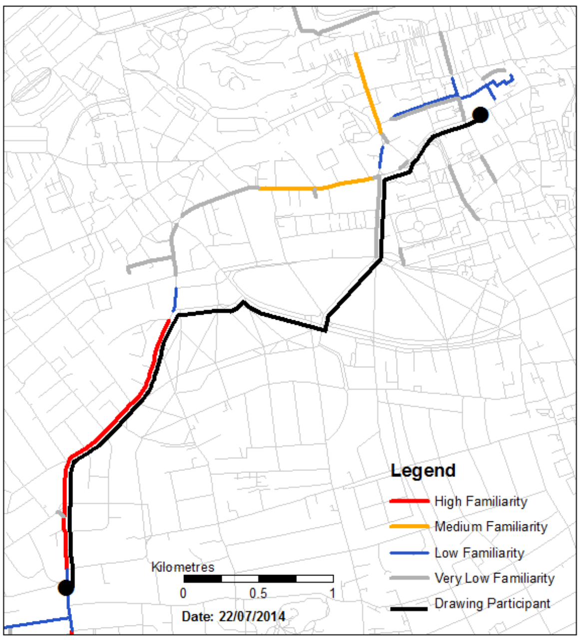

Figure 6 Comparing a participant's familiarity map with what they drew (in the mid-section, one would expect the black line to follow the familiarity path more closely)

5. Results

In order to search for patterns, the results of these comparisons were grouped based on the type of origin - destination combination (Situation 1, 2 and 3). Table 1 shows the averaged results for the ten participants in percentage length difference: Column 1 indicates that they are willing to walk 15% longer distances as compared with the shortest path in order to remain in familiar territory.

Table 1 Length Differences in Percentages

	SP & SPF	SP & HC	SPF & HC	Situation Average
Situation 1	104%	101%	100%	101%
Situation 2	120%	112%	93%	108%
Situation 3	122%	106%	102%	110%
Average	115%	107%	98%	-
	SPF longer	*HC longer*	*SPF longer*	

Table 2 shows the average Discrete Fréchet Distance in meters for the ten participants. The smallest Fréchet Distance, and therefore the smallest shape difference, can be seen when comparing the shortest path with the path calculated based on the Familiarity Index (column 1).

Table 2 Average Discrete Fréchet Distance in Metres

	SP & SPF	SP & HC	SPF & HC	Situation Average
Situation 1	1147	1432	1870	1483
Situation 2	868	513	809	730
Situation 3	707	793	842	781
Average	907	913	1174	-

From the interviews, three ideas emerged as to why participants did not take the shortest path. Firstly, people prefer walking through green spaces as much as possible. Secondly, people take routes which "feel" simple and require less thinking, or are directly related to familiarity. Thirdly, people try to avoid things like busy streets or steep gradients. One could of course argue that all three factors are implicitly reflected in the historical trajectories of the participants. These results furthermore suggest that "cognitive ease" of following familiar routes might not be the only reason as to why certain route choices are made and that many more factors influence this decision making process (Penn, 2001).

Three participants completed a Validation Walk (VW): first walking from a familiar origin to a familiar destination (Situation 1) based on recollection from memory, and then walking from the same familiar destination back to a familiar origin based on what was calculated by Dijkstra's algorithm based on shortest distance in metres only. Figure 7, 8 and 9 show the results of the VW compared to the earlier calculated and drawn routes for the three participants: none of the participants has the exact same VW as HC. Table 3 shows the averaged results for the VW. The difference in absolute length is a bit less for the Validation Walk, meaning that the SPF predicts length better. The Discrete Fréchet Distance also shows that the Validation Walk is more similar to the prediction of the SPF than the route drawn by the participants (HC).

Table 3 Results comparison Validation Walk and Human Choice (experiment 2)

	SPF & HC	SPF & VW
Length Comparison (%)	102%	99%
Discrete Fréchet Distance (m)	714	588

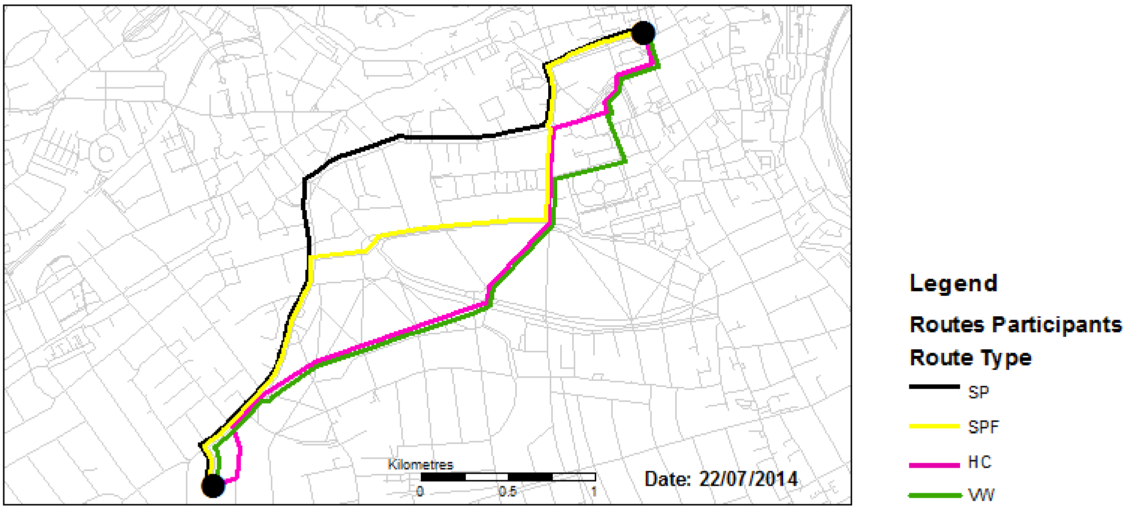

Figure 7 Participant 1 Validation Walks

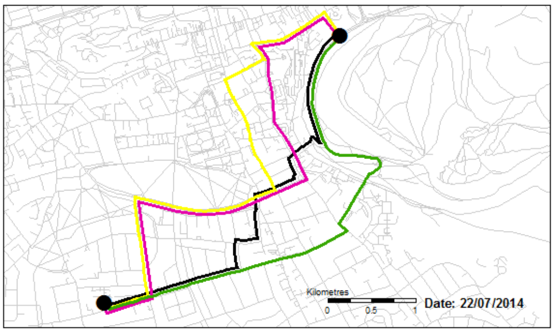

Figure 8 Participant 3 Validation Walks

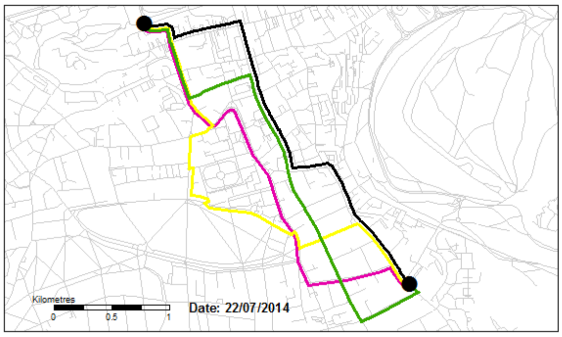

Figure 9 Participant 2 Validation Walks

5.1 Privacy Considerations

Undertaking this research brought into focus issues of privacy – as much can be inferred from both an individuals' digital footprint, and an individuals' data shadow (information others have generated about them) (Koops 2011; Kuebler et al. 2013). People's tolerance for privacy invasion varies with culture and age (Thomas et al. 2013), and people's understanding of how such information is shared (Post and Woodrow 2008). Privacy issues are of increasing concern, given the increasing precision with which LBS devices record location, and the ability to infer activities based on trajectory analysis (Android 2014; Hanson 2005).

6. Conclusion

The research has demonstrated that familiarity can be quantified and used as an edge weight multiplier to modify outputs from Dijkstra's Shortest Path Algorithm. The 'Shortest Path Familiarity' predicted the actual route more accurately when looking at the comparison of the Validation Walk cases versus the Human Choice cases. From the results, the following conclusions can be drawn 1) People are willing to travel 15% further if it means moving through familiar territory; 2) People see the world differently through maps as compared to the world they experience.

The validity of these conclusions is restricted to the constraints in the data set, sample size and other limitations (van Haeren 2014). To further explore these conclusions additional research is recommended: 1) development of a richer model of familiarity that incorporates data collection over a longer period of time, and gives priority to recency of visitation; 2) deeper understanding of the 'in filling' process by which people 'connect' familiar spaces together and how this influences their route choice; and 3) Inclusion of motive for choice of route (social dimension, time of day, urgency, and satisfying other 'on route' tasks).

7. Acknowledgements

We would like to acknowledge all ten participants who have dedicated their time and data to this research.

8. Biography

Maud van Haeren recently completed the MSc GIS at The University of Edinburgh and is now living and working in The Netherlands. William Mackaness is a senior lecturer at The University of Edinburgh in the School of GeoSciences.

9. References

Alt H and Godeau M (1995). Computing the Fréchet Distance Between Two Polygonal Curves. *International Journal of Computational Geometry & Applications*, 5(1&2), 75 – 91.

Android (2014) 'Google Android Location APIs, [online], available: http://developer.android.com/google/play-services/location.html

Corley H W and Moon I D (1985). Shortest Path in Networks with Vector Weights. *Journal of Optimization Theory and Applications*, 46(1).

Danziger Z (2011). Discrete Fréchet Distance Calculations for MATLAB.

Demirbas G U D (2001). Spatial Familiarity as a Dimension of Wayfinding. Unpublished thesis, Bilkent University.

Dijkstra E W (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik,* 1, 269 – 721.

Gale N, Golledge R G, Halperin W C, Couclelis H (1990). Exploring Spatial Familiarity, *The Professional Geographer*, 42(3), 299 – 313.

Gen M and Lin L (2005). Multiobjective Hybrid Genetic Algorithms for Bicriteria Netowrk Design Problem. *Graduate School of Information, Production & Systems Waseda University Japan.*

Golledge R G (1999). Wayfinding Behaviour: Cognitivie Mapping and Other Spatial Processes, Baltimore, Maryland USA: The Johns Hopkins University Press.

Google (2014). Google Location History [online], available: http://maps.google.com/locationhistory [accessed 02/05/2014].

Haklay M M (2008). How Good is OpenStreetMap information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets for London and the rest of England. *Under review in Environment & Planning B.*

Hanson, J. (2005) 'Mobile Data Device and Method of Locating Mobile Data Device'. Google Patents [online], available: http://www.google.co.uk/patents/US20050153681 [accessed 02/05/2014].

Joshi R R (2001). A New Approach to Map Matching for In- Vehicle Navigation Systems: The Rotational Variation Metric. *Translated by Oakland (CA),* USA.

Koops, B. J. (2011) ‘Forgetting Footprints, Shunning Shadows. A Critical Analysis of the “Right to be Forgotten” in Big Data Practice’, *SCRIPTed*, 8(3), 229-256.

Kuebler, K., Palm, D. and Slavec, A. (2014) ‘The Ethics of Personal Privacy and Location Based Services’ in *Confronting Information Ethics in the New Millenium*.

Lloyd C D (2010). Spatial Data Analysis: An Introduction for GIS users. Oxford: Oxford University Press.

Lovelace K L and Hegarty M (1999). Elements of Good Route Directions in Familiar and Unfamiliar Environments. *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science,* Berlin: Springer, 65 – 82.

Malczewski J (2011). Local Weighted Linear Combination. *Transactions in GIS,* 15(4), 439 – 455.

MATLAB R2013a (2013).

Meness F and Moreira A (2007). Using GSM CellID Positioning for Place Discovering. Portugal: University of Minho, Department of Information Systems.

Montello D R (1993). Scale and Multiple Psychologies of Space. *Spatial Information Theory: A Theoretical Basis for GIS,* Berlin: Springer- Verlag, 312 – 321.

Ochieng W Y, Quddus M, and Noland R B (2003). Map- Mathcing in Complex Urban Road Networks. *Revista Brasileira de Cartografia*, 55(02).

OpenStreetMap (2014). OpenStreetMap, [online], available: http://www.openstreetmap.org/ [accessed 02/05/2014].

Pahlavani P and Delavar M R (2014). Multi- Criteria Route Planning Based on Driver’s Preferences in Mutli- Criteria Route Selection. *Transportation Research Part C: Emerging Technologies,* 40, 14 – 35.

Papinski D and Scott D M (2011). A GIS- based Toolkit for Route Choice Analysis. *Journal of Transport Geography*, 19(3), 434 – 442.

Papinski D, Scott D M and Doherty S T (2009). Exploring the Route Choice Decision- Making Process: A Comparison of Planned and Observed Routes Obtained using Person- Based GPS. *Transportation Research Part F: Traffic Psychology and Behaviour,* 12(4), 347 – 358.

Penn, A. (2001). Space Syntax and Spatial Cognition. *Proceedings 3rd International Space Syntax Symposium Atlanta.*

Savage N S, Baranski M, Chavez M E and Hollerer T (2011). I’m Feeling LoCo: A Location Based Context Aware Recommendation System. Translated by Vienna: Springer.

Savage N S, Chun W, Chavez M E and Hollerer T (2012). Seems Familiar: An Algorithm for Inferring Spatial Familiarity Automatically. *Santa Barbara, California: Computer Science Department*, University of California.

Schmid F (2008). Knowledge- Based Wayfinding Maps for Small Display Cartography. *Journal of Location Based Services*, 2(1), 57 – 83.

Schmid F (2010). Personal Wayfinding Assistance. *Dissertation zur Erlangung des Grades eines Doktors der Ingenieurswissenschaften,* Bremen: Universitat Bremen.

Schmid F and Richter K F (2006). Extracting Places from Locational Data Streams. *UbiGIS 2006 – Second International Workshop on Ubiquitous Geographical Information Services.*

Sonnier D L (2006). Parallel Algorithms for Multicriteria Shortest Path Problems.

Thomas, L., Little, L., Briggs, P., McInnes, L., Jones, E. and Nicholson, J. (2013) 'Location Tracking: views from the older adult population', *Age Ageing*, 42(6), 758 – 63.

van Haeren M (2014). Technical Report Part II: A Personalised Routing Algorithm. *Unpublished thesis (MSc Geographical Information Science)*, The University of Edinburgh.

von Watzdorf S and Michahelles F (2010). Accuracy of Positioning Data on Smartphones. *LocWeb.*

Wise S (2002). GIS Basics. London: Taylor & Francis.

Worboys M and Duckham M (2004). Fundamental Spatial Concepts. *GIS: A Computing Perspective,* second ed. London: Boca Raton, 426.

Zandbergen P A and Barbeau S J (2011). Positional Accuracy of Assisted GPS Data from High Sensitivity GPS- enabled Mobile Phones. *Journal of Navigation*, 64(03), 381 – 399.