FINAL REPORT
Decipher Travel Behavior Using Longitudinal Data: The Psychology of Route Choice in Familiar Networks

Date of report: January 2018

Mohan Venigalla, Associate Professor, George Mason University
Xi Zhou, Graduate Research Assistant, George Mason University
Shanjiang Zhu, PhD, Assistant Professor, George Mason University

Prepared by:
Side and Reva Dewberry Department of Civil, Environmental, and Infrastructure Engineering, George Mason University
4400 University Drive, MS 6C1, Fairfax, VA, 22030

Prepared for:
Transportation Informatics Tier I University Transportation Center
204 Ketter Hall
University at Buffalo
Buffalo, NY 14260
**Abstract**

It is widely accepted that path choice of a trip is dependent on trip characteristics, network attributes, and traveler’s personal characteristics. The best-known network variables that influence route-choice are travel distance and travel time. This research attempts to study the influence of other network variables, namely signals, turns and roadway classification on route choice. Real world trip data from path trajectories tracked by Global Positioning System (GPS) in an urban area are used to isolate nearly 5,700 unique real paths. Procedures to compute theoretical shortest time path (STP) and shortest distance path (SDP) based on travel time and distance as impedance variables, respectively, are developed. Street network data is augmented with data on signalized intersections. Procedures to identify turns, quantify turn penalties and road classes along the real and theoretical paths are developed. The real paths are compared to their STP and SDP counterparts to identify discernible relationships between the network variables and the path choice.
Acknowledgements

This study is funded by the TransInfo University Transportation Center. The authors thank Dr. Mark Houck and Dr. Chaowei Yang of George Mason University for their help in reviewing the research work and offering valuable suggestions during the course of this study.

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation’s University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.
Contents

1. INTRODUCTION ............................................................................................................................................... 5

2. METHODOLOGY .................................................................................................................................................. 8
   2.1 Real Path Identification ................................................................................................................................. 10
   2.2 Identification of Trip Ends ............................................................................................................................ 11
   2.3 Map-matching Algorithm .............................................................................................................................. 13
   2.4 Path Screening .............................................................................................................................................. 13
   2.5 Computation of Shortest Paths ..................................................................................................................... 15
   2.6 Signals and Turns along Paths ..................................................................................................................... 16
   2.7 Turn Penalties ............................................................................................................................................... 17
   2.8 Road Classes along Paths ............................................................................................................................. 18

3. EFFECT OF NETWORK CHARACTERISTICS ON PATH CHOICE ................................................................. 20
   3.1 Effect of Signals ........................................................................................................................................... 23
   3.2 Effect of Turns Regardless of Signal Status .................................................................................................. 25
   3.3 Effect of Turns at Signalized Intersections ................................................................................................ 27
   3.4 Effect of Turns at Signalized Intersections ................................................................................................ 29
   3.5 Effect of Road Classification ....................................................................................................................... 30
   3.6 Primary and Secondary Road Usage ............................................................................................................ 32

4. CONCLUSIONS ................................................................................................................................................ 34

5. REFERENCES .................................................................................................................................................... 37
1. Introduction

On a typical street network, travelers are faced with a choice among multiple paths for travel between the origin and destination (O-D pair) of their trip. Given the characteristics of the trip (purpose, time, origin, destination, and travel mode, etc.), attributes of the alternative routes (network variables), and traveler’s personal characteristics, the trip maker chooses the “best” route through the transportation network following some criterion (Antonisse, Daly, and Ben-Akiva, 1989). This route-choice problem may also be stated in the form of a rhetorical question, “What is the most practical path from point A to point B?” However, it is widely accepted that there is no single answer to this question. The best-known network variables that influence route-choice are travel distance and travel time. Route-choice criterion in the real world is not necessarily based on the path with the shortest travel time, travel cost or a combination thereof (impedance). Rather, most drivers choose a route that they perceive to be the best according to their personal preferences, knowledge and experience (Liu, 1996).

The route choice behavior is of interest to transportation modelers. The traffic assignment methods used in the traditional four-step transportation planning process require an adherence to certain path or route-choice behavior. The framework proposed by Wardrop and Whitehead (1952) provided a theoretical basis for formulating and solving the traffic assignment problem as User Equilibrium (UE) and System Optimal (SO) optimization problems (Chatterjee and Venigalla, 2003). Wardrop’s first and second principles of equilibrium, popularly known as Wardrop Criteria, assume a generalized route choice behavior of individuals (Venigalla, Chatterjee and Bronzini, 1999).

Both UE and SO problems involve only the minimization of individual or system-wide impedance. Despite the popularity of the UE method in the present day travel demand modeling exercises, there is no empirical evidence that validates real world route-choice behavior conforming to Wardrop Criteria (Correa and Stier-Moses, 2011). Research has shown that numerous criteria, which could be used to formulate a route and route-choice application, are more stochastic than deterministic in nature. Therefore, assuming travel time (or distance) as the sole criterion of route-choice may be an overly simplistic abstraction of individual driver behavior and may result in an inaccurate representation of traffic in transportation models.

A number of probabilistic route-choice models have also been developed and studied on the basis of Wardrop’s Criteria. A few route-choice models are also derived from utility theory,
which states that each person tries to maximize utility when faced with a choice among competing routes. On the basis of the relative importance of factors of influence, the route-choice model first identifies the set of sufficiently attractive alternatives for specific travelers. Travelers make their choices from this set; with the chosen route being the one that best satisfies their needs and is consistent with their personal constraints and preferences. A variety of logit / probit route-choice models were developed that vary by the basic structure of the model. These models include multinomial probit (MNP) and multinomial logit (MNL) model (Daganzo and Sheffi, 1977); C-logit model (Cascetta et al., 1996); the implicit availability/perception (IAP) logit model (Cascetta and Papola, 1998); and the path-size logit model (Ben-Akiva and Bierlaire, 1999). Other commonly used logit-based models include the PCL model (Chu, 1989); the CNL model (Vovsha, 1997); and the Logit Kernel or Mixed Logit model (McFadden and Train, 2000; Ben-Akiva and Bolduc, 1996).

Other than the best-known factors, travel distance and travel time, research has shown that route-choice behavior is also influenced by demographic variables (such as age, gender, profession, or household structure), road and traffic conditions (such as time of day, travel cost, road classification, or congestion), trip characteristics (such as time of day, purpose or mode), and environment conditions (such as weather or accident) (Jan et al., 2000).

Studies have been conducted to identify the influence of travel time reliability (Jackson and Jucker, 1981); travel information (Stern et al., 1993; Abdel-Aty et al., 1994; Polydoropoulou et al., 1994; Srinivasan and Mahmassani, 2000; Chen and Jovanis, 2003); and usage of freeways (Li, 2004) on path choice behavior. Jackson and Jucker (1981) found that travel time reliability, defined as the difference between the 90th percentile and the median travel times, could be an important influence factor on commute route-choices. Travel time reliability may be positively correlated with criteria like number of traffic signals along a route or the safety of the route. However, the study did not explore this relevance further.

The decision-making process in route choice is a learning process, which is central to the driver’s cognition (Polydoropoulou, et. al., 1994). Therefore, information acquired through the experience of earlier travel choices is considered when a driver is making the next decision. Srinivasan and Mahmassani (2000) have discussed the influence of travel information on route-choice theoretically. Abdel-Aty et al. (1994) conducted a stated preference survey in Los Angeles that indicated that about 60 percent of drivers listen to a travel information report. The
study did not indicate how the drivers responded to this information. Two other surveys in Sweden and Israel found that on average two-third of commuters will change their travel behaviors based on real-time information (Stern, Holm and Maarseveen, 1993).

Chen and Jovanis (2003) investigated drivers’ responses to incident and congestion information by studying whether drivers would follow the guidance to change routes if they were advised to turn or switch to a freeway. Although the study found that the influence of advice on turns is significant, it did not reveal whether drivers tried to minimize the number of turns along the whole route, as the guidance suggests only which road-link drivers should take immediately next without showing guidance for the rest of the route.

An investigation of route-choice behavior on morning commutes (Li, 2004) compared routes that commuters most frequently chose over alternatives. Statistical analysis showed that the primary routes employed a higher freeway percentage and fewer signals than alternative routes. However, primary routes have not been compared to the computed shortest paths.

Jan, Horowitz and Peng (2000) have used Global Positioning System (GPS) data to investigate variations in path choice. They found that travelers often took paths that greatly deviated from the shortest paths, but they did not explain why travelers choose these routes.

Compared to suburban or rural areas, urban areas tend to have a much higher density of intersections. Therefore, the number of traffic lights and turning movements along the path may significantly affect route-choice. Zhuang, Gong, He, and Xu (2012) examined 50 trips between four O-D pairs, and found that the experienced routes (GPS tracking routes used by taxis) have lower frequencies of signalized intersections and turning movements than theoretical shortest path routes. However, there is still a deficiency in literature with respect to impacts of these two specific factors. Past studies did not reveal how these factors impacted path choice or qualitatively identify these impacts.

A number of studies based on stated preference surveys have been carried out to identify factors influencing route-choice other than travel time and distance. Some studies pointed out that drivers tend to minimize signals and turns along the path. However, very few studies presented any empirical evidence based on field data to support this observation. Among those studies that considered empirical data, either the sample sizes were too small, or these studies did not investigate influential factors in depth. Most recently, Zhou (2014) and Zhou and Venigalla
(2014) studied the impact of traffic signals and turning movements along the path on route choice.

This research aims at analyzing a large dataset of real-world trip data to study the influence of the aforementioned network variables on route-choice. The study objective is to determine the extent to which turn movements, signalized intersections and roadway type have influence on route choice. Furthermore, the study will explore the influence of roadway class on path choice. The scope of the study includes statistical analyses on real-world trips tracked by GPS equipment to examine the influence of signals and turning movements on drivers’ path choices in an urban street network. Originally, the research team would like to collect the real-world longitudinal travel trajectory data using smartphone apps that were made available in both iOS and Droid app stores. The apps, as shown in appendix, were designed to provide on campus parking information to users and in return, users provide location data to support travel behavior study. However, the authors found that the allowable density of such location data as regulated by iOS and Droid was reduced significantly during the past few years due to privacy concerns and can not support the objective of this study. Therefore, the research team used an alternative longitudinal travel trajectory datasets collected in previous study instead. Future studies will continue to explore the feasibility of using longitudinal data collected using smartphone to support travel behavior analysis.

In this study, real-world trips made in the metropolitan area of Minneapolis-St. Paul (Twin Cities), Minnesota, during the period from September to December 2008 and the associated network data are extracted and processed with customized Geographic Information Systems (GIS) applications.

2. Methodology

Figure 1 illustrates the study methodology employed for examining the factors influencing drivers’ real world path choices. As the schematic indicates, given a digitized street network in the form of a set of nodes and links, and trip trajectory data (as a series of points) collected by GPS devices mounted on vehicles, the real paths that drivers actually chose can be identified. It is further hypothesized that, by conflating the street network attributes such as signalized intersections and road classes with the paths, additional insights into the route choice behavior can be gleaned.
Figure 1: Study Methodology for Analysis of Factors Influencing Path Choice

GPS-tracked path trajectories collected for a study in the metropolitan area of Twin Cities (or study area) during the period from September to December 2008, are acquired. The dataset contains information on trips made by 44 randomly selected volunteers in the 21 to 65 year age group. All the drivers are local to Twin Cities and presumably have fairly good knowledge of the alternative routes in the network. The volunteers commuted alone and made travel choices without any instructions. A GPS device was installed in the vehicle of each study participant. The device recorded trajectories of each vehicle at a frequency of one GPS location point per second. The geographic location and time stamps of each point were documented and projected onto the ArcGIS shape files for post-processing.
The street network data covering seven counties in the study area are obtained from the U.S. Geological Survey’s (USGS) National Geospatial Program. The network contains full paths of all trips in the dataset. The GPS points are first snapped to the street network, generating the set of real paths. Each real path is then represented by a sequence of nodes, as well as a sequence of links. The first and last nodes of each path are identified as origin and destination of the trip, respectively. GPS data are analyzed to compute turn penalties for incorporation in path search algorithms. Using Python scripting, Dijkstra’s algorithm (1959) is implemented as the path search algorithm, which is then used for finding shortest path between each O-D pair in the data set. The GPS time-stamps are used to obtain travel speeds on road links, by hour of day. For links where GPS trajectory data are not available, posted speed limits are used to compute travel times. Thus computed travel times are used in path search algorithms where travel time is used as impedance.

The data on signals are acquired separately. An algorithm to identify left-turns, right-turns and through movements has been developed. By overlaying on the road network, network variables that are relevant for the paths in the solution space and turns are identified and tagged appropriately. Paired sample t-tests between real paths and shortest paths are then conducted on these factors to analyze their influence on path choice.

2.1 Real Path Identification
A three-step methodology is developed to identify real paths as shown in Figure 2. The first step generates individual trips from the dataset containing GPS-tracking points. Second, these trip points are snapped to the street network by a map-matching algorithm, yielding paths represented by both node sequence and link sequence. In the last step, further path screening eliminates invalid paths, producing the set of real paths for later analysis.
3.2 Identification of Trip Ends

The vehicle trajectories in the database are first distinguished trip-by-trip for subsequent analysis. Ideally, if the interval between two successive points is more than a few seconds, the two points may be treated as points belonging to two different trips. However, when a satellite signal is lost, while the driver is still on the trip, the data on trip trajectories may be broken. This could lead to erroneously splitting a single trip into more than two trips. Another exception also can occur. If the driver finished a trip when the GPS device was still on, the dwelling time would be included and the two trips would be treated as one.

Processing of trip data to identify individual trips from the large set of GPS data involved the following effort:
1) It is assumed that the vehicle may not stay on a digitized road at the end of a trip. This assumption is expected to help screen out the data pertaining to trips in which the GPS device is still on when the driver has completed a trip. As the trajectories are not aligned with a network link, a threshold is needed. With “trial and error” and a random manual check, it is determined that a threshold of 30 meters could provide a good measure. Thus identified GPS points that are away from the road network are removed from the dataset.

2) Determining the minimum possible time gap requires an identification process for the next trip. Intuitively, there must be a threshold below which a new trip is not possible. However, if the gap is larger than the threshold, further analysis would be needed to determine whether it is because of a trip-end or just signal loss period. Previous research showed that 30 seconds is a good threshold for the minimum time gap (Du and Aultman-Hall, 2007).

3) Distinguishing signal loss from trip-end for the time gap longer than the minimum threshold of 30 second is the third important step in identifying individual trips. If the time gap were caused by signal loss, the average speed of the vehicle during this gap would not be much different from average speeds before and after the gap, if the driving pattern were assumed constant. The average speed during the gap can easily be estimated from the time and distance recorded by GPS devices. The highest free flow speed is 70 mph, and average speeds on most streets are no more than 50 mph. Therefore, it would be reasonable to identify a trip-end when the average speed during the time gap is 50% less than speeds before and after the gap. Du and Aultman-Hall (2007) recommended using 20 points before and after the gap to calculate the assumed driving speed. This proved enough to obtain successful identification.

This methodology has proved to be very effective in identifying individual trips. However, certain special situations posed challenges to this methodology. For example, if the vehicle met with traffic congestion or traveled from an expressway to a local road when the GPS device was experiencing a signal loss, a continuous trip would be split because the average speed reduction could be over the threshold. Such instances are assumed to be rare and could not be avoided. A few random checks have shown that this identification process could yield better results than the results obtained only using the original time stamps recorded by GPS devices.
2.3 Map-matching Algorithm

A route can be obtained by connecting GPS points according to the time order in which they are recorded. However, the route may not match any links on the street network in many cases, due to either an error in the GPS location or an inaccurate digital road network. It is necessary to snap the GPS points of each trip to the digital street network by the map-matching technique. Real trip paths are identified in the format of node and/or link sequences between trips’ origins and destinations.

Numerous studies have developed procedures to perform map-matching effectively and accurately (Bernstein and Kornhauser, 1996; Quddus, Noland, and Ochieng, 2006; Quddus, Ochieng, and Noland, 2007; White et al., 2000). Based on the recommendations of these studies, embedded tools in ArcGIS are used extensively for map matching. Since GPS collects points second by second, multiple points may have a common nearest link, thus, the link will not be kept in the sequence repeatedly if it is the same as its previous link. Where necessary, custom tools within or outside the ArcGIS environment are developed using Python scripts.

The set of candidate links identified by ArcGIS may contain links that are not on real paths used by trips in the dataset. On the other hand, a few links forming a real path may be missing, because neither the GPS device nor the digital street map is 100% accurate and/or compatible. To screen out the incorrect links and find back the missing links, further processing of the digital network was needed. This processing required examination of connection between successive links to make sure the link sequence is consistent with the real travel route and direction. The map-matching algorithm developed, also in Python script, for this study eliminates wrong links, while at the same time retrieving missing links. The algorithm also generates a new node sequence that is consistent with the actual trip.

About 8% of total trips (1,668 out of 20,174 trips) failed in map matching due to errors attributable to the digital street network (for example, topological errors, missing links, or incorrect link configuration).

2.4 Path Screening

The map-matching process reduced the data to 18,560 trips from 20,174 trips. For identifying traversable paths in the network, the data are further screened by adhering to the following criteria and by eliminating the trips that do not fit the criteria:
a) A path requires at least two links. If a path contains fewer than three nodes, the trip is eliminated from the analysis dataset.

b) Trips with path lengths shorter than one minute in travel time indicate a potentially faulty GPS device and therefore are dropped from the analysis dataset.

c) If multiple paths made by the same driver are identical to each other, they are assumed to be commuting trips. Since such duplicate trips do not provide additional insights into the route-choice behavior of the driver, making those trips, only one of these multiple trips is included in the analysis dataset.

After this screening process is complete, the remaining 5,694 trips were identified with a valid path represented by both a link sequence and a node sequence. The first and last nodes in the path are tagged as the origin and the destination, respectively. This O-D node set is then used for computing theoretical shortest paths.

Figure 3 depicts the distribution of street links that are used by one or more paths. This map indicates that identified valid paths primarily occurred in the urban areas. Links used over 100 times by the trips in the analysis dataset (thick red lines) are mostly within downtown areas. In the suburban areas, the most used links are primary or secondary roads, and none of them was used more than 10 times.
2.5 Computation of Shortest Paths

Custom Python scripts are developed for implementation of shortest path algorithms. The script uses *Forward Star* notation to describe the street network. The notation represents the network as an adjacency list:

\[ G = \{ n \in N, n = Node(id, x, y, signal, FS(n)) \} \]  

Where \( G \) is the graph, \( n \) is the node that belongs to a set of nodes \( N \), with an identification tag \( id \), coordinates \( x \) and \( y \), signal existence, and a link set named forward star.

A forward star of a node is a set of all links starting from this node. The forward star notation is a representation of network connectivity. The Dijkstra algorithm (Dijkstra, 1959) was used in the script for finding shortest paths.

For a given O-D pair in the trip set derived from the real path identification (i.e. for each real trip in the dataset), the script generates two paths. The first path is based on shortest travel time (shortest time path, or STP), and the second path is based on shortest distance (shortest
As in the case of real paths, each of the two computed shortest paths is
represented by a node sequence as well as by a link sequence. Link sequence or node sequence
along these three paths (real path, STP and SDP) may or may not be the same. Comparing the
node/link sequences of any two paths can identify whether or not they are identical to each other.
Only when two paths have the same number of nodes/links in their sequences, and at every
corresponding place of these two sequences, node/link in one sequence, the node/link is exactly
the same as each other, can two paths be regarded identical.

2.6 Signals and Turns along Paths

The signal information was obtained from the local jurisdiction responsible for signal
operations in the Twin Cities. The signal data, which is a point feature among the GIS layers
used in the analysis, is conflated to the nearest intersection based on the coordinates. Each
intersection node in the network is tagged with a Boolean attribute to indicate whether a signal
exists at that location. Such tagging enables the tracking of number of signals in each path by
simply going through the node sequence for the path. It should be noted that for simplicity, the
data on signals is limited to tagging with mere presence of signals. The data contains no detailed
information such as number of phases, cycle lengths or approach delays.

The most challenging task in the development of the Python scripts for this research is
the development of a procedure that identifies left or right turning movements along a given
path. The procedure performs geometric calculation on nodes and the links along the paths (real
or computed) under examination.

A turning movement is identified when the angle between two successive links is greater
than 45 degrees. As shown in Figure 4, this angle $\theta$ is the absolute value of the difference by
subtracting angle $\alpha$ from angle $\beta$. As the figure illustrates, angle $\alpha$ and angle $\beta$ are angles
between links and the positive direction of x-axis. Using coordinates of the three nodes A, B, and
C, which form the two links, angles $\alpha$ and $\beta$ can be calculated using the equations shown.
Starting from the first node in the path sequence, for each three-node group the procedure
computed the angle $\theta$. If $\theta$ is greater than 45°, a turning movement is recognized to occur at the
node B. Depending on the direction of trip and turn, the movement is flagged as a left or right
turn. The procedure’s logic for identifying the left or right turn is also illustrated in Figure 4.
The procedure goes through the whole path-link sequence to count the number of left- and right-turns along the paths studied. Because length of the paths varies, the counts of signals and turns along the path are normalized as number of signals or turns per 1.61 km (per 1 mile) for later comparisons.

2.7 Turn Penalties

One of the key determinants of path selection process is delay associated with turns along the way. The delays can be due to the presence of signals and/or turn maneuvers. Search algorithms have to account for these turn-penalties for realistic finding paths between any given origin and destination. Literature on development and implementation of these algorithms has been fairly limited. Very few studies have focused on appropriate values for turn penalties and on how to quantify these penalties. Thériault et. al., (1999) have chosen penalties of 24 seconds for left turns and 12 seconds for right turns. In another study, 30 seconds and 7.5 seconds were thought more suitable penalties for left and right turns, respectively (Yiannakoulas and Svenson,
Using the GPS-tracked path trajectory data used in this study, a methodology was developed to ascribe turn penalties for signalized and non-signalized intersections (Zhou, 2014, Chapter 7). Table 1 presents a consolidated summary of turn-penalties obtained by applying this methodology to the path trajectory data. Shortest path search algorithms take into account the results of this analysis.

### Table 1: Average Turn-Penalties from Path Trajectories

<table>
<thead>
<tr>
<th>Turning Movements</th>
<th>Number of Turns</th>
<th>Average Turn Penalties (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signalized Left Turn</td>
<td>3,299</td>
<td>29.41</td>
</tr>
<tr>
<td>Signalized Right Turn</td>
<td>2,960</td>
<td>15.06</td>
</tr>
<tr>
<td>Non-signalized Left Turn</td>
<td>7,105</td>
<td>14.71</td>
</tr>
<tr>
<td>Non-signalized Right Turn</td>
<td>5,465</td>
<td>12.51</td>
</tr>
</tbody>
</table>

### 2.8 Road Classes along Paths

For analytical convenience, all the 23 road classifications in the network identified by Census Feature Class Codes (CFCC) are combined into three major classes: Primary roads (CFCC code A1 and A2), secondary roads (CFCC code A3) and local roads (CFCC codes A4, A51 and A73) In terms of number of links and total link length, the class of local roads is dominant in the network. The primary road class has values generally higher than, but comparable to, the values for secondary road class (Zhou, 2014).

To exclude the possibility that drivers have to choose a certain road class merely due to the composition of the network rather than their preferences, a concept of “road availability” was defined. This concept reflects the degree to which a specific class of facility can be chosen by drivers when they are making a trip. The road availability measure cannot be for the entire network, because only a portion of the network is relevant for a specific trip. This relevant portion, defined as *trip proximity*, is the smallest rectangular area being able to cover the trip path (Figure 5). This rectangular area is deemed appropriate for trip proximity based on one of the triangle laws of inequality. The law states that the length of each side is less than the sum of the lengths of the other two sides and greater than the difference between these two lengths. In Figure 5, the chosen path by the driver divides trip proximity into two approximately equal triangles.
The portion of trip represented in each road class is first computed. Total lengths of all link segments along all paths are then computed for each road class within the area of trip proximity. The following two measures are computed from these accumulation counters:

1. Percentage of trip length in each road class
2. Percentage of trip length in each road class normalized by length of road class within the trip proximity

The Python script examines link sequences of each path and classifies links, summing up the link lengths by CFCC. The total length of each roadway class are divided by total path lengths to obtain usages for this class. The comparison between availability and usage of a specific trip can rule out the possible effect of the network composition on path selection, so as to help identify drivers’ preferences among various road classifications.
3. Effect of Network Characteristics on Path Choice

For analytical conveniences, real paths, SDPs, and STPs are categorized in four groups by their length. On real paths the median value of path length is 2.47 km (1.54 miles) and the average length of all paths is 3.51 km (2.18 miles). Thus, it is reasonable to put all paths shorter than 1.61 km (one mile) into one group and paths between 1.61 and 8.05 km (1 and 5 miles) in another. The maximum path length in the dataset is 49.17 km (30.55 miles) and relatively long paths represent only a small percentage of the trips. For this reason, all paths longer than 16.1 km (10 miles) are categorized as one group. Therefore, the data on real paths are presented in four groups:

- a) less than 1.61 km (1 mile);
- b) 1.61-8.05 km (1-5 miles),
- c) 8.05-16.10 km (5-10 miles), and
- d) longer than 16.1 km (10 miles).

Shortest distance and shortest time paths are categorized in the same way.

As illustrated in Figure 6, most trips (3,721 or 65%) have a path length 1.61-8.05 km (1-5 miles). With paths shorter than 1.61 km (1 mile) together, over 90% of trips have a path length less than 8.05 km (5 miles), no matter if the path is real or computed. One of the reasons for relatively fewer longer trips in the dataset is that some of the longer trips were removed in previous steps because they were repetitive commuting trips made by the same driver.
Figure 6: Composition of Path-length Distributions

About 35% (1,997) shortest distance paths and 22% (1266) shortest time paths are found to be identical to their real path counterparts. Figure 7 further illustrates the breakdown of shortest paths that are identical to corresponding real paths in each path length category. The highest identical rates occur on paths with shorter lengths, for both shortest distance and shortest time paths. The rate decreases when the path length becomes longer. The category of “8.05-16.10 km (5-10 miles)” as the lowest identical rate. The identical rate analysis shows that drivers did not necessarily choose the SDPs or the STPs. This finding means that drivers must be influenced by other factors when they choose their real paths.
As shown in Table 2, both real and computed paths have average path lengths comparable to each other. The two computed paths, SDP and STP, have the same average length for paths less than 8.05 km (5 miles). Real paths have average length slightly longer than theoretical paths (SDP and STP), except for the trips between 8.05 and 16.10 km (5 and 10 miles). In this category, STPs have the longest average path length. For path lengths more than 16.10 km (10 miles), SDPs take the longest time. For all trips, STPs take less time than other two types, which is consistent with what’s expected. It is clearly evident from the table that in making a route choice, drivers are willing to spend longer time or travel longer distances that the time and distance via STP or SDP, respectively.

Table 2: Average Path Length and Path Times

<table>
<thead>
<tr>
<th>Path length category</th>
<th>&lt; 1.61 km (1 mile)</th>
<th>1.61-8.05 km (1-5 miles)</th>
<th>8.05-16.10 km (5-10 miles)</th>
<th>&gt; 16.10 km (10 miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average path length in km (miles)</strong></td>
<td>Real paths</td>
<td>0.74 (1.19)</td>
<td>2.20 (3.54)</td>
<td>6.36 (10.24)</td>
</tr>
<tr>
<td></td>
<td>Shortest distance paths</td>
<td>0.69 (1.11)</td>
<td>2.16 (3.48)</td>
<td>6.41 (10.32)</td>
</tr>
<tr>
<td></td>
<td>Shortest Time paths</td>
<td>0.69 (1.11)</td>
<td>2.20 (3.54)</td>
<td>6.50 (10.47)</td>
</tr>
<tr>
<td><strong>Average path time (minutes)</strong></td>
<td>Real paths (GPS times)</td>
<td>7.81</td>
<td>11.13</td>
<td>18.33</td>
</tr>
<tr>
<td></td>
<td>Shortest distance paths</td>
<td>1.33</td>
<td>3.81</td>
<td>10.17</td>
</tr>
<tr>
<td></td>
<td>Shortest time paths</td>
<td>2.61</td>
<td>5.81</td>
<td>11.92</td>
</tr>
</tbody>
</table>
As mentioned earlier, the volunteer trip makers are local to Twin Cities and therefore it is assumed that the trip makers have prior knowledge about the network variables along their chosen and alternative paths. It is further assumed that for each trip the trip-maker’s choice set included only three paths: the actual path chosen, SDP and STP. If the real path is a deviation from SDP or STP, the following network variables along the paths are able to explain the deviation.

- Presence of signals
- Overall number of turning movements
- Number of turning movements at signalized intersections
- Number of turning movements at unsignalized intersections
- Class of roadway links

It is also assumed that left- and right-turns have different impact on path choice and therefore treated separately.

In order to identify the influence of these network variables on path choice, a controlled statistical test is necessary. A paired t-test, which is used to compare two population means via two samples in which observations in one sample can be paired with observations in the other sample, suits this need. A series of paired t-tests are performed, which paired sample means of real paths vs. SDPs and real paths vs. STPs.

### 3.1 Effect of Signals

Figure 8 illustrates that both SDPs and STPs have fewer signals than real paths when the length is shorter than 16.1 km (10 miles). The paired t-test results shown in Table 4 are consistent with this observation. Most of p-values in Table 3 are more than 0.05, which means the null hypothesis that real paths have more signals than shortest paths cannot be rejected. The only two exceptions are in the comparison to SDPs and STPs for trips longer than 16.1 km (10 miles). Considering the small portion of long paths, these p-values could be regarded as the result of outliers. Therefore, it can be concluded that the number of signals along the path is not an influential factor that significantly impacts drivers’ path choice. The test indicates that the opposite is true for trips shorter than 16.1 km (10 miles). That is, drivers embraced more signals along their paths than the number of signals that are present along the SDP or STP.
Figure 8: Number of Signals per 1.61 km (one mile)

Table 3: Effect of Signals on Path Choice

<table>
<thead>
<tr>
<th>Path Category by Length</th>
<th>&lt; 1.61 km (1 mile)</th>
<th>1.61-8.05 km (1-5 miles)</th>
<th>8.05-16.10 km (5-10 miles)</th>
<th>&gt; 16.10 km (10 miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Path</td>
<td>3.27</td>
<td>2.43</td>
<td>1.61</td>
<td>0.22</td>
</tr>
<tr>
<td>Shortest Distance Path</td>
<td>2.92</td>
<td>2.11</td>
<td>1.33</td>
<td>0.22</td>
</tr>
<tr>
<td>Shortest Time Path</td>
<td>2.35</td>
<td>1.23</td>
<td>0.67</td>
<td>0.18</td>
</tr>
</tbody>
</table>

- **Real path**
  - Mean signals per mile along real path ($\mu_r$): 3.304, 2.540, 1.644, 0.245
  - Degrees of freedom: 1,611, 3,721, 304, 78
- **Real path Vs. Shortest distance Path**
  - Mean signals per mile along SDP ($\mu_s$): 3.059, 2.247, 1.471, 0.252
  - Mean difference ($\mu_r - \mu_s$): 0.236, 0.288, 0.111, -0.092
  - t-statistic: 7.390, 16.137, 2.290, -2.423
  - p-value (one-tailed): 1.000, 1.000, 0.989, **0.009**

- **Real path Vs. Shortest time path**
  - Mean signals per mile along STP ($\mu_t$): 2.369, 1.359, 0.591, 0.167
  - Mean difference ($\mu_r - \mu_t$): 0.935, 1.181, 1.053, 0.078
  - t-statistic: 15.815, 40.608, 11.645, 2.037
  - p-value (one-tailed): 1.000, 1.000, 1.000, **0.023**

- $H_0$: ($\mu_r - \mu_s$) $\geq 0$, $H_a$: ($\mu_r - \mu_s$) < 0
- At 95% confidence interval of the difference, the p-value less than 0.05 indicates that $H_0$ may be rejected (or, $H_a$ may be accepted).
- Values in bold indicate that real paths have statistically fewer signals per mile than do shortest paths.
3.2 Effect of Turns Regardless of Signal Status

Figure 9 illustrates the number of turns per 1.61 km (one mile) along paths for real paths, for SDPs, and for STPs. It can be generally ascertained that, the longer the path length, the fewer the number of turns. This pattern is the same for all sets and for both left and right turns.

![Figure 9: Compositions of Turns in Real, Shortest Distance and Shortest Time Paths](image)

It is clearly notable that real paths have lower average turns per 1.61 km (one mile) in all length categories than computed paths. Furthermore, 52.7% of all turns in real paths are right-turns and for trips shorter than 8.05 km (5 miles), the percentage of right-turns goes up to 54% (Table 5). Also, all paths have fewer left turns than right turns on average for trips shorter than 8.05 km (5 miles). These two observations are particularly important because they lead to the following questions:

1. Do real paths tend to have fewer left turns than right turns for shorter trips?
2. For shorter trips, do real paths tend to have fewer overall turns than the corresponding shortest distance path or shortest time path?

Presented in Table 4 are the results of the paired t-test analysis, which addresses the first question. In this analysis, the number of left-turns per 1.61 km (one mile) is paired against the
number of right-turns per 1.61 km (one mile) for each of the nearly 6000 real world trips in the data set.

Table 4: Left-turns vs. Right-turns in Real Paths

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Path length category</th>
<th>Degrees of freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt; 1.61 km (1 mile)</td>
<td>1,611</td>
</tr>
<tr>
<td></td>
<td>1.61-8.05 km (1-5 miles)</td>
<td>3,721</td>
</tr>
<tr>
<td></td>
<td>8.05-16.10 km (5-10 miles)</td>
<td>304</td>
</tr>
<tr>
<td></td>
<td>&gt; 16.10 km (10 miles)</td>
<td>78</td>
</tr>
<tr>
<td>Mean left turns per 1.61 km (one mile) along real path ($\mu_l$)</td>
<td></td>
<td>1.011</td>
</tr>
<tr>
<td>Mean right turns per 1.61 km (one mile) along real path ($\mu_r$)</td>
<td></td>
<td>1.211</td>
</tr>
<tr>
<td>Mean difference ($\mu_r - \mu_l$)</td>
<td></td>
<td>-0.199</td>
</tr>
<tr>
<td>t-statistic</td>
<td></td>
<td>-4.839</td>
</tr>
<tr>
<td>p-value (two-tailed)</td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>

- $H_0$: ($\mu_l$ - $\mu_r$) $\geq$ 0, $H_a$: ($\mu_l$ - $\mu_r$) $<$ 0
- At 95% confidence interval of the difference, the p-value less than 0.05 indicates $H_0$ may be rejected and $H_a$ may be accepted.
- Values in bold indicate that real paths have statistically fewer left turns than right turns.

The analysis in Table 4 indicates that for trip lengths 8.05 km (5 miles) or shorter, routes chosen by drivers will have more right turns than left turns. This observation statistically confirms the intuition that drivers tend to minimize or avoid left-turns. Furthermore, it also validates the case for a vehicle maintenance recommendation that is unrelated to route choice: tires must be rotated at a regular frequency because the majority of the turns drivers make are right turns.

The results of paired t-tests in Table 5 address the second question posed earlier. The t-tests indicated that in all path length categories the null hypothesis that real paths tend to have more left turns may not be rejected. Therefore, the statistical evidence suggests that real paths tend to have fewer overall turns than the corresponding shortest distance path or shortest time path. The same cannot be said about right turns.

Table 5: Effect of Turns on Path Choice

<table>
<thead>
<tr>
<th>Path</th>
<th>Statistic</th>
<th>Path length category</th>
<th>Degrees of freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>All left-turns (signalized or unsignalized)</td>
<td></td>
<td></td>
<td>1,611</td>
</tr>
<tr>
<td>Real path</td>
<td>Degrees of freedom</td>
<td>3,721</td>
<td>304</td>
</tr>
<tr>
<td>Path</td>
<td>Statistic</td>
<td>Path length category</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------------------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean left turns per 1.61 km (one mile) along real path ($\mu_r$)</td>
<td>&lt; 1.61 km (1 mile)</td>
<td>1.61-8.05 km (1-5 miles)</td>
</tr>
<tr>
<td>Real path vs. Shortest distance Path</td>
<td>Mean left turns per 1.61 km (one mile) along SDP ($\mu_s$)</td>
<td>1.182</td>
<td>0.859</td>
</tr>
<tr>
<td></td>
<td>Mean difference ($\mu_r - \mu_s$)</td>
<td>-0.171</td>
<td>-0.181</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>-5.570</td>
<td>-12.818</td>
</tr>
<tr>
<td></td>
<td>p-value (one-tailed)</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Real path vs. Shortest time path</td>
<td>Mean left turns per 1.61 km (one mile) along STP ($\mu_s$)</td>
<td>1.133</td>
<td>0.823</td>
</tr>
<tr>
<td></td>
<td>Mean difference ($\mu_r - \mu_s$)</td>
<td>-0.121</td>
<td>-0.144</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>-2.614</td>
<td>-8.376</td>
</tr>
<tr>
<td></td>
<td>p-value (one-tailed)</td>
<td>0.005</td>
<td>0.000</td>
</tr>
</tbody>
</table>

All right-turns (signalized or unsignalized)

|                                  | Degrees of freedom | 1,611 | 3,721 | 304 | 78 |
| Real path                        | Mean right turns per 1.61 km (one mile) along real path ($\mu_r$) | 1.211 | 0.729 | 0.413 | 0.080 |
| Real path vs. Shortest distance path | Mean right turns per 1.61 km (one mile) along SDP ($\mu_s$) | 1.272 | 0.865 | 0.616 | 0.138 |
|                                  | Mean difference ($\mu_r - \mu_s$)              | -0.066 | -0.136 | -0.219 | -0.067 |
|                                  | t-statistic                                    | -2.095 | -10.802 | -7.877 | -2.839 |
|                                  | p-value (one-tailed)                           | 0.018 | 0.000 | 0.000 | 0.003 |
| Real path vs. Shortest time path | Mean right turns per 1.61 km (one mile) along STP ($\mu_s$) | 1.190 | 0.835 | 0.433 | 0.112 |
|                                  | Mean difference ($\mu_r - \mu_s$)              | 0.021 | -0.106 | -0.020 | -0.033 |
|                                  | t-statistic                                    | 0.442 | -6.910 | -0.330 | -2.909 |
|                                  | p-value (one-tailed)                           | 0.671 | 0.000 | 0.371 | 0.002 |

- $H_0$: ($\mu_r - \mu_s$) $\geq$ 0, $H_a$: ($\mu_r - \mu_s$) $<$ 0
- At 95% confidence interval of the difference, the p-value less than 0.05 indicates that $H_0$ may be rejected (or, $H_a$ may be accepted).
- Values in bold indicate that real paths have significantly fewer left or right turns per 1.61 km (one mile) than do shortest paths.

### 3.3 Effect of Turns at Signalized Intersections

The results of the analysis of turns at signalized intersections along real paths, SDPs and STPs are shown in Table 6. As can be seen from the table, for all path lengths, mean signalized left turns along real paths are consistently higher than mean signalized left turns along SDPs and STPs. The same observation is true for signalized right turns on routes longer than 1.61 km (one...
mile). Paired t-tests also confirm that there is statistical evidence that drivers prefer a route where turns happen at signalized intersections. This evidence further supports the notion floated on the title of this research paper that drivers embrace signals in making a path choice.

**Table 6: Effect of Signalized Turns on Path Choice**

<table>
<thead>
<tr>
<th>Path</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Path length category</td>
</tr>
<tr>
<td></td>
<td>&lt; 1.61 km (1 mile)</td>
</tr>
<tr>
<td>Signalized left turns</td>
<td>Degrees of freedom</td>
</tr>
<tr>
<td>Real path</td>
<td>Mean signalized left turns per 1.61 km (one mile) along real path (µ&lt;sub&gt;r&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Real path vs. Shortest distance Path</td>
<td>Mean signalized left turns per 1.61 km (one mile) along SDP (µ&lt;sub&gt;s&lt;/sub&gt;)</td>
</tr>
<tr>
<td></td>
<td>Mean difference (µ&lt;sub&gt;r&lt;/sub&gt; - µ&lt;sub&gt;s&lt;/sub&gt;)</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
</tr>
<tr>
<td></td>
<td>p-value (one-tailed)</td>
</tr>
<tr>
<td>Signalized right turns</td>
<td>Degrees of freedom</td>
</tr>
<tr>
<td>Real path</td>
<td>Mean signalized right turns per 1.61 km (one mile) along real path (µ&lt;sub&gt;r&lt;/sub&gt;)</td>
</tr>
<tr>
<td>Real path vs. Shortest distance Path</td>
<td>Mean signalized right turns per 1.61 km (one mile) along STP (µ&lt;sub&gt;s&lt;/sub&gt;)</td>
</tr>
<tr>
<td></td>
<td>Mean difference (µ&lt;sub&gt;r&lt;/sub&gt; - µ&lt;sub&gt;s&lt;/sub&gt;)</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
</tr>
<tr>
<td></td>
<td>p-value (one-tailed)</td>
</tr>
</tbody>
</table>

- H<sub>0</sub>: (µ<sub>r</sub> - µ<sub>s</sub>) ≥ 0, H<sub>a</sub>: (µ<sub>r</sub> - µ<sub>s</sub>) < 0
- At 95% confidence interval of the difference, the p-value less than 0.05 indicates that H<sub>0</sub> may be rejected (or, H<sub>a</sub> may be accepted).
- Values in bold indicate that real paths have significantly fewer signalized left or right turns per 1.61 km (one mile) than do shortest paths.
3.4 Effect of Turns at Signalized Intersections

When normalized number of unsignalized intersections along RP, STP and SDP are compared in each distance category, real paths are observed to have fewer unsignalized right turns than the theoretical paths (Table 7). Average number of turns per 1.61 km (one mile) at unsignalized intersections is significantly more than that of STPs for path lengths longer than 8.05 km (5 miles).

Table 7: Effect of Unsignalized Turns on Path Choice

<table>
<thead>
<tr>
<th>Path</th>
<th>Statistic</th>
<th>Degrees of freedom</th>
<th>&lt; 1.61 km (1 mile)</th>
<th>1.61-8.05 km (1-5 miles)</th>
<th>8.05-16.10 km (5-10 miles)</th>
<th>&gt; 16.10 km (10 miles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsignalized left turns</td>
<td>Real path</td>
<td>Mean Unsignalized left turns per 1.61 km (one mile) along real path ($\mu_r$)</td>
<td>0.628</td>
<td>0.428</td>
<td>0.313</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>Real path vs. Shortest distance Path</td>
<td>Mean unsignalized left turns per 1.61 km (one mile) along SDP ($\mu_s$)</td>
<td>0.897</td>
<td>0.663</td>
<td>0.477</td>
<td>0.141</td>
</tr>
<tr>
<td></td>
<td>Mean difference ($\mu_r - \mu_s$)</td>
<td>-0.262</td>
<td>-0.234</td>
<td>-0.178</td>
<td>-0.065</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>-8.757</td>
<td>-17.707</td>
<td>-5.669</td>
<td>-2.960</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value (one-tailed)</td>
<td><strong>0.000</strong></td>
<td><strong>0.000</strong></td>
<td><strong>0.000</strong></td>
<td><strong>0.002</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Real path vs. Shortest time path</td>
<td>Mean Unsignalized left turns per 1.61 km (one mile) along STP ($\mu_s$)</td>
<td>0.918</td>
<td>0.652</td>
<td>0.342</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>Mean difference ($\mu_r - \mu_s$)</td>
<td>-0.289</td>
<td>-0.224</td>
<td>-0.029</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>-9.323</td>
<td>-17.049</td>
<td>-1.122</td>
<td>0.329</td>
<td></td>
</tr>
<tr>
<td></td>
<td>p-value (one-tailed)</td>
<td><strong>0.000</strong></td>
<td><strong>0.000</strong></td>
<td><strong>0.131</strong></td>
<td><strong>0.628</strong></td>
<td></td>
</tr>
</tbody>
</table>

| Unsignalized right turns | Real path | Mean Unsignalized right turns per 1.61 km (one mile) along real path ($\mu_r$) | 0.784 | 0.454 | 0.261 | 0.058 |
| | Real path vs. Shortest distance Path | Mean unsignalized right turns per 1.61 km (one mile) along SDP ($\mu_s$) | 0.983 | 0.667 | 0.509 | 0.122 |
| | Mean difference ($\mu_r - \mu_s$) | -0.203 | -0.210 | -0.256 | -0.075 |
| | t-statistic | -6.81 | -16.465 | -3.581 | -3.396 |
| | p-value (one-tailed) | **0.000** | **0.000** | **0.000** | **0.000** |
| | Real path vs. Shortest time path | Mean unsignalized right turns per 1.61 km (one mile) along STP ($\mu_s$) | 0.955 | 0.665 | 0.386 | 0.090 |
| | Mean difference ($\mu_r - \mu_s$) | -0.171 | -0.211 | -0.125 | -0.032 |
| | t-statistic | -5.109 | -17.608 | -6.059 | -3.105 |
| | p-value (one-tailed) | **0.000** | **0.000** | **0.000** | **0.001** |

- $H_0$: ($\mu_r - \mu_s$) $\geq$ 0, $H_a$: ($\mu_r - \mu_s$) $<$ 0
- At 95% confidence interval of the difference, the p-value less than 0.05 indicates that $H_0$ may be rejected (or, $H_a$ may be accepted).
- Values in bold indicate that real paths have significantly fewer unsignalized left or right turns per 1.61 km.
Similar to the findings for all turns regardless of signal status, unsignalized turns per 1.61 km (one mile) along real paths are fewer than unsignalized turns per 1.61 km (one mile) along STP and SDP intersections when path length is shorter than 8.05 km (5 miles). A reasonably simple way to summarize this analysis is that there is statistical evidence to indicate that, in choosing paths for trips shorter than 8.05 km (5 miles), drivers tend to minimize turns at unsignalized intersections.

### 3.5 Effect of Road Classification

Figure 10 compares road availability to road usage of real paths for four path lengths and three major road classes. Local roads within proximity dominate for all length groups. Although the combined availability of primary and secondary roads is only 10% or less, these two higher classes of roads account for a disproportionally larger portion of the road usage. Regardless of the network composition, drivers were willing (not forced) to choose roads with a higher level of functional class.
As would be expected, as path lengths increase, the portion of the primary road in a real path becomes larger, and the portion of the local road becomes smaller. The percentage of secondary roads in real paths first increases with path length and then decreases. For paths shorter than 16.1 km (10 miles), over 50% usage belongs to local road, and primary road only takes the smallest portion. Only when trip length is longer than 16.1 km (10 miles), does the percent share of local roads become lower than the other two classes. The percent share of primary roads is the highest for trips longer than 16.1 km (10 miles).

Figure 10: Availability vs. Usage of Various Road Classes in Real Paths

As would be expected, as path lengths increase, the portion of the primary road in a real path becomes larger, and the portion of the local road becomes smaller. The percentage of secondary roads in real paths first increases with path length and then decreases. For paths shorter than 16.1 km (10 miles), over 50% usage belongs to local road, and primary road only takes the smallest portion. Only when trip length is longer than 16.1 km (10 miles), does the percent share of local roads become lower than the other two classes. The percent share of primary roads is the highest for trips longer than 16.1 km (10 miles).
3.6 Primary and Secondary Road Usage

This section further explores commonalities and differences between real paths and shortest paths in terms of usage of primary and secondary roads. In each road class, SDPs and STPs trend the same way as real paths when path lengths increase (Figure 11). In other words, usage of primary roads increases while that of local roads decreases. The highest usage of secondary roads happens to paths between 8.05 and 16.1 km (5 and 10 miles). However, for trips longer than 16.1 km (10 miles), the smallest percentage is secondary road for SDP. For trips between 8.05 and 16.1 km (5 and 10 miles), the largest percentage is secondary road for STP.

Figure 11: Road Class Distribution in SDP and STP
Primary, secondary, and local road usage statistics on real paths are compared to that of STP and SDP. On average, STPs occurred more along primary and secondary roads than did the other two sets and by extension, fewer STPs occurred along local roads for all trips. This is reasonable because primary and secondary roads offer higher average traveling speeds and less travel time accordingly. Furthermore, for the shortest time path, the percentage of local roads declined more sharply than for the other two types of path with path length increasing. Meanwhile, SDPs used fewer primary roads than real paths in all categories.

Paired sample t-tests show that real paths have higher percentages of primary and secondary roads than shortest distance paths in almost all circumstances (Table 8). Only for paths shorter than one mile do real paths have less usage of secondary roads than shortest distance paths, where the p-value (0.123) is more than the level of significance (0.05). Proportion of primary roads in real paths is higher than the proportion in SDPs or STPs for the same OD pairs.

### Table 8: Effect of Primary Road on Path Choice

<table>
<thead>
<tr>
<th>Path</th>
<th>Statistic</th>
<th>Degrees of freedom</th>
<th>Path length category</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>&lt; 1.61 km (1 mile)</td>
</tr>
<tr>
<td>Usage of primary roads in paths</td>
<td>Degrees of freedom</td>
<td>1,611</td>
<td>3,721</td>
</tr>
<tr>
<td>Real path</td>
<td>Mean percentage of primary road along real path ((\mu_r))</td>
<td>0.011</td>
<td>0.078</td>
</tr>
<tr>
<td>Real path vs. Shortest distance path</td>
<td>Mean percentage of primary roads along SDP ((\mu_s))</td>
<td>0.013</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>Mean difference ((\mu_r - \mu_s))</td>
<td>0.0080</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>3.667</td>
<td>3.871</td>
</tr>
<tr>
<td></td>
<td>p-value (one-tailed)</td>
<td>(&lt; 0.001)</td>
<td>(&lt; 0.001)</td>
</tr>
<tr>
<td>Usage of secondary roads in paths</td>
<td>Degrees of freedom</td>
<td>1,610</td>
<td>3,721</td>
</tr>
<tr>
<td>Real path</td>
<td>Mean percentage of secondary road along real path ((\mu_r))</td>
<td>0.030</td>
<td>0.121</td>
</tr>
<tr>
<td>Real path vs. Shortest distance path</td>
<td>Mean percentage of secondary roads along SDP ((\mu_s))</td>
<td>0.038</td>
<td>0.131</td>
</tr>
<tr>
<td></td>
<td>Mean difference ((\mu_r - \mu_s))</td>
<td>0.002</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>t-statistic</td>
<td>1.162</td>
<td>3.352</td>
</tr>
<tr>
<td></td>
<td>p-value (one-tailed)</td>
<td>(&lt; 0.001)</td>
<td>(&lt; 0.001)</td>
</tr>
<tr>
<td>Real path vs. Shortest time path</td>
<td>Mean percentage of secondary road along STP ((\mu_s))</td>
<td>0.042</td>
<td>0.116</td>
</tr>
<tr>
<td>time path</td>
<td>Mean difference (µ_r - µ_s)</td>
<td>t-statistic</td>
<td>p-value(one-tailed)</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------</td>
<td>-------------</td>
<td>--------------------</td>
</tr>
<tr>
<td></td>
<td>-0.012</td>
<td>-4.180</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>1.593</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td>-0.003</td>
<td>-0.215</td>
<td>0.415</td>
</tr>
<tr>
<td></td>
<td>0.020</td>
<td>1.438</td>
<td>0.923</td>
</tr>
</tbody>
</table>

- H0: µ_r - µ_s ≤ 0, H1: µ_r - µ_s > 0
- At 95% confidence interval of the difference, the p-value less than 0.05 indicates it’s safe to reject H0 and H1 may be accepted.
- Values in bold indicate that real paths have significantly higher percentage of primary or secondary roads than the shortest paths.

Paired t-tests also indicate that real paths gravitate more towards the higher usage of both primary and secondary roads than theoretical paths computed using shortest time. On the other hand, there is no significant difference between real paths and SDPs with regards to usage of roads of higher classes.

However, real paths did not have more primary and secondary roads than shortest time paths. This confirms intuition that people prefer major roads because they have higher posted speed limits and fewer interruptions, which leads to less traveling time. For this reason only there is no significant difference between real paths and shortest time paths regarding the distribution of road classifications.

For shorter trips the majority of road usage is in the local road category. Local roads have more intersections, and consequently may generate more turning movements. However, there is no obvious pattern for trips of more than 8.05 km (5 miles) because on these trips the influence of turns declined due to the increasing percentage of primary and secondary roads.

4. Conclusions
This research examined the influence of street network and turn variables on drivers’ path choices in a major metropolitan area, Minneapolis-St. Paul, Minnesota. Controlling for these variables, real world path trajectories are compared to the computed shortest distance and shortest time paths. The analysis indicated that the drivers are willing to spend longer time or travel longer distances on paths that have fewer turning movements. The study confirms the intuition that drivers tend to avoid turning movements during their travel. There is statistical evidence to indicate that real paths have fewer turns per 1.61 km (one mile) than both shortest time and shortest distance paths. When they must make a left-turn or right-turn to complete their trips, drivers seem more prone to making the turn at a signal controlled intersection, while at the same time trying to minimize the number of turns occurring at unsignalized intersections. Most
notably, for trips shorter than 8.05 km (5 miles) in length, real paths have a statistically significant fewer left-turns per 1.61 km (one mile) than right-turns per 1.61 km (one mile). This leads to the conclusion that drivers tend to minimize left turns while selecting a path.

The research also presented statistical evidence that the number of traffic signals along the chosen route vs. its theoretical counterparts is not a significant factor in path choice processing. Statistical analyses also revealed that, in terms of the number of signals per mile, real paths contain more signals than their theoretical paths. Prior to this effort, most studies relating path choice behavior to network and path characteristics were based on stated preference surveys. This research used a large dataset of paths with trajectories tracked by GPS to identify the impacts of certain network and path characteristics on drivers’ route-choice. Compared to stated preference surveys, the GPS tracking data are a better representation of how drivers choose route in actual practice.

Study findings confirm, with statistical evidence, some intuitive and some not so intuitive expectations and hypotheses. The results of this study may be generalized with caution for route choice behavior of drivers who are familiar with the network characteristics of street network, signals and traffic patterns. The methodologies used in this research will make it easier to find paths more consistent with drivers’ real choices and consequently provide more sound and solid solutions to traffic assignment problems and other problems in transportation planning. Therefore, the main contribution from this study may be seen as the methodologies that are developed to addressing the research questions, rather than the end results. By applying these methods to a wide range of vehicle trajectory data sets currently available, thanks to the prevalence of smart phones and advanced vehicle technologies, the effect of network variables on path choice behavior can be clearly understood. Similar studies done for different urban areas will lead to developing solutions to network management, transportation planning problems at a local jurisdictional level. A collective assessment of similar studies using different data sets will lead to developing philosophical underpinnings of groundbreaking theories similar to Wardrop Criteria, which led to the development of user equilibrium assignment.

While the trends observed from the data used in this research may be applicable to other population groups and geographical areas comparable to the study area, care must be taken in generalizing the results or using the numerical values in traffic impact and transportation planning studies. Certain observations made in this study may be applicable only to the trip data
collected from 44 volunteers in St. Paul-Minneapolis, Minnesota. More real-world observation data from other metropolitan areas and rural areas would provide more statistically significant results and findings.

It is not clear if any of the path-choices in the data set are influenced by real time travel time information received prior to or during the trip. However, the control variables of this study are network and turn characteristics, which are independent of travel time information. Therefore, the findings of this study are independent of the actions taken by the drivers based on any real time traveler information. As with any study, expansion of the data available for analysis would improve the significance and robustness of the results of this research.
5. References
Li H. “ Investigating Morning Commute Route Choice Behavior Using Global Positioning Systems and Multi-day Travel Data”. Doctoral Dissertation, 2004
Polydoropoulou A, Ben-Akiva M, And Kaysi I (1994). “Influence of Traffic Information on Drivers’ Route Choice Behavior” Transportation Research Record; 1453:p56-65


Appendix 1: Interface of the Smartphone App, GMU Parking Helper that was developed for longitudinal travel trajectory data collection. The density of location data collected were restricted by iOS system and can range to about 100 meters apart.

Figure 1-a Main Interface Interface  
Figure 1-b Navigation Interface  
Figure 1-c Reporting