Modelling of the tactical path selection of bicyclists at signalized intersections

Sasan Amini, M.Sc., Graduate Student (Corresponding Author)
Chair of Traffic Engineering and Control
Technische Universität München, Arcisstraße 21, 80333 Munich
Tel: +49-157-51369662, Email: sasan.amini@tum.de

Heather Twaddle, M.Sc., Graduate Research Associate
Chair of Traffic Engineering and Control
Technische Universität München, Arcisstraße 21, 80333 Munich
Tel: +49-89-28922436, Email: heather.twaddle@tum.de

Axel Leonhardt, Ph.D., Professor
Transportation Planning and Traffic Engineering
Erfurt University of Applied Science, Altonaer Straße 25, 99085 Erfurt
Tel: +49-361-6700572, E-Mail: axel.leonhardt@fh-erfurt.de

Word Count: 4923 Words (excluding the references) + 3 Figures + 5 Tables = 6923 Words

Submitted for presentation and publication to the Transportation Research Board 95th Annual Meeting

Submission Date: July 31th 2015

ABSTRACT

Bicycling is becoming more and more prevalent due to its societal and personal benefits. Consequently, understanding bicyclists' behavior and considering bicyclists as relevant elements in transport and traffic modelling is essential. To assess operational aspects, bicyclists' behavior at intersections is particularly important, as intersections have a large impact on overall system performance and safety. In contrast to motorized vehicles, bicyclists typically have multiple (legal and illegal) path options to travel through an intersection. This study presents a discrete choice model to predict the path on which left-turning bicyclists travel through signalized intersections. To accomplish this objective, revealed preference data from busy intersections in Munich, Germany, has been collected through video observations. The exhibited left-turning maneuvers are categorized in three types: bicycle turn, pedestrian turn and vehicular turn. After a careful analysis of the initial set of explanatory variables, unnecessary variables are omitted from the model. For the data analysis, a multinomial logit model is developed in order to identify the influence of the individual factors. A field effect variable is examined, which reflects the influence of the choice of the peer decision-makers. The results of the study reveal that among the selected variables, seconds passed since the beginning of the red phase of the signal is the most influential parameter followed by the approaching speed of the bicyclist. Ultimately, an external validation was performed with an independent dataset from the same intersection, and the result shows 86% accuracy in the model prediction.

Keywords: Bicyclists' Behavior, Tactical Path Selection, Multinomial Logit Model, Revealed Preference

INTRODUCTION

The personal and societal benefits of bicycling has captured the interest of professionals and policy makers to facilitate the bicycle commuting trips by improving the safety and convenience of bicycling (*1*, *2*). As a result, bicycling is becoming increasingly prevalent for daily commuting trips, which has led to a heterogeneous traffic stream composition in urban areas. Microscopic traffic simulation is a widely used instrument in evaluation of the transportation and traffic control measures before their implementation. However, the reliability of these evaluations strongly depend on the realistic modelling of the road users' attributes and dynamics (*3*).

The existing literature on the simulation models for bicycle traffic, in contrast to motor vehicles, is scarce. Previous research in the field of bicycle transportation has mainly focused on the influential factors on the bicyclists' route choice. For instance, (*4*) have proposed a route choice model with revealed preference GPS data. Many studies have conducted a stated preference survey, which asks bicyclists to rank their preferences for different facility types (*5*–*7*). Bicyclists are difficult to capture in conventional models since they often share the rights of way with motor vehicles but their behavior is quite different due to their different physical and dynamic characteristics. Bicycles are narrower and have a greater lateral flexibility letting them to utilize the lateral space within a traffic lane and switch easily between different types of available infrastructure (*8*). Because of these challenges and the complexity, the majority of the available traffic simulation software on the market still lack a realistic modelling of bicyclists' behavior, particularly the interaction of the bicyclists with other road users. Even if there is the possibility to include bicycles in the simulation, they are modeled through a simplistic approach assuming the bicycles are smaller and low-power vehicles or fast moving pedestrians (*9*).

Nevertheless, a number of recent studies have proposed models, which take bicycles into consideration as a separate mode of travel. For instance, as one of the first attempt, Faghri and Egyhaziova (*10*) developed a computer simulation model called BICSIM (BICycle SIMulator), which is applicable to car-bicycle, bicycle-car and bicycle-bicycle following. The higher degree of lateral flexibility of bicycles was first studied by Oketch (*11*), who investigated the idea of simultaneous utilization of two lanes and gradual lane changing, as opposed to an instantaneous one. These efforts have tried to integrate non-motorized vehicles into the simulation models, especially the short term decision-making. However, these models are not properly calibrated due to lack of empirical data.

In most transportation studies three levels of human behavior are distinguished: strategic, tactical and operational level (*12*–*14*). In this hierarchy, expected utilities at a lower level affect choices at a higher level and choices at a higher level govern behavior at a lower level. The difference between the tactical level of behavior and the operational and strategic level is rooted in the complexity of the goal and duration of the activity. The tactical level of behavior modelling focuses on human behavior when seeking short-term goals in a time scale of seconds to minutes, while operational behavior models assume singular goals that are achieved at a time frame of one second; whereas the strategic behaviors have more complicated goals and involve decision making at more than one level (*15*). For instance, in case of a bicyclist, at the operational level, one would model the obstacle avoidance ability; at the tactical level the decisions to stop or cross a red light would be modeled, and the strategic level of behavior would focus on route choices.

The focus of this research is on bicyclists' path choice behavior at the tactical level. Previous research on the tactical behavior of bicyclists has focused on topics such as red light compliance, infrastructural preferences and gap acceptance. Developing logit models based on revealed preference or stated preference surveys is the most common approach deployed by the researchers in order to find the most influential factors on these decisions. The key findings and the methodology employed in some of these studies is discussed below and is used as the basis for selecting influential factors for this research.

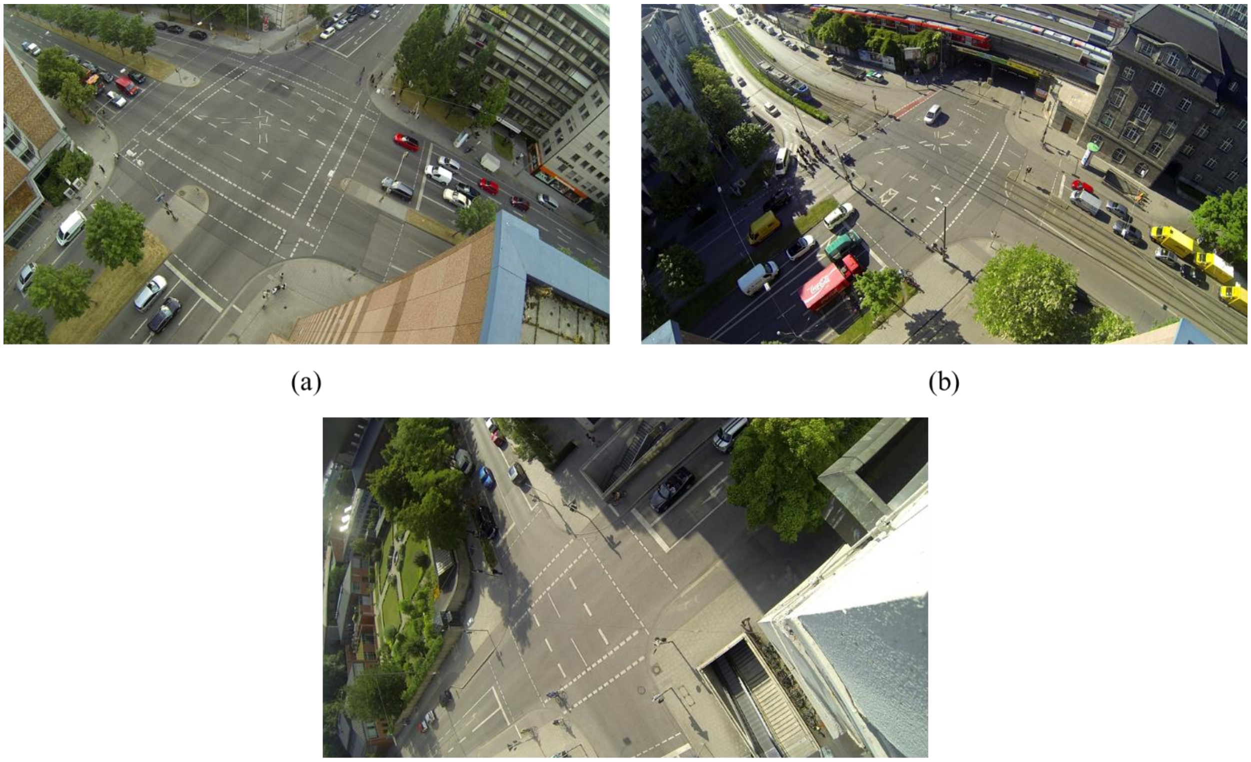
Several studies have been conducted to investigate the behavior characteristics and associated factors of red light violation. The results of these studies reveal that age group, gender (*16*, *17*) and helmet use (*18*) as well as the duration of the red phase and the geometry of the intersection (*19*) have significant influence on red light violation behavior. Regarding the infrastructure selection of bicyclists, the purpose of travel, riding skills and gender are found as the most important factors (*20*, *21*). In another study, (*22*) investigated the influence of weather condition, time of day and segment characteristics in addition to socio-demographic variables on the probability of riding in bicycle lane. To the knowledge of authors, there is no previous research on the path selection at signalized intersections; however, some studies on crossing behavior of bicyclists at unsignalized intersections provide valuable insight of the problem. For instance, Huang and Wu (*23*) developed a fuzzy logic model to describe a bicyclist's path planning in mixed-traffic flow at an unsignalaized intersection in China. The results of this study shows that bicyclists first try to gain rough information of intersection situation and then sketch their preferred path; comfort, directness and efficiency are major criteria for path sketch.

METHODOLOGY

The methodology employed in this study is used to investigate bicyclists' decision making process while approaching a signalized intersection and the translation of these parameters into an algorithm that is suitable for integration in a microscopic simulation tool. Thus, a revealed preference dataset from more than 18 hours of recorded videos from three intersections in Munich, Germany, has been created. A portion of these videos are only used to set the analysis framework and draw reasonable assumptions. This preliminary analysis is performed on five hours of video data from all three intersections, which are different in terms of geometric design and traffic volumes (FIGURE 1). This is an essential step in order to understand the choice situation. Then, the primary dataset is collected from 9 hours of video data from one of the intersections (intersection (b) in FIGURE 1) to estimate the model. The high number of bicyclists, clear overview of the area as well as availability of accurate traffic signal data are the decisive factors to select this intersection as the study site. In the second stage of data collection, six hours of video of another day from the same intersection is collected, which is called the secondary dataset, and is used to validate the model.

Analysis Framework

Path selection problem arises when more than one path exists to reach the destination. In general, there are many physically possible paths to turn left at an intersection, some of which are depicted in FIGURE 2. Note that for the sake of simplicity, some paths from/to vehicle lane or sidewalk are not shown in the figure. This choice set will grow even further if the type of the behavior, for example walking on the pedestrian crosswalk instead of riding or crossing against the red light, is also considered.



(a) (b)

(c)

FIGURE 1 Observed intersections in preliminary analysis phase (a) Marsstrasse-Seidlstrasse (b) Arnulfstrasse-Seidlstrasse (c) Karlstrasse-Luisenstrasse.

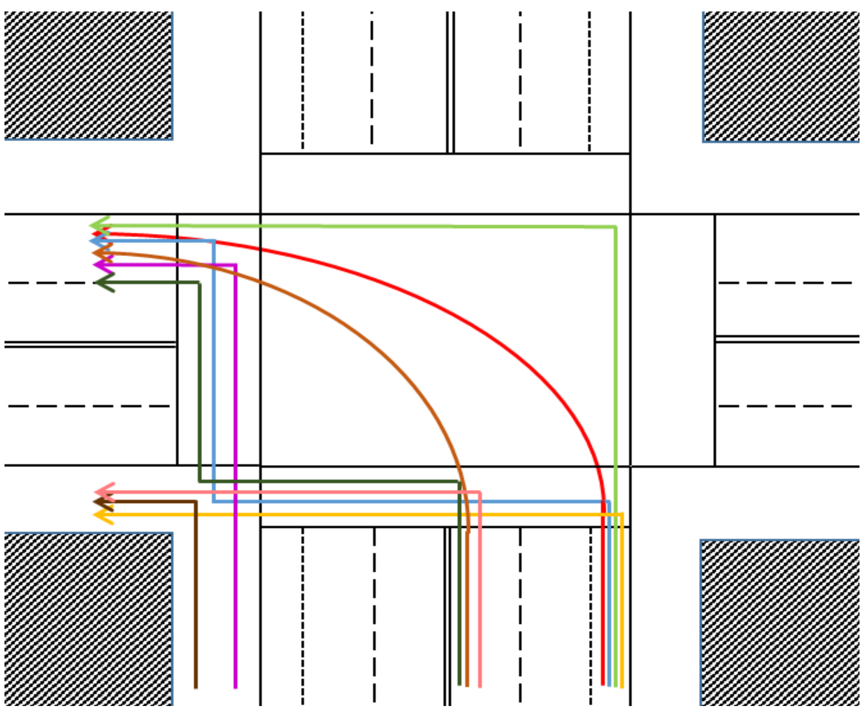


FIGURE 2 Possible paths for turning left at intersections.

The target group of this study is left-turning bicycles regardless of their lateral position on the road (bicycle lane, vehicle lane or sidewalk) and including those that exhibit law-breaking behaviors. The analysis framework is set based on the results of the preliminary observation on left-turning bicyclists. The main conclusions made at this step are:

- Bicyclists who arrive during the green phase of the straight-through signal follow the same path. If they are in bicycle lane, first they go straight and then wait for the green light of the second signal, and if they are in the left most vehicle lane, they turn diagonally. This implies that two groups of bicyclist are confronted with two different sets of choices and have to be studied separately.
- Bicyclists riding on the sidewalk as well as the red light violators are rarely observed and can therefore be neglected.
- Bicyclists change their approaching lane and speed in accordance to their preferred path around 30 meters upstream the stop line. Those who want to stop at the straight-through signal smoothly reduce their speed, but other riders that want to cross the pedestrian crosswalk are less patient and ride faster to avoid two stops at the intersection. In order to address this issue in this research, left-turning path selection is defined as a sequence of actions; as the bicyclist gets closer to the stop line, starts to analyze the intersection, in particular, the traffic signals status, speed and the position of other road users.
- Bicyclists assess all the available paths continuously for a short period of time and make their decision a few meters (three meters in this case) prior to the stop line. This is where they are able to observe the pedestrian signal status and have enough space to adjust their speed and lateral position in line with their decision. An "observation zone" on each segment of the intersection has been defined that is similar to the concept of "dilemma zone" which has been used for analyzing drivers' reaction to signal change at intersections (*24*). Since there is not a clear view of 30 meters on all segments of the intersections, the beginning of the observation zone is considered to be 20 meters upstream of the stop line; bicyclists' speed would be recorded onset of entering the observation zone and they choose their desired path as they reach the decision making point. In case this distance is not covered by the camera, the farthest observable distance will be used.

Data Collection

To study bicyclists' behavior, a collection of video data is gathered by mounting a high-definition camera in 25fps format at the top of a building that has a full view of the intersection. As the automated extraction of variables describing the tactical behavior is very difficult, the video frames are manually analyzed. The intersection of two important arterial roads, Arnulfstrasse and Seidlstrasse, in the Central Business District (CBD) of Munich close to the Central Train Station (Hauptbahnhof) create a four leg-intersection as shown in FIGURE 1b. It is a large, high-volume intersection with a fully actuated traffic signal control. Three approaches of the intersection have two vehicle lanes and a dedicated bicycle lane. The east approach does not have a bicycle lane. A tram line running through the middle of Arnulfstrasse is prioritized over the private car traffic. The duration of the cycle time varies between 70 to 90 seconds. The south approach leads to the intersection through an underpass that limits the camera view range. This causes limitations on speed measurements on this approach. For this reason a binary variable that indicates the relative speed in comparison with other bicyclists is used in place of a measures speed value.

Modelling Approach

The selection of a path from a set of possible paths is a special case of discrete choice problem. The mathematical method which is by far the most commonly used approach to represent the characteristics of discrete choice problem is the logit model. Therefore, Multinomial Logit (MNL) model is deployed, which has shown to be effective in traffic simulation tools as well (*25*).

The first and foremost step in the estimation process of the logit model is to identify the choice set. As shown in FIGURE 2, there are an extremely large number of finite paths that can be selected by a bicyclist. In order to simplify the choice situation, left-turning maneuvers are categorized in three types:

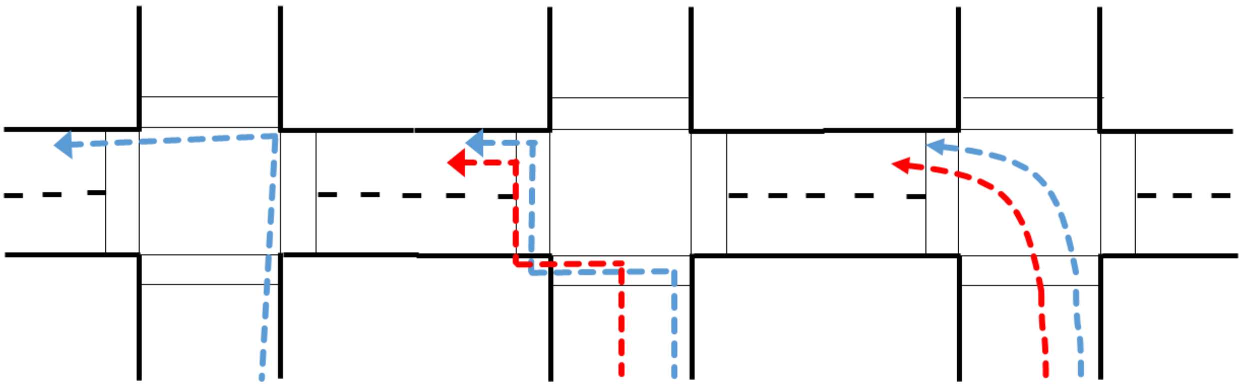


FIGURE 3 Categorization of possible alternatives for turning left at an intersection; blue paths are for bicycle lane group and red ones for the vehicle lane group of riders.

The fundamental difference between two groups of bicyclists, bicycle lane and vehicle lane group, implies that a standard MNL model is not applicable for this problem, and two independent MNL models are needed. This is similar to the approach followed by Bierlaire et al. (*26*) to combine the results of two revealed preference and stated preference surveys. The idea is to formulate two independent MNL models with the same explanatory variables.

After identifying the choice set, the utility functions that are assigned to the alternatives should be formulated. In general, a linear in parameter function is used (*27*):

$$U_{in} = V_{in} + \varepsilon_{in} = \sum_{k} \beta_k X_{ink} + \varepsilon_{in} \tag{1}$$

Where V_{in} is the deterministic portion of the utility, which is defined by β, a vector of k parameters influencing the utility and ε_{in} is the error term. The crucial limitation of this model is that only the observable attributes of the alternatives and characteristics of the decision maker are considered. Many psychologists and economists, however, have specifically emphasized the influence and constraints imposed by the peer group to which the decision maker belongs (*28*). One solution is to integrate a field effect variable into the utility function that captures the average effect of peer group choice on the probability of an alternative (*29*). Thus, the utility function is formulated as:

$$U_{in} = V(X_{in}, S_n; \beta) + \gamma F_{in} + \varepsilon_{in} \tag{2}$$

Where F_{in} is the proportion of the people in the peer group of agent *n* who chose alternative *i* and γ is the coefficient that reflects the strength of the influence. This choice model includes the social influence and can be estimated by conventional methods. However, it may suffer from endogeneity resulting in inconsistent estimate of the model parameters. The endogeneity arises because the effect of other decision makers is captured by both ε_{in} and F_{in}, and therefore, these two variables might be correlated. A correction algorithm known as BLP method has been proposed by Berry, Levinsohn and Pakes (*30*). This approach solves the endogeneity at a market level, which includes a group of similar decision makers that influence each other. The error term is decomposed into two portions: the endogenous-causing part and a random term. Since the endogeneity occurs at market level, the utility function is modified by considering *m* markets. The equation then becomes:

$$U_{in_m} = V(X_{in}, S_{n_m}; \beta) + \dot{\varepsilon}_{in_m} + [\gamma F_{im} + \ddot{\varepsilon}_{im}] \qquad (3)$$

The term $[\gamma F_{in} + \ddot{\varepsilon}_{in}]$ in this equation reflects both the observable and unobservable part of the utility relevant to the decision maker's peer group. This term will be replaced by a market constant α_{im} which will be estimated by using a two-stage linear regression model (readers are referred to (*29*) for more information). The ultimate form of utility function is expressed as:

$$U_{in_m} = V(X_{in}, S_{n_m}; \beta) + \dot{\varepsilon}_{in_m} + \alpha_{im} \qquad (4)$$

It should be noted that if the estimated coefficient for this field effect variable is close to zero or statistically insignificant, then the distribution of the probability of that particular alternative is not in relation with other decision makers' choices. In other words, it is not beneficial to integrate the social influence in the model and consequently, the traditional approach should be used.

Once the model is defined, selecting the influential parameters which must be recorded during data collection phase plays an important role. A list of selected factors is given in the following table. The video data does not provide adequate information for the selection of the personal explanatory variables (i.e. age group, gender, helmet use) due to height of the camera:

TABLE 1 Coded Parameters

Parameter Symbol	**Parameter Definition**
X_{red}	Seconds passed since the beginning of the red phase of the signal Continuous variable
X_{rs}	Pedestrian crosswalk signal status Dichotomous variable: 1 = green and 0 = red
X_s	Bicyclists' approaching speed at the beginning of the observation zone Continuous variable
X_{ph}	Peak-hour Dichotomous variable: 1 = peak-hour and 0 = non-peak

X_{st}	Stop line direct accessibility (bicyclists that are already waiting at the stop line have blocked the direct access to the pedestrian crosswalk for other bicyclists) Dichotomous variable: 1 = blocked and 0 = direct access
X_{cc}	Number of the conflicting cars from the opposing traffic stream Continuous variable
X_q	Number bicyclists waiting at the stop line Continuous variable
X_{ped}	Number of conflicting pedestrians Continuous variable
X_w	Effective width of pedestrian crosswalk (sum of the body width of present pedestrians is deducted from actual width of the pedestrian crosswalk and converted into a ratio value between 0 and 1) Continuous variable

DATA ANSLYSIS

Choice Distribution

In total, 315 left-turning bicyclists arriving during the red phase of the straight-through signal were observed; 261 riding in bicycle lane and 54 riding in the left most vehicle lane. No other paths apart from the ones included in the initial choice set, which were illustrated in FIGURE 3, have been observed; therefore, the universal choice set remains the same. However, among the riders in the bicycle lane group no one selected the vehicular turn, and as a result this path will be omitted from the choice set. The choice distribution for bicyclists riding in bicycle lane is almost the same on all segments of the intersection. It is likely due to the infrastructural similarities on the segments. Almost 40% of bicyclists have chosen bicycle turn and 60% have selected the pedestrian turn. However, in the vehicle lane group, the vehicular turn is more attractive than the pedestrian turn. The descriptive statistics reveals that for the bicycle lane group, 79% of those who ride faster than the average speed of the approach and 67% of peak-hour riders have selected the bicycle turn versus the bicycle turn. These values are 64% and 38% for the vehicle lane group respectively. Below, the choice distribution per segment for both group of riders is summarized:

TABLE 2 Choice Distributions

Approach	**Bicycle Lane**			**Vehicle Lane**		
	Total	Bicycle turn	Pedestrian turn	Total	Pedestrian turn	Vehicular turn
North	43	17	26	18	3	15
East	12	5	7	34	12	22
South	131	50	81	0	0	0
West	75	33	42	2	2	0
Total	261	105	156	54	17	37

Explanatory Variables Analysis

In this section the initial set of explanatory variables is analyzed to discover which variables strongly influence the bicyclists' choice (dependent variable) and to investigate potential correlation between the independent variables. First, the field effect variable is discussed; the value of F_{2n} is the portion of bicyclists selecting path 2 (pedestrian turn) and is 0.61, 0.58, 0.62 and 0.56 for north, east, south and west approach of the intersection respectively. The value of this variable is more or less 0.6 in all segments for the bicycle-lane group, which is a first indication that the field effect variable could be insignificant. For the vehicle-lane group, due to small sample size of each segment, it is not possible to include the field effect variable into the utility functions.

Pearson's correlation coefficient is computed for each pair of independent variables in order to investigate their linear correlation. High values of correlation coefficient (e.g. 0.8 or 0.9) indicates a strong linear correlation, and thus, only one of the two intercorrelated variables should be kept in the model. Developing the correlation matrix, reveals that the queue length (X_q) is strongly correlated to a number of independent variables like X_{red}, X_{cc}, X_{st}. Thus, X_q is removed from the model and X_{st} (the stop line accessibility) is used to reflect the influence of the waiting bicyclists in the model. Furthermore, including X_{cc} in the bicycle-lane model is not necessary as none of the bicyclist selected the vehicular turn.

In addition to the statistical analysis, observational conclusions have been made to add or drop some variables from the utility functions. Conflicts with pedestrians mostly occur on the north crossing, and additionally, in most cases even if there is a huge crowd walking across the crosswalk, bicyclists do not change their path, but try to slightly adjust their trajectory to avoid collision with the crowd. Therefore, X_{ped} and X_w will be discarded from the model.

To conclude the findings of the descriptive statistics, the models which will be estimated in the next section are formulated below:

- Bicycle lane model (including the social influence):

Choice set = {bicycle turn (path 1), pedestrian turn (path 2)}

$$U_{1_m} = \alpha_{1_m} + \beta_{red} X_{red} + \dot{\varepsilon}_{1_m} \qquad (5\text{-}1)$$

$$U_{2_m} = \alpha_{2_m} + \beta_{rs} X_{rs} + \beta_s X_s + \beta_{st} X_{st} + \beta_{ph} X_{ph} + \dot{\varepsilon}_{2_m} \qquad (5\text{-}2)$$

If the field effect variable is statistically insignificant or close to zero, the standard format of the model will be used, which includes the Alternative Specific Constants (ASC) as well:

$$U_1 = ASC_1 + \beta_{red} X_{red} + \varepsilon_1 \qquad (6\text{-}1)$$

$$U_2 = ASC_2 + \beta_{rs} X_{rs} + \beta_s X_s + \beta_{st} X_{st} + \beta_{ph} X_{ph} + \varepsilon_2 \qquad (6\text{-}2)$$

- Vehicle lane model:

Choice set = {pedestrian turn (path 2), vehicular turn (path 3)}

$$U_2 = ASC_2 + \beta_{rs} X_{rs} + \beta_s X_s + \beta_{st} X_{st} + \varepsilon_2 \qquad (7\text{-}1)$$

$$U_3 = ASC_3 + \beta_{red} X_{red} + \beta_{cc} X_{cc} + \varepsilon_3 \qquad (7\text{-}2)$$

It is worth to note that these are not the final form of the functions; the ultimate model will be obtained after conducting the necessary statistical tests, external validation and refining the explanatory variables to reach the best fit model.

Logit Model Estimation

In this section, first a logit model would be estimated using BIOGEME freeware (*31*) and then, the backward approach is employed to remove the insignificant variables from the model. More specifically, variables with lower p-value will be omitted stepwise until all p-values are smaller than the defined threshold. Even though modern models criticize the usefulness of hypothesis testing approaches (*32*), for this specific case study due to limited number of explanatory variables these methods are simple and useful. The outcome of the estimation are assessed by three criteria:

1. Model performance: it is evaluated based on rho-squared and adjusted rho-squared values which is a measure for goodness of fit.
2. Model significance: this measure analyzes the significance of the coefficients. This is done via t-test at 95% confidence interval, which means the value of t-test must be equal or higher than 1.96 to be able to reject the null hypothesis.
3. Model correctness: correctness of the coefficients is the last measure to check if the sign of the coefficients are the same as the expected sign.

The bicycle-lane model was estimated first with consideration of the field effect variable. However, the t-test value for all the market specific constants shows they are insignificant at the 90% interval. Consequently, the field effect variable is excluded and the model is estimated without considering social influence.

At this step, the model is estimated by considering the utility functions as formulated in Equations (6-1) and (6-2). Then, the model is refined by employing the backward elimination procedure, in which the less significant coefficients are omitted one by one to see if the performance of the model decreases considerably. For the bicycle lane model, β_{ph} and β_{st} have the greatest p-value and despite the fact that they both have the expected sign, they should be discarded stepwise (model significance criterion is violated). Ultimately, a likelihood ratio test, which is a useful test to compare a full model with a restricted model, was conducted to select the best model. The following table summarizes the results of the parameter estimation:

TABLE 3 Estimation Results for the Bicycle Lane Model

Coefficient	Value	Standard Error	T-Test	P-Value
ASC_1	0	fixed	-	-
ASC_2	1.45	0.394	3.68	0.000
β_{red}	0.083	0.0116	7.11	0.000
β_{rs}	1.64	0.354	4.61	0.000
β_s	1.3	0.41	3.17	0.000
Number of observations	261			
Null log-likelihood	-180.91			
Final log-likelihood	-104.245			
Likelihood ratio test	153.33			
Rho-squared	0.424			
Adjusted rho-squared	0.402			

From the table it can concluded that β_{red} is the most influential parameter. The positive sign of this factor means that as the end of red phase of the straight-through signal nears, bicyclists are more likely to select the bicycle turn. Moreover, faster bicyclists are expected to choose the pedestrian turn, possibly because they are less patient. The value of β_{rs} implies that if the pedestrian crosswalk signal is green upon arrival, the probability of choosing the pedestrian turn is higher than the bicycle turn.
Regarding the vehicle lane model, as discussed earlier, the sample size is too small to provide enough variation among the selected variables. This issue can be observed better when the parameters are estimated:

TABLE 4 Estimation Results for the Vehicle Lane Model

Coefficient	Value	Standard Error	T-Test	P-Value
ASC_2	-3.94	5.91	-0.67	0.51
ASC_3	0	fixed	-	-
β_{cc}	-0.429	0.198	-2.17	0.03
β_{ph}	0.748	0.99	0.76	0.45
β_{red}	0.103	0.0371	2.77	0.01
βrs	1.43	0.932	1.53	0.13
β_s	0.85	0.995	0.85	0.39
β_{st}	3.94	5.79	0.68	0.5
Number of observations	54			
Null log-likelihood	-37.43			
Final log-likelihood	-17.49			
Likelihood ratio test	39.887			
Rho-squared	0.533			
Adjusted rho-squared	0.346			

All parameters except β_{red} and $\beta_{cc}$are statistically insignificant. Therefore, the estimation procedure of the vehicle lane model is terminated and this model will not be discussed further.

Model Validation

Two types of validation are performed: first a face validity by computing the "adjusted percentage of right" value as presented by (27):

$$\overline{PR} = \frac{100}{N}\sum_{n}\sum_{i} P_n(i)^{y_{in}} \tag{8}$$

Where $P_n(i)$ is the probability of choosing alternative i for person n, and y_{in} is 1 if the highest predicted probability by the model corresponds to the chosen alternative, and 0 otherwise. This adjusted statistic reflects the value of log likelihood function better and is more sensitive to low values of probabilities for the chosen alternative. The bicycle lane model predicts 219 out of 261 correctly which means a $\overline{PR}$ value of 71.1.

Second type of validation is analyzing the accuracy of the model on an independent dataset at both aggregate and disaggregate levels, which is also known as external validation. Six hours of video data from the same intersection but from another day with similar conditions (weather, weekday, traffic, etc.) is collected and analyzed with the same approach. In total 200 bicyclists are recorded, 164 of them rode in bicycle lane. The external validation is performed at both the aggregate and disaggregate level. At aggregate level, the choice distribution of the model is compared with the actual choice distribution of the empirical data. At disaggregate level, the choice distribution is compared for each path separately. More specifically, the accuracy of prediction for each path is checked with individuals who have chosen that specific path. The following table summarizes the results of the external validation:

TABLE 5 External Validation of the Bicycle Lane Model

Path	Empirical Data		Aggregate		Disaggregate	
	Total	Percentage	Number	Percentage	Correct	Percentage
Bicycle turn	62	37.8%	53	32.3%	46	74.2%
Pedestrian turn	102	62.2%	111	67.7%	95	93.1%
Total	164	100%	-	-	141	85.9%

CONCLUSION

This study presented an MNL model to predict the path selection of left-turning bicyclists at signalized intersections. As a case study, video data was collected at a busy intersection in Munich, Germany and data extraction was performed manually to create a revealed preference dataset. Discrete choice analysis was employed to develop the MNL model, which is successful in accurately predicting the choice behavior of the bicyclists riding in bicycle lane. The predicted path selection ratio is only 5.5% different from the empirical data. However, due to the small portion of bicyclists that ride in the vehicle lane, it was not possible to estimate a consistent model. Nevertheless, from the bicycle lane model it can be concluded that X_{red} is the most influential variable. Another key finding of this research indicates that bicyclists who are riding faster than the average of a segment, are more likely to select the pedestrian turn. Moreover, external validation denotes that the model correctly predicts the selected path almost 86% of the times. With further studies, the impact of gender and helmet use as well as infrastructural factors such as signal cycle time and crossing distances can be investigated. The nature of the observation zone that has been considered in this study easily suits the route decision algorithms on the existing traffic simulation tools. The integration of the presented choice model into an existing microscopic traffic flow simulation tool and its deployment will be presented in a future work.

REFERENCES

1. Noland, R. B., and H. Kunreuther. Short-run and long-run policies for increasing bicycle transportation for daily commuter trips. Transport Policy, Vol. 2, No. 1, 1995, pp. 67–79.
2. Dill, J. Bicycling for transportation and health: the role of infrastructure. *Journal of public health policy*, Vol. 30, 2009, pp. S95–S110.

3. Barceló, J. *Fundamentals of Traffic Simulation*. Springer-Verlag, New York, 2010.
4. Broach, J., J. Dill, and J. Gliebe. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, Vol. 46, No. 10, 2012, pp. 1730–1740.
5. Yang, C., and M. Mesbah. Route Choice Behaviour of Cyclists by Stated Preference and Revealed Preference. *Proceedings of 36th Australasian Transport Research Forum (ATRF)*, Brisbane, Queensland, Australia 2013.
6. Krizek, K. J. Two approaches to valuing dome of bicycle facilities' presumed benefits. *Journal of the American Planning Association*, Vol. 72, No. 3, 2006, pp. 309–320.
7. Stinson, M. A., and C. R. Bhat. Commuter Bicyclist Route Choice. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1828, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 107–115.
8. Twaddle, H., T. Schendzielorz, and O. Fakler. Bicycles in Urban Areas: Review of Existing Methods for Modeling Behavior. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2434, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 140–146
9. Kosonen, I. HUTSIM: Simulation tool for traffic signal control planning, Helsinki University of Technology, Transportation Engineering. Publication 89, Finland, 1996.
10. Faghri, A., and E. Egyháziová. Development of a Computer Simulation Model of Mixed Motor Vehicle and Bicycle Traffic on an Urban Road Network. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1674, Transportation Research Board of the National Academies, Washington, D.C., 1999, pp. 86–93.
11. Oketch, T. New Modeling Approach for Mixed-Traffic Streams with Nonmotorized Vehicles. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1705, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 61–69.
12. Doshi, A., and M. M. Trivedi. Tactical driver behavior prediction and intent inference: A review. *Proceedings of IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2011, pp. 1892-1897.
13. Rudigier, M., and M. Horn. Concepts for modeling drivers of vehicles using control theory, In *Data and Mobility*, Springer Berlin Heidelberg, 2010, pp. 27-38.
14. Hoogendoorn, S. P., and P. H. L. Bovy. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*, Vol. 38, No. 2, 2004, pp. 169–190.
15. Sokolowski, John A., and Catherine M. Banks. *Modeling and simulation fundamentals: theoretical underpinnings and practical domains*. John Wiley & Sons, 2010.
16. Huan, M., X. Yang, and B. Jia. Modeling Cyclist Violation Behavior at Signalized Intersection in China. *Proceedings of 5th International Joint Conference on Computational Sciences and Optimization,* 2012, pp. 489-493.
17. Johnson, M., S. Newstead, J. Charlton, and J. Oxley. Riding through red lights: The rate, characteristics and risk factors of non-compliant urban commuter cyclists. *Accident Analysis and Prevention*, Vol. 43, No. 1, 2011, pp. 323–328.
18. Zangenehpour, S., L. F. Miranda-Moreno, and N. Saunier. Impact of Bicycle Boxes on Safety of Cyclists: A Case Study in Montreal. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.

19. Pai, C. W., and R. C. Jou. Cyclists' red-light running behaviors: An examination of risk-taking, opportunistic, and law-obeying behaviors. *Accident Analysis and Prevention*, Vol. 62, 2014, pp. 191–198.
20. Taylor, D. B., and H. S. Mahmassani. *Behavioral models and characteristics of bicycle-automobile mixed-traffic: planning and engineering implications*. Report No. SWUTC/98/60056-1, 1998.
21. Hunt, J. D., and J. E. Abraham. Influences on bicycle use. *Transportation*, Vol. 34, No. 4, 2007, pp. 453–470.
22. Kang, L., and J. D. Fricker. Bicyclist commuters' choice of on-street versus off-street route segments. *Transportation*, Vol. 40, No. 5, 2013, pp. 887–902.
23. Huang, L., and J. Wu. Cyclists' path planning behavioral model at unsignalized mixed traffic intersections in China. *IEEE Intelligent Transportation Systems Magazine*, Vol. 1, No. 2, 2009, pp. 13–19.
24. Gates, T. J., D. A. Noyce, L. Laracuente, and E. V. Nordheim. Analysis of Driver Behavior in Dilemma Zones at Signalized Intersections. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2030, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 29–39.
25. PTV Vissim 7 User Manual, PTV AG, 2014, pp. 509-512
26. Bierlaire, M., K. Axhausen, and G. Abay. The acceptance of modal innovation: The case of Swissmetro. Presented at 1st Swiss Transport Research Conference, Ascona, Switzerland, 2001.
27. Ben-Akiva, M., and S. R. Lerman. *Discrete Choice Analysis: Theory and Applications to Travel Demand*. MIT Press, Cambridge, 1985.
28. Hess, S., and A. John Daly. *Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice Modelling Conference: The State-of-the-art and the State-of-practice: Proceedings from the Inaugural International Choice Modelling Conference*. Emerald Group Publishing, 2010.
29. Walker, J. L., E. Ehlers, I. Banerjee, and E. R. Dugundji. Correcting for endogeneity in behavioral choice models with social influence variables. *Transportation Research Part A: Policy and Practice*, Vol. 45, No. 4, 2011, pp. 362–374.
30. Berry, S., J. Levinsohn, and A. Pakes. Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market. *Journal of Political Economy*, Vol. 112, No. 1, 2004, pp. 68–105.
31. Bierlaire, M. BIOGEME: A free package for the estimation of discrete choice models, *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland, 2003.
32. Lukacs, P. M. et al. Concerns regarding a call for pluralism of information theory and hypothesis testing. *Journal of Applied Ecology*, Vol. 44, No. 2, 2007, pp. 456–460.