Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections

By Heather Twaddle[a] and Fritz Busch[b]

[a,b] Chair of Traffic Engineering and Control
Technical University of Munich
Arcisstrasse 21, Munich, Germany, 80333

Abstract

Bicyclists are extremely flexible road users who employ various tactical behaviours to optimise comfort, directness and time efficiency while crossing a signalised intersection. Tactical choices faced by bicyclists at signalised intersections include whether to use the bicycle lane, roadway or sidewalk, to stop at or violate a red traffic signal, to ride with or against the mandatory direction of travel and the method of executing a left turn. The outcome of these choices has a direct impact on traffic safety and efficiency at intersections. In this paper, revealed choice data from 4710 bicyclists at four intersections in Munich, Germany are used to estimate binomial and multinomial logistic regression models to predict tactical choice outcomes. Optimal predictor sets are selected from the main and two-way interaction effects of 43 independent variables describing the situation, strategic behaviour and prior tactical choices of bicyclists using recursive feature elimination. A simplified model is estimated using the statistically significant variables of the optimal predictor set. The prediction power of the resulting regression model is assessed using k-fold cross validation. The models to predict response to a red signal and the type of left-hand turn exhibit high predictive power while the prediction of infrastructure selection and the direction of travel proves to be difficult.

Keywords: Bicyclist behaviour; Tactical choice modelling; Regression analysis

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

1. Introduction

Bicycling is an inexpensive, non-polluting transportation mode that is often the fastest alternative for trips under 5 km in cities and towns (Dekoster et al., 2000). Benefits can be realised on both the personal and societal level by capitalising on the economic, health and mobility advantages of bicycling. However, persisting concerns with bicyclist safety and the challenge of maintaining traffic efficiency for all modes while encouraging bicycling have underlined the need for quantitative research in the field of bicycle transport. Knowledge concerning the tactical behaviour of bicyclists, factors motivating this behaviour and the relationship between bicyclist behaviour and overall traffic safety and efficiency is necessary to support future transport planning and engineering endeavours. Furthermore, the development of driver assistance systems and Intelligent Transport Systems that aim to protect bicyclists requires detailed knowledge about the behaviour of bicyclists. Models for predicting the behaviour of bicyclists based on previous movements and the current situation are imperative inputs for such systems.

According to the framework defined by Michon (1985), tactical behaviour of road users comprises conscious decisions made on a time horizon of seconds to minutes, such as path selection and response to a red signal. Tactical behaviour is guided by strategic choices, such as route choice, and is constrained by feedback from the operational level, including current speed and necessary evasive actions. Tactical decisions are also influenced by situational factors, such as the geometry of the intersection and the traffic signal control. The sociodemographic characteristics and preferences of the bicyclist influence the interpretation of the situation and determine the response.

To address the lack of definitive knowledge concerning the factors motivating tactical behaviour, four choices faced by bicyclists at signalised intersections are investigated in this paper; the choice between riding on the bicycle lane, roadway or sidewalk (infrastructure selection), the reaction to a red signal, the decision to ride with or against the mandatory direction (direction of travel) and the path selection when executing a left turn manoeuvre. Logistic regression models are estimated to predict the outcome of these choices using strategic choices, prior tactical

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

behaviour and situational variables as predictors. The aim of this research is to identify relationships between the environment and the tactical behaviour of bicyclists in order to design infrastructure and traffic signal control that encourages overall rule-conform and safe behaviour. As such, the personal attributes of the bicyclist, including gender, age and bicycling experience, are not included in the literature review or model development.

2. Literature Review

There is a large body of research that has examined the link between various tactical behaviours of bicyclists and the occurrence of a collision. For example, riding against the mandatory direction of travel has been found in many studies to increase the risk to bicyclists (Alrutz and Meewes, 1980; Gerstenberger, 2015; Herslund and Jørgensen, 2003; Ortlepp, 2009; Räsänen and Summala, 1998; Summala et al., 1996). In addition, a number of studies have investigated the type of infrastructure available for bicyclists and the risk of injury (Aultman-Hall and Hall, 1998; Lusk et al., 2011; Moritz, 1998; Reynolds et al., 2009; Rivara et al., 1997; Rodgers, 1995; Teschke et al., 2012). The tactical use of different parts of the road infrastructure is very likely also associated with risk, although no studies were found that directly examine this link. Other tactical behaviours, including violating red traffic signals and riding on the sidewalk, have been found by researchers to pose significant risk to bicyclists and other road users (Lusk et al., 2011; Moritz, 1998; Reynolds et al., 2009; Rivara et al., 1997; Rodgers, 1995; Schramm et al., 2008; Teschke et al., 2012).

Although the relationship between bicyclists' actions and resulting safety risk is relatively well understood, few studies have investigated the motivating factors of these behaviours. The findings of the few studies identified in the course of this review are summarised below.

- **Infrastructure Selection:** The choice between using a bicycle lane, the roadway or the sidewalk is motivated by the type and width of an available bicycle lane (Alrutz et al., 2009; Guo et al., 2013) as well as the quality and possible obstruction of the facility (Kuller et al., 1986). Researchers have investigated the role of bicycle, pedestrian and motor vehicle traffic volumes in infrastructure selection, but have come to contradictory conclusions (Alrutz et al.,

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

2009; Guo et al., 2013; Kuller et al., 1986). Non-observable motivating factors, such the desire to ride two abreast, preparation for upcoming manoeuvres and a feeling of safety were given by bicyclists who decided not to use an available bicycle lane (Kuller et al., 1986).

- **Response to red signal:** In their review of 16 studies pertaining to red light violations of bicyclists, Richardson & Caulfield (2015) found percentages of red light violation (or violators for self-reported surveys) to range between 6.9% and 87.5%. The infrastructure used by a bicyclist is linked to the likelihood of a red light violation (Allen et al., 2005; Johnson et al., 2011; Richardson and Caulfield, 2015). In addition, the traffic flow on the current approach as well as on the crossing road of the intersection plays an important role in red light compliance (Johnson et al., 2013, 2011; Pai and Jou, 2014). The desired manoeuvre of the bicyclist (straight, right or left turn) has been found to be related to red light violation (Johnson et al., 2013, 2011), as is the length of the signal phase (Pai & Jou, 2014).
- **Direction of travel:** Two main reasons for riding against the given direction of travel were identified by Kuller et al. (1986). First, bicyclists are more likely to violate traffic rules upon approaching their final destination or an intermediate goal and consequently ride the last few meters of the trip against the mandatory direction of travel. The second reason given for riding against the mandatory direction of travel is route simplification. No other literature was found that investigates this behaviour.
- **Left turn manoeuvre:** In comparison to motor vehicles, which are constrained to one method of carrying out a left turn, bicyclists have a number of rule conform and non-conform options for implementing this manoeuvre (depicted in Figure 1):
  1. Direct left turn – turn with the motor vehicle traffic in one signal phase (l in Figure 1)
  2. Indirect left turn – turn over two phases using a pedestrian style turn (m in Figure 1)
  3. Indirect left turn (wrong way) – similar to the indirect left turn but moving against the mandatory direction of travel during both of the turning phases (n in Figure 1)

The selection of one of these paths appears to be influenced by the state of the traffic signal and the infrastructure used upon arrival (Amini et al., 2016).

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

3. Methodology

Revealed choice data from bicyclists at four research intersections in Munich, Germany are used to develop logistic regression models. Intersections were selected with differing types of bicycle infrastructure (on-road bicycle lane, separated bicycle path and no specific bicycle infrastructure), volumes of bicycle and motor vehicle traffic and road geometry. A two-hour segment of video data recorded during the morning peak hour at each intersection was selected for a detailed behavioural analysis. Detailed data describing the tactical behaviour of the bicyclists and the situation at the intersection were extracted. Finally, timing information from the traffic actuated signals was supplied by the City of Munich and was linked to the observed data using a corrected time stamp.

A total of 37 independent variables describing the strategic and prior tactical choices of the bicyclist as well as the situational factors are used as predictors in the logistic regression models. A visual representation of selected independent variables and dependent variables (tactical choices) is shown in Figure 1.

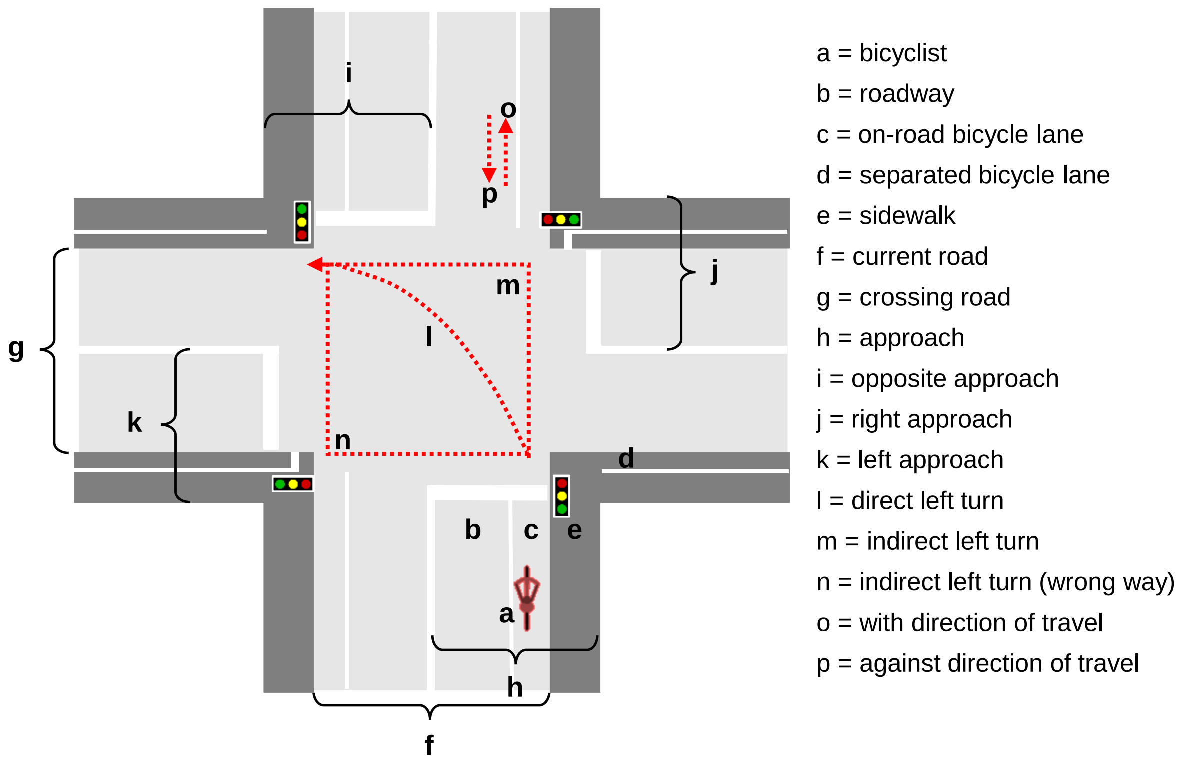

Figure 1 Graphical representation of selected tactical choice options and situational factors

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

Adequate variation in the independent variables is assured through the selection of research intersections with approaches differing from one another in terms of traffic flow and geometry. Variables describing the presence of other road users as well as the state of the signal control at the moment a bicyclist arrives provide further variation between cases. A list of the categorical independent variables with observed frequencies is shown in Table 1. The continuous independent variables with descriptive statistics are given in Table 2.

Independent variable	**Category 1**	**Category 2**	**Category 3**
Strategic / prior tactical choice			
Manoeuvre	Straight N=4040 (80.4%)	Right N=454 (9.0%)	Left N=534 (10.6%)
Infrastructure selection	Bicycle lane N=3532 (94.8%)	Roadway N=67 (1.8%)	Sidewalk N=128 (3.4%)
Geometry			
Bicycle lane	None N=634 (12.4%)	Bicycle lane N=4485 (87.6%)	
Bicycle lane type	None N=634 (12.4%)	On-road N=2070 (40.4%)	Separated N=2415 (47.2%)
Parking	None N=2268 (44.3%)	Parking N=2851 (55.7%)	
Left turn lane	None N=2484 (48.5%)	Left turn lane N=2635 (51.5%)	
Centre island	None N=1729 (33.8%)	Centre island N=3390 (66.2%)	
Traffic			
Right lane occupancy	No N=1617 (33.5%)	Yes N=3214 (66.5%)	
Signal control			
Signal phase	Red N=2817 (55.6%)	Green N=2253 (44.4%)	
Specific bicycle signal	Shared signal N=2438 (47.6%)	Bicycle signal N=2681 (52.4%)	

Table 1 Description of categorical independent variables

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

Independent variable	**Unit**	**Mean**	**Std. Dev.**	**Min**	**Max**
Geometry					
Bicycle lane width	m	1.5	0.7	0.0	2.2
Sidewalk width	m	3.5	1.6	0.9	9.3
Roadway width (approach)	m	7.9	3.0	0.0	12.0
Roadway width (opposite approach)	m	5.3	1.7	0.0	10.9
Driving lanes (approach)	-	2.4	0.9	0	4
Driving lanes (opposite approach)	-	1.6	0.6	0	2
Total roadway width (current road)	m	18.3	7.0	8.8	28.6
Total roadway width (crossing road)	m	16.8	5.5	8.8	28.6
Total driving lanes (current road)	-	4.0	1.4	2	6
Total driving lanes (crossing road)	-	3.6	1.2	2	6
Traffic					
Cars in approach	-	2.5	2.0	0	10
Trucks in approach	-	0.1	0.4	0	3
Pedestrians in approach	-	1.3	1.8	0	20
Bicyclists in approach	-	1.6	2.1	0	16
Traffic volume (approach)	veh/h	646.6	309.7	0	1800
Traffic volume (crossing road)	veh/h	523.9	286.6	0	1800
Percentage HDV and buses (approach)	%	4.9	4.0	0	16
Percentage HDV and buses (crossing road)	%	5.6	4.2	0	16
Bicyclist volume (approach)	bicycles/h	337.7	201.3	0	660
Bicyclist volume (crossing road)	bicycles/h	199.4	159.2	0	660
Pedestrian volume (approach)	ped/h	255.9	238.2	0	1160
Pedestrian volume (crossing road)	ped/h	329.5	358.3	0	1160
Signal control					
Time since last phase change (red)	s	25.5	16.2	0	98
Time since last phase change (green)	s	14.8	11.8	0	102
Time until next phase change (red)	s	28.0	17.9	0	91
Time until next phase change (green)	s	17.2	16.0	0	103
Phase length	s	43.3	17.2	7	104

Table 2 Description of continuous independent variables

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

A logistic regression model for each of the tactical choices is specified and calibrated using recursive feature elimination, which combines k-fold cross validation and predictor selection based on the log likelihood of the model. The main effects and two-way interactions between the situational variables listed in Table 1 and Table 2 are used as an initial set of explanatory variables. The following steps are taken to identify the optimal set of explanatory variables for each of the tactical choice models and estimate the corresponding β parameters:

1. **Data pre-processing:** Relevant cases are extracted from the complete dataset to analyse each of the tactical choices. For example, to estimate the regression model for predicting the response to a red signal, only data from bicyclists who encountered a red signal are selected (N=1935). Each of these data subsets is unique and is pre-processed prior to model estimation. This is done by removing variables that contain zero or near to zero variance. The pair-wise correlations between the remaining variables are assessed to identify inter-correlated variables. If a pair-wise correlation greater than 0.6 is identified, the variable with the largest mean correlation with all other variables is removed from the dataset. Data pre-processing is carried out in two phases. In the first phase, the main effects of variables listed in Table 2 are assessed and variables with near to zero variance and high correlations with other variables are removed from the dataset. Pair-wise interaction terms for the remaining variables are created and the pre-processing procedure is repeated.

2. **Recursive feature elimination:** The resulting dataset is divided into $k = 10$ equal subsets for the k-fold cross validation. The model is estimated using $k - 1$ of the subsets and is validated using the remaining subset. This is repeated k times using each of the data subsets once for validation. The backwards elimination process is carried out within each fold. The model is estimated using all of the variables remaining after the pre-processing step. The predictive power of the model, which is assessed using the Area Under the Curve (AUC) for binomial regression and accuracy for multinomial regression, is assessed using the held back dataset and the variables are ranked based on their importance. The least important variable

is removed and the model is re-estimated with the remaining variables. This is repeated until only one variable remains in the model. The optimal set of predictors (largest AUC or accuracy) is identified for each fold. The performance profiles of the variable subsets are calculated over all the held back samples of the k-fold cross validation and the optimal set of predictors is determined.

3. **Full model estimation:** The entire data subset is used to estimate the β values for the identified optimal set of predictors. In order to improve the interpretability of the regression models, the main effect of both variables in two-way interaction terms are added to the optimal set of predictors for the final model. This is done even if the main effects do not improve the predictive power of the model.

4. **Simplified model estimation:** The β values are re-estimated for a reduced model comprised of only the predictors found to be statistically significant ($p < 0.01$) in the full model. In cases where a main effect can replace an interaction term, the main effect predictor is given preference to maintain model simplicity.

Binomial and multinomial regression models are estimated and evaluated using the statistics software package R (The R Foundation, 2016). The recursive feature elimination (RFE) function of the classification and regression training package caret (Kuhn, 2016) is used to identify the most powerful set of predictors from the 37 variables using combined backward selection and k-fold cross validation. The reduced models are presented in this paper because the detail offered is deemed sufficient for application in microscopic traffic simulation.

The Receiver Operating Characteristic (ROC) curve, which compares the false positive rate with the true positive rate of a binary predictor at various classification thresholds, is used to assess the binomial regression models and identify the optimal classification threshold. According to Hosmer et al. (2013, p. 174), “this measure has now become the standard for evaluating a fitted model’s ability to assign, in general, higher probabilities of the outcome to the subgroup who develop the outcome (y=1) than it does to the subgroup who do not develop the outcome (y=0)”

The predictive power of the model can be deduced from the Area Under the Curve (AUC), with larger areas signifying higher predictive power. AUC values range between 0.5 and 1.0, where 0.5 indicates that the model is no better at predicting the outcome than random chance and 1.0 a perfect prediction. In general, AUC values between 0.5-0.7 indicate poor discrimination that is marginally better than random change, 0.7-0.8 indicates acceptable discrimination, 0.8-0.9 signifies excellent discrimination and above 0.9 shows outstanding discrimination (Hosmer et al., 2013).

In addition to evaluating the power of the logistic regression model, the ROC Curve is useful for selecting a well-suited cut-off point for the classification. Typically the cut-off point for a classification model is set at 0.5 such that if $P(y = 1) \geq 0.5$, the outcome is predicted to be one. This value can be shifted, however, to maximize the sensitivity and specificity of the regression model. Here, a cut-off point is selected for each of the models that is plotted on the upper most left corner of the ROC Curve.

Along with AUC, the following metrics derived from classification tables are used to evaluate the predictive power of the models:

- Accuracy ($\sum True\ positve + \sum True\ negative / \sum Predictions$)
- Sensitivity ($\sum True\ positive / \sum Condition\ positive$)
- Specificity ($\sum True\ negative / \sum Condition\ negative$)
- Positive Predictive Value ($\sum True\ positive / \sum Test\ outcome\ positive$)
- Negative Predictive Value ($\sum True\ negative / \sum Test\ outcome\ negative$)

To evaluate multinomial logistic regression models, these evaluation parameters are generalised to the Mean Sensitivity, Mean Specificity, Mean Positive Predictive Value and Mean Negative Predictive Value across all choice categories.

4. Results

The resulting reduced regression models are presented in this section. The tactical choices selected for analysis, all of which are described using nominal variables with two or three

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

categories, are listed in Table 3 along with the number and percentage of bicyclists observed carrying out each tactical option.

Tactical choice	Category 1	Category 2	Category 3
Infrastructure selection (no bicycle lane) N=451	Roadway N=428 (94.9%)	Sidewalk N=23 (5.1%)	
Infrastructure selection (bicycle lane) N=3727	Bicycle lane N=3532 (94.8%)	Roadway N=67 (1.8%)	Sidewalk N=128 (3.4%)
Response to red signal N=1935	Stop N=1552 (80.2%)	Violate N=383 (19.8%)	
Direction of travel N=4710	With direction N=4651 (98.7%)	Against direction N=59 (1.3%)	
Left turn manoeuvre N=426	Direct turn N=66 (15.5%)	Indirect turn N=166 (39.0%)	Indirect turn (wrong way) N=194 (45.5%)

Table 3 Tactical choices with categories and observed counts and percentages

For each of the tactical choices, the optimal set of predictors is identified using recursive feature elimination and k-fold cross validation. The final model is estimated using predictors found to be statistically significant ($p \leq 0.01$). The predictors are sorted by their predictive power within the main effects and interaction effects. The most important predictors in each model are discussed and compared with the findings of previous studies.

4.1. Infrastructure selection without bicycle lane

Over 95% of the observed bicyclists use the roadway on approaches with no bicycle lane. A reduced model consisting of only two predictors is found to provide acceptable predictive power (AUC = 0.76). Traffic attributes on the approach have an important influence on infrastructure choice; the likelihood of using the roadway decreases by 1.76 (0.58^{-1}) times for each additional car in the approach. This finding echoes that of Kuller et al. (1986), who found that high traffic volumes discourage roadway use. The manoeuvre of the bicyclist also affects the choice outcome. According to this model, bicyclists turning left are 6.09 (0.16^{-1}) times less likely to use the roadway.

This finding seems counterintuitive but is due to the fact that many bicyclists turning left ride against the mandatory direction of travel (route simplification) and therefore use the sidewalk rather than the roadway.

N = 451 **Sidewalk use = 0, Roadway use = 1**	**β**	**Odds ratio**	**Sig.**
Intercept	4.24	69.20	0.000
Cars in approach	-0.54	0.58	0.000
Manoeuvre (left turn)	-1.81	0.16	0.000
Classification threshold:		0.95	
AUC		0.76	
Accuracy		0.73	
Sensitivity		0.73	
Specificity		0.78	
Positive Predictive Value		0.98	
Negative Predictive Value		0.13	

Table 4 Binomial logistic regression model and k-fold cross validation for infrastructure selection without bicycle lane

While roadway use is predicted with considerable success, the prediction of sidewalk use proves to be less reliable. This could indicate that bicyclists choose to use the sidewalk for reasons that are unobservable, such as a feeling of safety or the anticipation of upcoming manoeuvre. Additionally, the low number of sidewalk use observations limits the potential to identify patterns between the independent variables and this choice outcome. A high classification threshold of 0.95 is identified, which addresses the observed skewness in decision outcomes by shifting predictions into the sidewalk category.

4.2. Infrastructure selection with bicycle lane

If a bicycle lane is provided, bicyclists tend to use this facility. Over 95% of bicyclists observed on approaches with a bicycle lane selected this infrastructure, which is slightly higher than the 90% found by Alrutz et al. (2009). Infrastructure selection can be framed as a discrete choice with three possible outcomes, bicycle lane, roadway or sidewalk. Initially, a multinomial logistic regression model was estimated to predict infrastructure use. Although correlations were found between the predictors and the choice outcome, these correlations were not strong enough to estimate a model capable of predicting roadway or sidewalk use.

In order to capitalise on the simplicity of binomial logistic regression as well as the adjustable classification threshold, the model shown in Table 5 is developed to predict whether a bicyclist will use an available bicycle lane or not. The strongest predictor of bicycle lane use is a right turn manoeuvre, which decreases the likelihood of using the bicycle lane by 6.03 (0.17^{-1}) times, due to increased sidewalk use. The width of the bicycle lane plays an important role in the choice, with the likelihood of bicycle lane use increasing by 1.25 times for each additional cm of width ($e^{\frac{21.56}{100}}$). This effect is moderated by the volume of bicycle traffic on the approach, which decreases bicycle lane use by 1.09 times for each increase in one bicycle per hour. The presence of other road users in the approach has an interesting effect on bicycle lane use. If there are only cars or only pedestrians present, the likelihood of bicycle lane use is reduced. If both are present, however, the interaction term increases the probability of bicycle lane use. This make intuitive sense as the presence of other road users on the sidewalk and roadway likely push bicyclists into an available bicycle lane. The presence of other bicyclists on the other hand, propels bicyclists from the bicycle lane, particularly on separated facilities.

The model predicts bicycle lane use with acceptable overall accuracy. However, the prediction of bicycle lane use is more reliable than that of not using the bicycle lane. This is likely due to the overrepresentation of bicycle lane observations in the sample and the potential role of personal attributes and unobservable factors in the choice to use the roadway or sidewalk when a bicycle lane is available. The high classification threshold of 0.96 coerces the prediction of not using the bicycle lane, but these predictions are often incorrect (low negative predictive values).

The finding of previous studies indicate that the width and type of the bicycle lane are decisive in infrastructure selection while traffic conditions are unimportant (Alrutz et al., 2009; Guo et al., 2013). According to the findings here, the number and type of road users in the approach

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

have a strong influence on infrastructure choice. Previous studies found that wider bicycle lanes have a higher rate of acceptance, which is confirmed. However, unlike Alrutz et al. (2009), on-road bicycle lanes are found here to have a higher acceptance than separated facilities.

N = 3727 **No bicycle lane use = 0, Bicycle lane use = 1**	**β**	**Odds ratio**	**Sig.**
Intercept	-30.86	0.00	0.000
Manoeuvre (right turn)	-1.80	0.17	0.000
Bicyclist volume – approach (bicycles/h)	0.09	1.09	0.000
Bicycle lane width (m)	21.56	2.32e9	0.000
Bicycle lane type (separated)	-5.65	0.00	0.001
Driving lanes (same direction)	1.24	3.47	0.000
Sidewalk width (m)	-1.50	0.22	0.002
Centre island	1.81	6.10	0.010
Parking	-1.67	0.19	0.000
Pedestrians in approach	-0.16	0.85	0.033
Bicyclists in approach	-0.04	0.96	0.421
Cars in approach	-0.01	0.99	0.828
Bicycle lane width (m) * Bicyclist volume – approach (bicycles/h)	-0.05	0.95	0.000
Bicycle lane type (separated) * Bicyclists in approach	-0.28	0.76	0.000
Bicycle lane type (separated) * Sidewalk width (m)	0.75	2.11	0.056
Cars in approach * Pedestrians in approach	0.06	1.06	0.047
Classification threshold:		0.96	
AUC		0.76	
Accuracy		0.73	
Sensitivity		0.73	
Specificity		0.72	
Positive Predictive Value		0.98	
Negative Predictive Value		0.13	

Table 5 Binomial logistic regression model and k-fold cross validation for infrastructure selection with bicycle lane

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

4.1. Response to a red signal

When faced with a red traffic signal, roughly a fifth of the observed bicyclists violated the signal. The manoeuvre carried out by a bicyclist plays a very strong role in whether or not he or she will stop at a red light. Bicyclists turning right are 134.15 times more likely to run a red light than those riding straight across the intersection. On one-way roads, bicyclists turning left are 13.76 times more likely to violate a red light (Roadway width – opposite = 0). For each meter of roadway width in the opposite direction, this probability increases by 1.33 times. This is due to the fact that carrying out an indirect left turn against the mandatory direction of travel includes the violation of the first traffic signal. The time elapsed since the signal became red has a deterring effect on red light violations; bicyclists become 1.35 ($\frac{1}{e^{-0.03*10}}$) times less likely to violate the signal for each ten seconds passed since the signal became red.

N = 1935 **Stop = 0, Violate = 1**	**β**	**Odds ratio**	**Sig.**
Intercept	-1.22	0.29	0.000
Manoeuvre (right turn)	4.90	134.16	0.000
Time since signal change (s)	-0.03	0.97	0.000
Manoeuvre (left turn)	2.62	13.76	0.000
Roadway width – opposite (m)	-0.28	0.76	0.000
Manoeuvre (left turn) * Roadway width – opposite (m)	0.28	1.33	0.008
Classification threshold:		0.46	
AUC		0.92	
Accuracy		0.91	
Sensitivity		0.85	
Specificity		0.93	
Positive Predictive Value		0.74	
Negative Predictive Value		0.96	

Table 6 Binomial logistic regression model and k-fold cross validation for response to red signal

The estimated binomial logistic regression model estimates the choice outcome with high accuracy. The prediction of signal compliance is slightly more reliable than that of signal violation. However, the prediction rates for both suggest that this behaviour is highly influenced by observable situational factors. The resulting model supports previous studies that found that turning right increases the probability of violating a red light (Johnson et al., 2011). The influence of additional parameters, such as the signal phase, infrastructure selection and left turn manoeuvre, are identified here.

4.2. Direction of travel

The vast majority of the observed bicyclists rode in the mandatory direction of travel (98.7%). According to the reduced model, bicyclists turning left are 9.25 times more likely to travel against the direction of travel than those carrying out other manoeuvres. Interestingly, the availability of a left turn lane discourages travelling against the direction of travel to a large extent (6.36 ($\frac{1}{e^{-1.85}}$) times decrease in likelihood). Parking increases the likelihood of riding against the mandatory direction of travel by 2.67 times while the presence of a separated bicycle lane decreases the likelihood by 9.87 ($\frac{1}{e^{-2.29}}$) times. Together, these two factors increase the probability of riding the wrong way by 8.60 times.

Far fewer bicyclists were observed riding against the mandatory direction of travel in this study than reported by Kuller et al. (1986) and Wachtel & Lewiston (1994). The only finding that could be verified in this study is that bicyclists turning left are more likely to ride the wrong way, reflecting the goal of path simplification that was noted by Kuller et al. (1986).

The cross-validation indicates a high success rate for predicting riding with the direction of travel and a low success rate for predicting those riding against the given direction of travel. The very low classification threshold of 0.02 reflects the skewing in the choice observations and the low positive predictive value of 0.04 reflects the inaccuracy caused by manipulating the classification threshold.

N =4710 **With direction = 0, Against direction = 1**	**β**	**Odds ratio**	**Sig.**
Intercept	-4.50	0.01	0.000
Manoeuvre (left turn)	2.23	9.25	0.000
Left turn lane	-1.85	0.16	0.000
Bicycle lane type (separated)	-2.29	0.10	0.019
Parking	0.98	2.67	0.202
Parking * Bicycle lane type (separated)	2.15	8.60	0.035
Classification threshold:		0.02	
AUC		0.77	
Accuracy		0.76	
Sensitivity		0.78	
Specificity		0.76	
Positive Predictive Value		0.04	
Negative Predictive Value		0.99	

Table 7 Binomial logistic regression model and k-fold cross validation for riding direction

4.3. Left turn manoeuvre

A multinomial regression model with three choice outcomes, direct left turn, indirect left turn and indirect left turn against the mandatory direction of travel, is estimated to predict the type of manoeuvre carried out by the bicyclist. The three choice outcomes are described in Section 2 and are shown graphically in Figure 1. The estimated β parameters are in reference to the base category, which is the direct left-hand turn.

The most important predictor for the type of left turn is roadway use; bicyclists using the roadway are more than 60 ($\frac{1}{e^{-4.12}}$) times less likely to carry out an indirect left turn and 20.5 ($\frac{1}{e^{-3.02}}$) times less likely to carry out an indirect left turn against the given direction of travel. Two characteristics of the infrastructure design, the type of bicycle lane and the presence of car parking, influence the choice outcome. The probability of an indirect left turn increases by 3.42 times if there is only parking and 6.05 times if there is only a separated bicycle lane. If both these features are present, the likelihood of this manoeuvre increases by 4.43 ($e^{1.80+1.23+-1.54}$) times. A similar

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

mechanism is at play for the choice to execute an indirect left turn against the mandatory direction of travel (5.18 times increase with parking only, 9.54 times increase with a separated bicycle lane only and 9.21 ($e^{2.26+1.64+-1.68}$) times with both features). The signal phase and the presence of other road users also play an important role in the left turn choice.

N = 426 **Base category = Direct left turn**		**β**	**Odds ratio**	**Sig.**
Indirect left turn	Intercept	-0.40	0.67	0.704
	Infrastructure selection (roadway)	-4.12	0.02	0.000
	Bicycle lane type (separated)	1.80	6.05	0.052
	Parking	1.23	3.42	0.165
	Signal phase (green)	1.17	3.23	0.023
	Bicyclists in approach	0.53	1.69	0.008
	Parking * Bicycle lane type (separated)	-1.54	0.22	0.158
	Signal phase (green) * Bicyclists in approach	-0.72	0.49	0.003
Indirect left turn (wrong way)	Intercept	0.56	1.74	0.606
	Infrastructure selection (roadway)	-3.02	0.05	0.000
	Bicycle lane type (separated)	2.26	9.54	0.020
	Parking	1.64	5.18	0.076
	Signal phase (green)	-2.13	0.12	0.000
	Bicyclists in approach	0.30	1.34	0.121
	Parking * Bicycle lane type (separated)	-1.68	0.19	0.136
	Signal phase (green) * Bicyclists in approach	-1.11	0.33	0.022
Accuracy			0.73	
Mean Sensitivity			0.70	
Mean Specificity			0.85	
Mean Positive Predictive Value			0.73	
Mean Negative Predictive Value			0.86	

Table 8 Multinomial regression model and k-fold cross validation for left turn manoeuvre

In contrast to the multinomial regression model for infrastructure selection, which failed to predict roadway and sidewalk use, the multinomial regression model for the left turning manoeuvre provides exceptional predictions for all three types of turn. The predictive power of this model suggests that this decision is greatly influenced by observable situational factors at the intersection.

The resulting model supports the findings of a previous study that found that bicyclists using the roadway often carry out direct left turns while bicyclists on the sidewalk and bicycle lane do not (Amini et al., 2016). The signal stage and the time since the last stage change are also found to play a role in manoeuvre selection, as previously found.

5. Discussion and Conclusion

The findings presented in this paper are useful for understanding the relationships between the tactical choices of bicyclists at signalised intersections and the situational factors, strategic and prior tactical choices of the bicyclist. This knowledge is useful for predicting the behaviour of bicyclists and designing infrastructure and traffic control that takes this knowledge into consideration. For example, the manoeuvre of a bicyclist is found to have a strong influence on a number of tactical choices at intersections, including the reaction to a red light, direction of travel and infrastructure selection. If turning rates are known for given intersections, the levels of rule breaking behaviour such as red light violations and riding against the mandatory direction of travel can be predicted. Additional information describing the static attributes of the intersection, such as the geometry and average traffic volumes, provides further input for predicting tactical behaviour without using dynamic attributes. If it is possible to determine the dynamic state of the intersection, including the number of road users currently present and the state of the traffic signal, the choice outcomes can be predicted with more accuracy. Possible applications of dynamic models include traffic flow simulations, driver assistance systems and autonomous driving systems.

There is considerable variation between the predictive power of the four regression models. The models estimated to predict the reaction to a red signal and the type of left turn manoeuvre are capable of predicting each of the choice categories with exceptional accuracy. In contrast, the infrastructure selection models and the direction of travel model have difficulty predicting seldom occurring choice outcomes. Two possible explanations for this variation are suggested. First, the tactical behaviours may be motivated by different types of factors. The tactical behaviours that are

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

highly explainable using the developed regression models are likely motivated by situational factors that are externally observable. The difficulty in predicting at least one of the outcomes for the remaining behaviours, the direction of travel and infrastructure selection, suggests that the motivation factors for these choices are intrinsic or cannot be observed in the restricted observation area of one intersection. The predictive power of the estimated models is indicative of which rule breaking behaviours of bicyclists can be addressed through infrastructure design and traffic signal control and which are rooted in non-observable factors. The high predictive power of the models for the reaction to a red signal and the type of left turn manoeuvre suggest that these behaviours can be modified by altering the situation at the intersection. Riding against the given direction of travel and infrastructure selection, on the other hand, may be more responsive to softer measures such as safety campaigns and traffic rule enforcement.

A second possible explanation for the variation in predictive power is the number of observations for each of the choice outcomes. The observational data for the models with higher predictive power (left turn manoeuvre and red signal reaction) contained a relatively balanced distribution of the choice outcomes. Conversely, the tactical choices that are dominated by one outcome (e.g. using a bicycle lane and riding with the mandatory direction of travel) are found to be more difficult to predict using regression models. The low number of observations for seldom occurring events makes it very difficult to discern patterns in the independent variables that lead to this outcome. This is compensated in this paper by selecting varying classification thresholds and coercing predictions into the seldom occurring category. However, these coerced predictions are often incorrect, leading to very low negative (or positive) predictive values. Integration of further observations from additional intersections with differing geometric and traffic characteristics may increase the predictive power of these regression models.

Nevertheless, in light of the high accuracy and relatively balanced prediction success for all the possible outcomes, the regression models developed for the selection of a left turn manoeuvre and response to a red signal could be used by researchers and practitioners to predict the outcome of these choices. The effects of intrinsic factors, such as the socio-economic characteristics of the

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

bicyclist, personality traits (e.g. aggressiveness and nervousness), as well as factors that are observable over a long distance (e.g. route choice) on the tactical choices of bicyclists at signalised intersections would be an interesting extension of this work for future research.

6. Acknowledgements

This research is supported by the Federal Ministry of Economics and Technology on the basis of a decision by the German Bundestag. Research was carried out within the framework of the project UR:BAN (Urban Space: User oriented assistance systems and network management). Thank you to Prof. Dr. Constantinos Antoniou for his advice and Dr. Stephan Haug for his statistical guidance.

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

7. References

Allen, D., Bygrave, S., Harper, H., 2005. Behavior at Cycle Advanced Stop Lines, Transport for London. London.

Alrutz, D., Bohle, W., Müller, H., Prahlow, H., Hacke, U., Lohmann, G., 2009. Unfallrisiko und Regelakzeptanz von Fahrradfahrern. Berichte der Bundesanstalt für Straßenwesen. Unterr. Verkehrstechnik. 184, 1–127.

Alrutz, D., Meewes, V., 1980. Untersuchungen zum Radverkehr in Köln. Mitteilung der Beratungsstelle für Schadensverhütung. Cologne.

Amini, S., Twaddle, H., Leonhardt, A., 2016. Modeling of the Tactical Path Selection of Bicyclists at Signalized Intersections, in: Transportation Research Board 95th Annual Meeting Proceedings.

Aultman-Hall, L., Hall, F., 1998. Research Design Insights from a Survey of Urban Bicycle Commuters. Transp. Res. Rec. 1636, 21–28. doi:10.3141/1636-04

Dekoster, J., Schollaert, U., Bochu, C., 2000. Cycling: the way ahead for towns and cities. Office for Official Publications of the European Communities.

Gerstenberger, M., 2015. Accidents at Intersections: Basic Analysis of Causation and Approaches for Improvements. Doctoral dissertation Technical University of Munich.

Guo, H., Wang, W., Guo, W., Zhao, F., 2013. Modeling lane-keeping behavior of bicyclists using survival analysis approach. Discret. Dyn. Nat. Soc. 2013. doi:http://dx.doi.org/10.1155/2013/197518

Herslund, M.-B., Jørgensen, N.O., 2003. Looked-but-failed-to-see-errors in traffic. Accid. Anal. Prev. 35, 885–891. doi:http://dx.doi.org/10.1016/S0001-4575(02)00095-7

Hosmer, D.W.J., Lemeshow, S., Sturdivant, R.X., 2013. Applied logistic regression. John Wiley & Sons.

Johnson, M., Charlton, J., Oxley, J., Newstead, S., 2013. Why do cyclists infringe at red lights? An investigation of Australian cyclists’ reasons for red light infringement. Accid. Anal. Prev. 50, 840–847. doi:doi:10.1016/j.aap.2012.07.008

Please cite: Twaddle, Heather, & Busch, F. (2019). Binomial and multinomial regression models for predicting the tactical choices of bicyclists at signalised intersections. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 47–57. https://doi.org/https://doi.org/10.1016/j.trf.2018.10.002

Johnson, M., Newstead, S., Charlton, J., Oxley, J., 2011. Riding through red lights: The rate, characteristics and risk factors of non-compliant urban commuter cyclists. Accid. Anal. Prev. 43, 323–328. doi:10.1016/j.aap.2010.08.030

Kuhn, M., 2016. The caret Package [WWW Document]. URL http://topepo.github.io/caret/index.html (accessed 7.18.16).

Kuller, E.C., Gersemann, D., Ruwenstroth, G., 1986. Regelabweichendes Verhalten von Fahrradfahrern. Forschungsberichte der Bundesanstalt für Straßenwesen, Bereich Unfallforsch.

Lusk, A.C., Furth, P.G., Morency, P., Miranda-Moreno, L.F., Willett, W.C., Dennerlein, J.T., 2011. Risk of injury for bicycling on cycle tracks versus in the street. Inj. Prev. 17, 131–135. doi:doi:10.1136/ip.2010.028696

Michon, J.A., 1985. A critical view of driver behavior models: what do we know, what should we do? Hum. Behav. traffic Saf. 485–520. doi:10.1007/978-1-4613-2173-6

Moritz, W., 1998. Adult bicyclists in the United States: characteristics and riding experience in 1996. Transp. Res. Rec. J. Transp. Res. Board 1–7. doi:http://dx.doi.org/10.3141/1636-01

Ortlepp, J., 2009. Verbesserung der Verkehrssicherheit in Münster. Straßenverkehrstechnik 53.

Pai, C.-W., Jou, R.-C., 2014. Cyclists' red-light running behaviours: An examination of risk-taking, opportunistic, and law-obeying behaviours. Accid. Anal. Prev. 62, 191–198. doi:http://dx.doi.org/10.1016/j.aap.2013.09.008

Räsänen, M., Summala, H., 1998. Attention and expectation problems in bicycle-car collisions: An in-depth study. Accid. Anal. Prev. 30, 657–666. doi:10.1016/S0001-4575(98)00007-4

Reynolds, C.C.O., Harris, M.A., Teschke, K., Cripton, P. a, Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. Environ. Heal. 8. doi:10.1186/1476-069X-8-47

Richardson, M., Caulfield, B., 2015. Investigating traffic light violations by cyclists in Dublin City Centre. Accid. Anal. Prev. 84, 65–73. doi:http://dx.doi.org/10.1016/j.aap.2015.08.011

Rivara, F.P., Thompson, D.C., Thompson, R.S., 1997. Epidemiology of bicycle injuries and risk factors for serious injury. Inj. Prev. 3, 110–114.

Rodgers, G.B., 1995. Bicyclist deaths and fatality risk patterns. Accid. Anal. Prev. 27, 215–223. doi:doi:10.1016/0001-4575(94)00063-R

Schramm, A.J., Rakotonirainy, A., Haworth, N.L., 2008. How much does disregard of road rules contribute to bicycle-vehicle collisions? Proc. High risk road users - Motiv. Behav. Chang. what Work. what doesn't Work. Natl. Conf. Australas. Coll. Road Saf. Travel. Comm. Queensl. Parliam. 2008.

Summala, H., Pasanen, E., Räsänen, M., Sievänen, J., 1996. Bicycle accidents and drivers' visual search at left and right turns. Accid. Anal. Prev. 28, 147–153. doi:doi:10.1016/0001-4575(95)00041-0

Teschke, K., Harris, M.A., Reynolds, C., Winters, M., Babul, S., Chipman, M., Cusimano, M., Brubacher, J., Hunte, G., Friedman, S., 2012. Route infrastructure and the risk of injuries to bicyclists: a case-crossover study. Am. J. Public Health 102, 2336–2343. doi:doi: 10.2105/AJPH.2012.300762

Wachtel, A., Lewiston, D., 1994. Risk factors for bicycle-motor vehicle collisions at intersections. ITE J. (Institute Transp. Eng. 64, 30–35.