Rethinking Route Choices! On the Importance of Route Selection in Wayfinding Experiments

Bartosz Mazurkiewicz
Geoinformation, TU Wien, Austria

Markus Kattenbeck
Geoinformation, TU Wien, Austria

Ioannis Giannopoulos
Geoinformation, TU Wien, Austria

Abstract

Route selection for a wayfinding experiment is not a trivial task and is often made in an undocumented way. Only recently (2021), a systematic, reproducible and score-based approach for route selection for wayfinding experiments was published. However, it is still unclear how robust study results are across all potential routes in a particular experimental area. An important share of routes might lead to different conclusions than most routes. This share would distort and/or invert the study outcome. If so, the question of selecting routes that are unlikely to distort the results of our wayfinding experiments remains unanswered. In order to answer these questions, an agent-based simulation study with four different sample sizes (N = 15, 25, 50, 3000 agents) comparing Turn-by-Turn and Free Choice Navigation approaches (between-subject design) regarding their arrival rates on more than 11000 routes in the city center of Vienna, Austria, was run. The results of our study indicate that with decreasing sample size, there is an increase in the share of routes which lead to contradictory results regarding the arrival rate, i.e., the results become less robust. Therefore, based on simulation results, we present an approach for selecting suitable routes even for small-scale in-situ studies.

2012 ACM Subject Classification Information systems → Decision support systems; Computing methodologies → Agent / discrete models; Information systems → Location based services

Keywords and phrases Route Selection, Route Features, Human Wayfinding, Navigation, Experiments, Experimental Design

Digital Object Identifier 10.4230/LIPIcs.COSIT.2022.6

1 Introduction

Novel navigation system paradigms for wayfinders are still the subject of ongoing research. Regardless of the target group, i.e., whether it would be pedestrians [5, 7, 12], cyclists [20, 16] or car drivers [11] many decisions during experimental design must be made. While these decisions may impact the study results, this impact is often neither evident nor easy to estimate. One of these decisions relates to the selection of a route suitable for a particular wayfinding study. Given a potential experimental area of non-trivial size, there are at least thousands of potential routes researchers can select from (see Section 3). The potential influence of different routes on study results, however, has not been scrutinized systematically. Given the myriad of potential routes and the different characteristics they come with, there might be an important share of routes that lead to study results deviating from the mean calculated over all possible routes (population mean). By means of an agent-based simulation study comparing two different navigation approaches for all potential routes in a selected experimental area, we will provide evidence that with decreasing sample size, the share of routes which lead to contradicting results increases. Given these differences in results, we

15th International Conference on Spatial Information Theory (COSIT 2022).
Editors: Toru Ishikawa, Sara Irina Fabrikant, and Stephan Winter; Article No. 6; pp. 6:1–6:13
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

propose a selection process of appropriate routes, i.e., routes that provide stable results across sample sizes and lead to results congruent with the population mean. Hence, our approach is useful for route selection in comparative wayfinding studies, even for smaller sample sizes.

We will provide evidence that route selection is a crucial step in experimental design, as it shows the potential to turn around the study results. Therefore, more attention should be given to this phase of experimental design. Our approach can be combined (see Section 5.3) with route selection methods for wayfinding experiments (see e.g., [15]), which have been proposed so far.

2 Related Work

In this section, two branches of related work will be discussed. First, we review systematic approaches for route selection during experimental design and route justification. Second, we will discuss comparative wayfinding studies which involve at least two routes and examine whether the route itself was treated as an independent variable in the analysis.

2.1 Systematic Route Selection for Wayfinding Studies

In our previous work, we did an exhaustive search of 'six major venues (conferences and journals) in the broader area of geographic information science and related fields' [15, p. 2] between 2010 and early 2020 regarding route descriptions and/or justifications in studies involving wayfinding tasks with a predefined route. In total, 32 papers fell into this category. The conclusion was that, in general, route choice was poorly justified and that only half of the selected publications mentioned the route length, which was considered a basic feature. This leaves the impression that route selection in wayfinding experiments tends to lack appropriate justification, given the potential impact a route may have on results. In very recent studies (i.e. from 2020 onwards), examples of both missing and explicit route selection justification can be found. Dong and colleagues [3] compared augmented reality (AR) and 2D navigation electronic maps in pedestrian wayfinding. The selection of three experiment routes was not explicitly justified. There are as well examples of explicit and elaborated route justifications. Benelia compared paper maps with audiovisual Turn-by-Turn (TBT) instructions in the context of spatial learning for car drivers [2]. While selecting the route, Benelia tried to maximize personal safety, to avoid high levels of stress in participants and to have sufficient stimuli along the route. Another example of explicit route justification can be found in [20] comparing TBT and ACTF (As-The-Crow-Flies) navigation approaches for cyclists. Both routes used in this publication were designed to contain a segment on which the participant had to cycle contrary to the compass direction pointing to the destination. This feature was crucial to the experimental design.

Although both examples present an explicit justification for route selection, they are not necessarily reproducible because several routes with those characteristics are possible and they might lead to different results. In order to tackle this problem, we previously proposed a methodological average-based framework for systematic and reproducible route selection [15]. All possible routes are ranked according to criteria and corresponding weights, which the researcher must set. This flexibility allows finding routes that exactly fit the requirements of the study. However, our framework does not provide any information on how the routes may impact the study results.

2.2 Comparative Wayfinding Studies and the Importance of Route as Independent Variable

This section will review comparative wayfinding studies and verify whether the route was used as an independent variable, thereby providing examples for both cases.

It is not new to consider the route itself an important variable in comparative wayfinding studies. Savino and colleagues [20] considered the potential influence of the two selected routes in their comparative wayfinding study for cyclists and, in consequence, analyzed the data for each route separately. For both routes, the authors came to the same conclusion regarding differences in route length, task load and orientation. However, the number and the type of errors committed differed. Dong et al. compared two navigation systems on three different routes [3]. In their ANOVA analysis, the route was treated as a factor. For none of the compared eye-tracking metrics, `route` yielded a significant effect. It was only significant for the metric wayfinding duration, which is expected as route lengths differed and were not normalized. Moreover, without justification, the authors do not include route as a predictor (logistic regression) when analyzing the sketch maps. Richter et al. compared consistent and inconsistent navigation instructions on eight routes in a desktop virtual environment [18]. The selected routes had a similar number of turns and a landmark at every intersection. In the analysis, the potential influence of the route was not considered. Kuo and colleagues compared four different navigation systems on four different roads in a virtual reality (VR) environment [9]. Here, the route was also stated to be used as one of four predictors (linear regression). However, this variable, as well as two further ones, were not mentioned in the analysis. Therefore, it remains unclear if the route had an effect on the results, although this expectation was made explicit. Another study conducted in VR compared three AR-based navigation interfaces on three different routes [21]. The routes were designed to have the same length, number of turning points and street crossings. To each interface, exactly one route was assigned. The route was not treated as a factor, and in the end, it is unclear whether the observed effects come from the navigation system, the route or a combination of both.

Generally speaking, only a few routes are compared within a single wayfinding study, which seems reasonable from a research economics perspective. Simulation studies, however, are a notable exception. Amores and colleagues [1] proposed a novel navigation paradigm *most recoverable path*. Their approach was tested by means of a simulation study in Quito, Paris and Melbourne in which 13500 routes per city were selected. However, they analyzed the influence of network topology on their approach but did not analyze the data on a route level. Another example of a simulation study in which a novel navigation paradigm was proposed is our previously published work [14]. We tested our approach with 100 routes in Vienna, Djibouti City and Mexico City, respectively. Differences between those cities were found, but route-wise differences were not analyzed.

Taken together, these examples give the impression that route selection is not always given sufficient relevance, even though it might have an impact on study results. There is no systematic approach, first, to show that different routes may lead to different results, and second, how to select routes for a wayfinding study congruent with the population mean of all routes. This paper aims to fill these gaps.

3 Experimental Setup

In this section, the agent-based simulation study is described in detail. We will elaborate on the experimental area and all potential routes with pre-defined features, such as route length. Furthermore, the sample sizes and both navigation approaches, namely Turn-by-Turn (TBT) and Free Choice Navigation (FCN), will be described. The study follows a between-subject design comparing two navigation systems.

3.1 Experimental Area and Potential Routes

As the experimental area, the city center (surface area 2.5 km^2) of Vienna, Austria is chosen (see Figure 1). According to the classification by Thompson et al. [22], the network layout is of type *high transit*. For this area, the raw network data were downloaded from OpenStreetMap (OSM)[1]. The intersections and their characteristics were calculated using the Intersections Framework [4], whereas street segments were extracted with a custom script. Taking these pieces together, the city center is represented as a networkx graph having 1848 nodes and 2722 edges.

For every experimental design, several decisions regarding route choice have to be made (e.g., route length, sequence of left, right and non-turns, number of decision points and experimental area). In order to reduce the search space of potential routes, we will consider only shortest path routes (see e.g., [19]) with 12 decision points [15] and a length between 550 m and 1000 m (see e.g., [17, 19]) in order to avoid trivial route length on the one hand and, on the other hand, to ensure a reasonable duration for an in-situ study (1000 m would result in a duration of 12.5 minutes based on an average walking speed of 4.85 km/h [10]). In order to find all possible routes sharing these characteristics and comprising no loops, SageMath 9.1 with its SubgraphSearch function[2] was used, as in our previous work about the route selection framework [15]. The resulting $N_r = 11373$ routes are the whole population of routes being shortest paths and matching the mentioned lengths and number of decision points and were used for the simulation, which was implemented in Python 3.6.

3.2 Sample Size

The simulation is run with four different samples sizes ($n = 15, 25, 50, 3000$ agents) following a between-subject design. The first three sample sizes can be considered realistic in wayfinding studies [21, 18, 20, 9, 6]. The largest sample size ($n = 3000$) is considered to be representative for the whole population of participants of such studies. Different sample sizes are tested in order to investigate whether the sample size impacts the results for both a single route and the whole route population. Each group navigates each of the 11373 routes.

3.3 Navigation Systems

The presented simulation approach will work for any two navigation systems, as we want to demonstrate that the comparison results may vary depending on the route choice. However, we continue our previous simulation study [14] and compare Free Choice Navigation (FCN) and Turn-by-Turn (TBT). While the particular figures will likely change for other navigation approaches, the proposed route selection process (see Section 5.2) based on the results remains unchanged.

---

[1] `https://www.openstreetmap.org`, last access February 4th, 2022

[2] `https://doc.sagemath.org/html/en/reference/graphs/sage/graphs/generic_graph_pyx.html`, last access January 30th, 2022

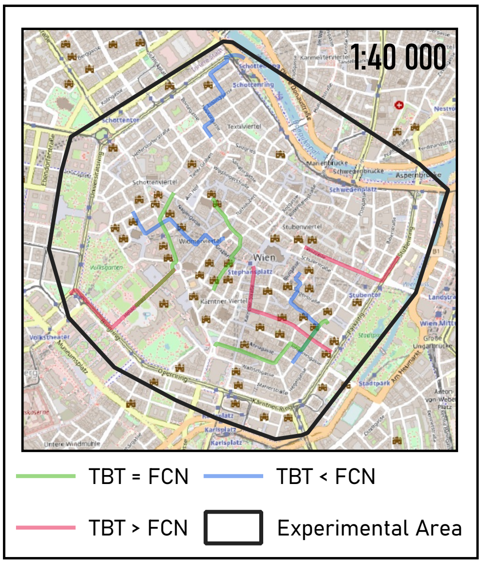

Figure 1 The experimental area in Vienna, Austria and 9 sample routes on which one navigation system performed better or they performed equally well across sample sizes. Basemap OpenStreetMap.

As the primary purpose of navigation systems is to assist wayfinders in reaching the destination, we choose arrival rate as the success metric. However, any other suitable success metric can be chosen by the researcher. As in our previous work [14], an agent is considered successful if it reaches the destination within 150% of the shortest path length. In the same work, we compared these two navigation systems in three different cities [14]. TBT lead between 5% and 10% more agents to their destination in all cities. Now, the main features and mechanics of both navigation approaches will be described.

3.3.1 Turn-by-Turn (TBT)

By analogy with commercial wayfinding assistance systems for pedestrians, the agent is supposed to follow the shortest path between origin and destination and receives navigation instructions at turning points only. If agents have to go straight ahead at a junction, then no instruction is issued and the agent continues straight.Going straight ahead is considered walking in a direction that does not deviate by more than 10 degrees to either side from the current one. Every agent has a probability to interpret a generic navigation instruction correctly. If an instruction is issued, the agent interprets it based on a weighted random choice: The branch to follow, indicated in the instruction and following the shortest path from the current junction, is assigned a weight equal to the agent's probability to interpret generic navigation instructions correctly. The remaining probability is split equally over all remaining branches (excluding the one indicated in the navigation instruction and the one the agent has come from). The trial ends when the agent reaches the destination.

3.3.2 Free Choice Navigation (FCN)

Free Choice Navigation is a novel navigation paradigm aiming for more freedom of choice during navigation [14]. The system allows the agent for some exploration but, on the other hand, tries to avoid costly mistakes by weighing the number of free choices, the number of given instructions and a maximum allowed route length. The working mechanism can be seen in the following example: Alice, a good wayfinder, is navigating to a museum. Before the navigation starts, the system gives her information about the beeline direction and distance to the museum. At the first two junctions, the system does not issue any instructions because it is assumed that the beeline direction is still clear to the user. In consequence, Alice decides on her own which branch to take. The upcoming junction, however, is rather complex as it has five branches. Alice is quite sure about the beeline direction, but there are two branches that seem equally well suited to her. The system detects this difficulty based on internal computations that take the environmental structure and spatial abilities of the user into account and issues an instruction because one of the branches leads to a considerable deviation from the allowed maximum route length. Alice interprets it correctly and continues her walk.

This example shows that the navigation system issues an instruction based on environmental spatial abilities of a user, the characteristics of the current junction and the already walked route. If an instruction is issued, then the same procedure as above applies with the difference that the branch the agent has come from is not excluded but has a lower probability of being taken. Again, the probabilities of available branches to be taken depend on the agent's probability of interpreting generic navigation instructions correctly, which in turn depends linearly on its environmental spatial abilities. Furthermore, FCN has six parameters that steer when an instruction is given. We used the best parameter set for Vienna, which is a trade-off between the percentage of successful trials and the number of given instructions [14].

For every agent, regardless of the condition, the ability to interpret navigation instructions correctly ranges between 0.8 and 1 and is fixed before the experiment. Please refer to our previous work for further modeling details regarding the agents and their decision mechanism [14].

4 Simulation Results

In this section, we, first, present descriptive statistics for each of the systems separately and, second, discuss the differences originating from different routes. Differences between both conditions are calculated using bootstrapping ($B = 10000$ runs) and 95% confidence intervals (CIs) are reported in square brackets. As mentioned above, the arrival rate (each agent walked each route) for both systems is compared (see Section 3.3). In order to ensure that the common ability of agents to interpret navigation instructions correctly (co-domain [.8; 1] [14]) did not influence the results, a Wilcoxon Signed-Rank Test was performed for every sample size. No significant ($\alpha = .05$) differences between both conditions were found ($n = 15$ ($Z = 1.14$, $p = .25$, $r = 0.29$), $n = 25$ ($Z = 1.17$, $p = .24$, $r = 0.23$), $n = 50$ ($Z = .05$, $p = .96$, $r = 0.01$), $n = 3000$ ($Z = .00$, $p = .99$, $r = 0.00$)). Furthermore, this ability defines good and weak wayfinders. We assured that agents from both groups are present in every sample size, which is a realistic scenario for real-world wayfinding studies. The presented figures are computed based on all potential routes, which were walked by all agents of a given sample size. There are 11 373 potential routes in the experimental area. This is an exhaustive sample considering the selected route features (see Section 3.1).

4.1 Turn-by-Turn

In the case of the Turn-by-Turn condition, the simulation for all four sample sizes yields similar results regarding the arrival rate (co-domain [0; 1])(see Table 1). The mean arrival rate is around 0.97 across all four sample sizes. This is in contrast to the minimum arrival

Table 1 Descriptive statistics of the arrival rate [0; 1] for the TBT condition for all four sample sizes tested in the simulation. The figures are rounded to 3 decimals.

Sample Size	Mean	SD	Median	Min	Max
15	.976 [.975; .977]	.048 [.047; .05]	1 [1;1]	.66 [.66; .66]	1 [1; 1]
25	.975 [.974; .976]	.043 [.042; .044]	1 [1;1]	.68 [.68; .72]	1 [1; 1]
50	.973[.972; .973]	.038 [.037; .039]	.98 [.98; .98]	.72 [.72; .74]	1 [1; 1]
3000	.97 [.97; .97]	.033 [.032; .034]	.982 [.982; .983]	.788 [.788; .798]	1 [1; 1]

rate, which shows considerable variation between sample sizes: For sample size $n = 15$, the minimum arrival rate for a route is 0.66 (only $\frac{2}{3}$ of the agents reached the destination), whereas for sample size $n = 3000$ it is 0.788. The range of the arrival rate decreases with increasing sample size. In order to see whether there are route-wise differences between sample sizes, the range ($max - min$) for every route is calculated (see Figure 2). Over 20% of the routes have a range greater than or equal to 0.05. There is no difference across sample sizes for 716 routes (6.2%), whereas the biggest difference encountered for a single route across sample sizes is 0.25.

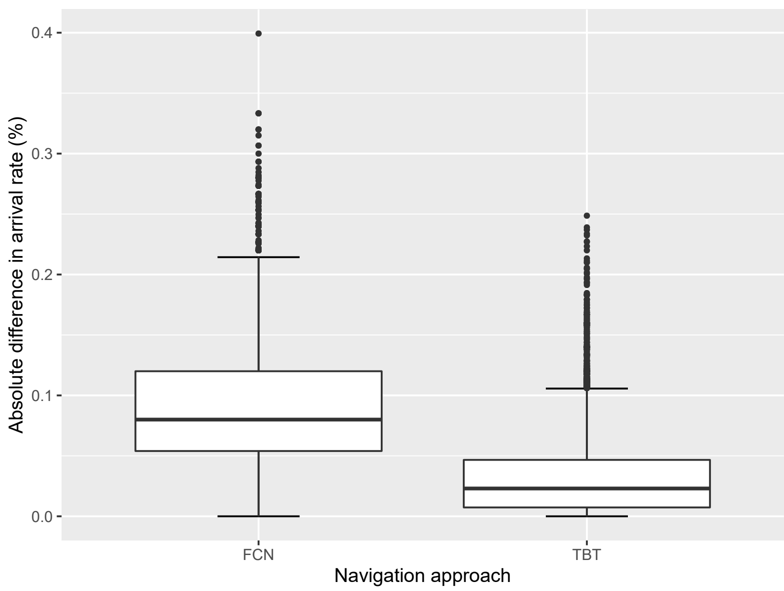

Figure 2 Route-wise ranges ($max - min$) across all sample sizes for the condition FCN (left) and TBT (right).

4.2 Free Choice Navigation

Analyzing all routes together, the four sample sizes yield, again, similar results (see Table 2). The mean arrival rate is approx. 0.90; in contrast to the TBT condition, the range remains almost identical across sample sizes. Route-wise ranges ($max - min$), however, reveal a higher variance in the FCN condition (see Figure 2): More than 80% of the routes show a difference greater than or equal to 0.05 and the biggest difference for a route across sample sizes is 0.4. For 34 routes (0.3%), there is no difference across sample sizes.

Table 2 Descriptive statistics of the arrival rate [0; 1] for the FCN condition for all four sample sizes tested in the simulation. The figures are rounded to 3 decimals.

Sample Size	Mean	SD	Median	Min	Max
15	.919 [.917; .921]	.104 [.099; .109]	.933 [.933; .933]	0 [0; 0]	1 [1; 1]
25	.908 [.906; .91]	.096 [.091; .101]	.92 [.92; .92]	0 [0; 0]	1 [1; 1]
50	.896 [.894; .897]	.091 [.086; .096]	.92 [.92; .92]	0 [0; 0]	1 [1; 1]
3000	.902 [.901; .904]	.08 [.074; .086]	.918 [.917; .919]	0 [0; 0]	.996 [.994; .996]

4.3 Differences within both Systems

In both approaches, of course, the ability to interpret navigation instructions plays a role, but as it is constant for all routes, it is not mentioned as a factor. As indicated by the figures in tables 1 and 2, arrival rates differ between both navigation systems. These differences may stem from navigation system mechanics and street layout. In the TBT condition, routes with less turning points likely lead to a higher arrival rate, as the agent has to make fewer decisions and, in consequence, has lower chances to commit an error. On the other hand, in the FCN approach, route features like junction complexity or junction skewness [4] are likely to play a role. A detailed analysis of route features leading to differences is beyond the scope of this paper (see Section 6).

4.4 Differences between both Systems

Based on the within-system results, both navigation systems will be compared regarding the arrival rate. Again, first, the whole population is analyzed, and second, route-wise differences will be inspected in order to investigate whether sample size impacts the share of routes that lead to contradicting results. Across all sample sizes, the TBT approach leads, on average, more agents to the destination (see Table 3). The sample size with the highest mean difference in arrival rate across routes is $n = 50$, whereas the lowest value can be observed for $n = 15$. Mean, standard deviation, median and maximum values are similar in all simulation runs; however, there are differences in the minimum: All minimum values are negative, meaning that there is at least one route on which the FCN approach performed better than TBT. Therefore, we will inspect per route differences between both conditions by subtracting FCN from TBT arrival rates for the respective sample size.

For every sample size, we count the number of routes which lead to a congruent result with the population mean (TBT performs better), as well as routes on which FCN performed better than or as good as TBT (see Table 4). For the sample size $n = 3000$, which is the most representative one, there are around 8% of routes on which FCN performed better or as good as TBT. For smaller sample sizes, this figure increases, reaching around 47% for $n = 15$. Contrary to the within-system results (see Sections 4.1 and 4.2), here, considerable differences between sample sizes can be observed.

5 Discussion and Limitations

This section will discuss the results, which suggest that route selection is an important part of experimental design and should be given more importance. Furthermore, a methodology that supports informed route selection is proposed. Finally, limitations that apply to our work are addressed.

Table 3 Descriptive statistics for the route-wise difference ($TBT - FCN$) in arrival rate [0; 1] for all four sample sizes tested in the simulation. Positive values mean that the TBT condition performed better and negative values indicate a better performance of the FCN navigation approach. The figures are rounded to 3 decimals.

Sample Size	Mean	SD	Median	Min	Max
15	.057 [.055; .059]	.111 [.107; .116]	.067 [.067; .067]	- .333 [- .333; - .333]	1 [1; 1]
25	.067 [.065; .069]	.101 [.096; .105]	.04 [.04; .04]	- .28 [- .28; - .24]	1 [1; 1]
50	.077 [.075; .079]	.094 [.089; .098]	.06 [.06; .06]	- .24 [- .24; - .18]	1 [1; 1]
3000	.067 [.066; .069]	.081 [.075; .086]	.057 [.056; .058]	- .119 [- .119; - .104]	1 [.983; 1]

Table 4 Shares of routes on which TBT performed better than, as good as and worse than FCN regarding the arrival rate. The figures are rounded to 1 decimal.

Sample Size	TBT Better	TBT = FCN	FCN Better
15	53 %	35.6 %	11.5 %
25	70.4 %	19.3 %	10.3 %
50	84 %	7.5 %	8.5 %
3000	92.1 %	0.1 %	7.8 %

5.1 Discussion

Regardless of the sample size, the simulation, which considers all potential routes in the experimental area, yields similar results, indicating the superiority of TBT over FCN regarding the arrival rate (see Table 3). Looking at the results for the whole population, one might think that route selection is not so critical because, independently of the sample size, the big picture is preserved. This picture is, however, somewhat misleading as wayfinding studies, of course, are conducted with a small-sized subsample of the whole population, considering both routes and participants. By means of keeping the population of routes constant across sample sizes, our simulation results indicate that different routes can lead to contradicting results (see Table 4). In consequence, ad-hoc decisions on route selection can lead to contrary results compared to the whole population of routes. This situation worsens with decreasing sample size as the chance of selecting such a route increases (see Table 4). The results, therefore, suggest that selecting a route is all the more important in the case of small numbers of participants. For samples sizes (n = 15, 25, 50), which can be considered realistic for comparative wayfinding experiments (see e.g. [21, 18, 20, 9, 6]), the probability to select a route that will yield results incongruent with the population mean varies between 16% and 47%. This means, if we planned an experiment with two groups, with 15 participants each, and we randomly picked a route from our experimental area, we would have a 47% chance to conclude that TBT is not superior regarding the arrival rate, although it actually is (see Table 3). Almost every second route would lead to the contrary conclusion in the case of $n = 15$, whereas, for sample sizes $n = 25$ and $n = 50$, it would be every third and sixth route, respectively. Given this high share, we want to draw attention to the importance of the route selection process as it can influence study results, in particular, given the relatively small number of participants, which is quite common in the wayfinding domain.

Taken together, in the selected experimental area, the lower the number of agents, the higher the probability of choosing a route which leads to results that are contradictory to the population mean, i.e., the route becomes more crucial with decreasing sample size. This is a problem as routes often seem to be selected in an ad-hoc manner during experimental design in wayfinding studies (see Section 2.1). Given that wayfinding studies are not conducted with 3000 participants, a method to select those routes which are likely to lead to a conclusion corresponding to the whole population is proposed.

5.2 Route Selection

Depending on the sample size, the chance of selecting a route that leads to conclusions that are not in line with the whole route population may be considerable. This section suggests an approach that allows for an informed route selection. This process is based on the simulation results, i.e., the simulation for the compared navigation systems needs to be run beforehand. It is a two-step approach. First, routes without great variance across all sample sizes are selected, and second, those which lead to results congruent with the population mean (based on the simulation) are chosen. In both steps, the researcher needs to select a filtering threshold depending on the selected performance metric and observed differences. The underlying idea is to select routes that lead to similar results across all sample sizes and are compatible with the population mean. In our example, two navigation systems are compared. Therefore, their differences in arrival rate are used in the presented filtering process. The same approach can be applied with one navigation system only by using the arrival rates directly instead of the differences or any other success metric chosen by the researcher.

5.2.1 Consistent Routes

In this step, routes will be selected which are *consistent* regarding differences in arrival rates, i.e., they do not vary considerably in arrival rates across sample sizes. Given that there are four values (one per sample size) for each route to consider, we refrain from calculating the standard deviation and will consider the range as the measure of variability. The applied measure with a corresponding threshold can be adapted according to the number of tested sample sizes and the researcher's needs. For every route, we calculate the range across all four sample sizes and select those routes whose range is not greater than 0.03, which means that the biggest allowed difference across sample sizes is 3%. This value can be set according to the simulated data. The smaller the value, the more restrictive this filtering step will be. In this case study, 618 (5.4%) routes have a range smaller than or equal to 0.03. By this filtering step, routes with high variance across sample sizes are excluded. However, routes that are not *close* to the population mean are still possible, or even routes on which the drawn conclusion is contrary to the population mean. Therefore, a second filtering step is necessary.

5.2.2 Routes in Concordance with the Population Mean

In order to find routes that are congruent with the most representative sample size (n = 3000), they are filtered by their mean across sample sizes. Routes whose means do not differ considerably from the population mean (see Table 3) are selected for being considered suitable routes. For this step, another threshold needs to be selected by the researcher. In consequence, routes are selected whose means do not deviate by more than the selected threshold from the population mean. Given that the population mean difference is 0.067, we set this threshold to 0.02. Therefore, routes with an average between 0.047 and 0.087 are considered in our case study as acceptable. With this second filtering step, 304 routes are left. This is 2.67% of the whole population. For this proof of concept, the exact threshold values are of less relevance. The smaller both thresholds are set, the more restrictive the filtering process is, i.e., less routes are considered suitable. This has to be decided based on the simulation results at hand and the researcher's needs.

5.3 Route Ranking

Our approach delivers a list of suitable routes with regard to the most representative sample size but does not state explicitly which one to choose. Our approach can, however, be combined with our route selection framework [15]. In doing so, potential route biases can be further mitigated. First, the routes are ranked according to features selected by the researcher [15], e.g., mean segment length, traffic, average number of branches or number of left, right and non-turns along the route. Second, the ranked routes are filtered according to our proposed approach. This results in routes that satisfy both the researcher's needs regarding route characteristics and being close to the global mean across sample sizes.

5.4 Limitations

Running a simulation implies simplifying certain aspects of the real world. In our simulation, the street network and the agent's spatial abilities are used to model the agent's behavior. Compared to our previous work [14], the agents, their reasoning mechanism and the environment could have been adapted regarding complexity (see e.g., [13, 8]). In addition to that, there may be relationships that have not been yet discovered and, therefore, are not considered in the simulation process. Given that randomness plays a role in our simulation, running the simulation once is a limitation. However, several seeds were tested with a subset of routes during a pretest and the results were quite consistent.

6 Conclusion and Future Work

By means of an agent-based simulation study, which was run on all potential routes in a selected experimental area, it was shown that depending on the route selection, the study results can be contradictory. Although the results for the whole population lead, on average, to the same conclusion, there is an important share of routes that lead to contrary results. Given that the route selection process usually does not receive much attention in wayfinding studies, with this simulation, we direct researchers' attention to the potentially harmful effects of ad-hoc route selections. Therefore, we propose a selection method based on running the same simulation with different sample sizes. The resulting selection of routes should lead to results that are congruent with the population mean.

Furthermore, our proposed simulation approach with different sample sizes allows for detecting weak points of a given navigation system. Researchers will find routes on which their proposed navigation system does not perform as good as expected and their examinations will lead to further improvements. Moreover, our simulation approach makes it possible to identify spatial configurations (routes and their neighborhood) favorable or adverse to the navigation system at hand by analyzing route features that cause differences in the selected performance metric. This analysis would provide valuable feedback in order to improve the tested navigation approach. The in-depth analysis of route properties and their influence on the success metric is part of our future work. One could improve the navigation system until it is robust on all routes, i.e., it performs equally well on the whole population of routes.

A series of simulation studies in different geographic areas is planned in order to see whether different network types [22] lead to the same results. *Motor cities* might be less vulnerable to ad-hoc route selection. In addition, the route properties which caused differences in arrival rates will be examined in depth. Moreover, we plan as well to increase the complexity of the models to increase the validity of the simulation. Adding more complexity would expand the search space for possible explanations because the differences in the selected success metric could be explained by additionally modeled features like points of interest, buildings or terrain slope.

We are aware that implementing a simulation is not a trivial task and not every researcher has the resources to do it. Therefore, another research direction could be the prediction of route suitability based on route features and the characteristics of the navigation approach without running a simulation study.

Our approach still needs to be verified in real-world environments. Therefore, a series of human subject experiments will be conducted. In these experiments, the results of several routes selected with our approach will be compared with the population mean resulting from a simulation study. Following our selection approach, we expect that those routes considered suitable will lead to consistent and congruent with the population mean results and the routes considered non-suitable will more likely lead to contrary conclusions. However, this hypothesis needs to be verified in a real-world setting as the routes are selected based on simulation results.

References

1 David Amores, Egemen Tanin, and Maria Vasardani. A proactive route planning approach to navigation errors. *International Journal of Geographical Information Science*, 35(6):1094–1130, 2021. doi:10.1080/13658816.2020.1820508.

2 Eran Ben-Elia. An exploratory real-world wayfinding experiment: A comparison of drivers' spatial learning with a paper map vs. turn-by-turn audiovisual route guidance. *Transportation Research Interdisciplinary Perspectives*, 9:100280, 2021. doi:10.1016/j.trip.2020.100280.

3 Weihua Dong, Yulin Wu, Tong Qin, Xinran Bian, Yan Zhao, Yanrou He, Yawei Xu, and Cheng Yu. What is the difference between augmented reality and 2d navigation electronic maps in pedestrian wayfinding? *Cartography and Geographic Information Science*, 48(3):225–240, 2021. doi:10.1080/15230406.2021.1871646.

4 Paolo Fogliaroni, Dominik Bucher, Nikola Jankovic, and Ioannis Giannopoulos. Intersections of Our World. In Stephan Winter, Amy Griffin, and Monika Sester, editors, *10th Int. Conf. on Geographic Information Science (GIScience 2018)*, volume 114 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.GISCIENCE.2018.3.

5 Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. Gazenav: Gaze-based pedestrian navigation. In *Proc of the 17th Int Conf on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '15, pages 337–346. ACM, 2015.

6 Antonia Golab, Markus Kattenbeck, Georgios Sarlas, and Ioannis Giannopoulos. It's also about timing! when do pedestrians want to receive navigation instructions. *Spatial Cognition & Computation*, 0(0):1–33, 2021. doi:10.1080/13875868.2021.1942474.

7 Haosheng Huang, Thomas Mathis, and Robert Weibel. Choose your own route – supporting pedestrian navigation without restricting the user to a predefined route. *Cartography and Geographic Information Science*, 0(0):1–20, 2021. doi:10.1080/15230406.2021.1983731.

8 Peter M. Kielar, Daniel H. Biedermann, Angelika Kneidl, and André Borrmann. A unified pedestrian routing model for graph-based wayfinding built on cognitive principles. *Transportmetrica A: Transport Science*, 14(5-6):406–432, 2018. doi:10.1080/23249935.2017.1309472.

9 Ting-Yu Kuo, Hung-Kuo Chu, and Yung-Ju Chang. Comparing the effects of reference-based, orientation-based, and turn-by-turn navigation guidance on users' independent navigation. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*, UbiComp-ISWC '20, pages 63–66, New York, NY, USA, 2020. Association for Computing Machinery. `doi:10.1145/3410530.3414424`.

10 R. V Levine and A. Norenzayan. The pace of life in 31 countries. *J.of Cross-Cultural Psychology*, 30(2):178–205, 1999.

11 Xiangdong Ma, Mengting Jia, Zhicong Hong, Alex Pak Ki Kwok, and Mian Yan. Does augmented-reality head-up display help? a preliminary study on driving performance through a vr-simulated eye movement analysis. *IEEE Access*, 9:129951–129964, 2021. `doi:10.1109/ACCESS.2021.3112240`.

12 Charlotte Magnusson, Kirsten Rassmus-Gröhn, and Delphine Szymczak. Navigation by pointing to GPS locations. *Personal and Ubiquitous Computing*, 16(8):959–971, 2012.

13 Tsubasa Maruyama, Satoshi Kanai, Hiroaki Date, and Mitsunori Tada. Simulation-based evaluation of ease of wayfinding using digital human and as-is environment models. *ISPRS International Journal of Geo-Information*, 6(9), 2017. `doi:10.3390/ijgi6090267`.

14 Bartosz Mazurkiewicz, Markus Kattenbeck, and Ioannis Giannopoulos. Navigating Your Way! Increasing the Freedom of Choice During Wayfinding. In K. Janowicz and J. Verstegen, editors, *11th Int. Conf. on Geographic Information Science (GIScience 2021) - Part II*, volume 208 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 9:1–9:16. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. `doi:10.4230/LIPIcs.GIScience.2021.II.9`.

15 Bartosz Mazurkiewicz, Markus Kattenbeck, Peter Kiefer, and Ioannis Giannopoulos. Not Arbitrary, Systematic! Average-Based Route Selection for Navigation Experiments. In K. Janowicz and J. Verstegen, editors, *11th Int. Conf. on Geographic Information Science (GIScience 2021) - Part I*, volume 177 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 8:1–8:16. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020. `doi:10.4230/LIPIcs.GIScience.2021.I.8`.

16 Martin Pielot, Benjamin Poppinga, Wilko Heuten, and Susanne Boll. Tacticycle: Supporting exploratory bicycle trips. In *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '12, pages 369–378, New York, NY, USA, 2012. Association for Computing Machinery. `doi:10.1145/2371574.2371631`.

17 Karl Rehrl, Elisabeth Häusler, Sven Leitinger, and Daniel Bell. Pedestrian navigation with augmented reality, voice and digital map: final results from an in situ field study assessing performance and user experience. *Journal of Location Based Services*, 8(2):75–96, 2014. `doi:10.1080/17489725.2014.946975`.

18 Kai-Florian Richter, Róisín Devlin, and Filippo La Greca. Investigating wayfinding under inconsistent information. In Jurģis Šķilters, Nora S. Newcombe, and David Uttal, editors, *Spatial Cognition XII*, pages 191–195, Cham, 2020. Springer International Publishing.

19 S. Robinson, M. Jones, P. Eslambolchilar, R. Murray-Smith, and M. Lindborg. "I Did It My Way": Moving Away from the Tyranny of Turn-by-Turn Pedestrian Navigation. In *Proc. of MobileHCI '10*, pages 341–344, 2010.

20 Gian-Luca Savino, Laura Meyer, Eve Emily Sophie Schade, Thora Tenbrink, and Johannes Schöning. Point me in the right direction: Understanding user behaviour with as-the-crow-flies navigation. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 1–11, 2020.

21 Tram Thi Minh Tran and Callum Parker. Designing exocentric pedestrian navigation for ar head mounted displays. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pages 1–8, New York, NY, USA, 2020. Association for Computing Machinery. `doi:10.1145/3334480.3382868`.

22 J. Thompson, M. Stevenson, and J. S. Wijnands et al. (8). A global analysis of urban design types and road transport injury: an image processing study. *The Lancet Planetary Health*, 4(1):e32–e42, 2020.