Wayfinding Stages: The Role of Familiarity, Gaze Events, and Visual Attention

Negar Alinaghi
Geoinformation, TU Wien, Austria

Ioannis Giannopoulos
Geoinformation, TU Wien, Austria

Abstract

Understanding the cognitive processes involved in wayfinding is crucial for both theoretical advances and practical applications in navigation systems development. This study explores how gaze behavior and visual attention contribute to our understanding of cognitive states during wayfinding. Based on the model proposed by Downs and Stea, which segments wayfinding into four distinct stages: self-localization, route planning, monitoring, and goal recognition, we conducted an outdoor wayfinding experiment with 56 participants. Given the significant role of spatial familiarity in wayfinding behavior, each participant navigated six different routes in both familiar and unfamiliar environments, with their eye movements being recorded. We provide a detailed examination of participants' gaze behavior and the actual objects of focus. Our findings reveal distinct gaze behavior patterns and visual attention, differentiating wayfinding stages while emphasizing the impact of spatial familiarity. This examination of visual engagement during wayfinding explains adaptive cognitive processes, demonstrating how familiarity influences navigation strategies. The results enhance our theoretical understanding of wayfinding and offer practical insights for developing navigation aids capable of predicting different wayfinding stages.

2012 ACM Subject Classification Applied computing → Psychology; General and reference → Empirical studies; Computing methodologies → Interest point and salient region detections; Computing methodologies → Video segmentation

Keywords and phrases Eye-tracking, Wayfinding, Spatial Familiarity, Visual Attention, Gaze Behavior

Digital Object Identifier 10.4230/LIPIcs.COSIT.2024.1

Supplementary Material *Dataset*: `https://geoinfo.geo.tuwien.ac.at/resources/`

1 Introduction

Navigation, a fundamental human cognitive skill, involves goal-oriented movement through space and is comprised of two components: wayfinding and locomotion [35]. Wayfinding in particular requires strategic planning and decision-making to reach the destination. Understanding the cognitive resources underlying this component is important to better grasp how people perceive their environment and interact with it (see [14] for a review). This knowledge is also essential for the development of tools and technologies to aid navigation in today's increasingly complex environments.

Numerous theories have explored the cognitive processes behind wayfinding [7, 12, 32, 38], contributing significantly to our understanding of its cognitive dimensions. However, empirical validation remains necessary, highlighting the importance of testing these theories in practice. Among these proposed theories, the model by Downs and Stea [12] has become a key part of wayfinding research, providing a detailed breakdown of the process into four stages: Orientation (also referred to as Self-Localization), Route Selection (or Route Planning), Route Control (or Monitoring), and Recognition (or Destination/Goal Recognition). Its

16th International Conference on Spatial Information Theory (COSIT 2024).
Editors: Benjamin Adams, Amy Griffin, Simon Scheider, and Grant McKenzie; Article No. 1; pp. 1:1–1:21
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

fine-grained distinction between the processes involved in wayfinding not only advances our theoretical understanding but also enhances the potential for practical application by providing a structured approach to empirically investigate these underlying processes.

Empirical research on wayfinding has evolved considerably, especially with the advent of eye-tracking technology. These studies focus on a variety of tasks related to wayfinding (e.g. self-localization, direction finding, searching for an object on the map or in the environment, etc. [10, 11]) and investigate eye movements to decode cognitive processes during navigation. Recent developments in deep learning have further improved the analysis of eye-tracking data and provide a more detailed understanding of what and how people focus visually during wayfinding [23]. While these studies have made important contributions by focusing on specific aspects of wayfinding, there remains room for comprehensive empirical research that spans the entire wayfinding stages.

In this work, we report on the behavioral data from 56 participants who performed wayfinding tasks in a real urban environment based on Downs and Stea's four-stages model. The focus lies in analyzing participants' eye-tracking recordings during wayfinding in familiar and unfamiliar areas. Using a deep learning-based software system [3], the eye-tracking data was automatically annotated, enabling us to examine the participants' visual attention and gaze behavior in a content-dependent manner, particularly in relation to their familiarity with the environment. We primarily investigated how participants' gaze behavior, focusing on patterns of gaze movements irrespective of the semantics of observed objects, and visual attention, concerning the actual objects observed by the participants, evolve across the four stages of wayfinding. Additionally, considering the known effects of familiarity on wayfinding and gaze behavior (see e.g., [1, 4, 30, 50, 54]), we investigated the influence of familiarity with the environment to see how these visual and cognitive processes differ between the familiar and unfamiliar groups during wayfinding.

The results of our study show that distinct gaze patterns emerge at different stages of wayfinding, some of which are significantly influenced by how familiar individuals are with the environment. These results improve our understanding of the visual cognitive mechanisms that are crucial for wayfinding. By shedding light on the stages of wayfinding from the perspective of visual cognition, this study expands our understanding of the dynamic interplay between vision and cognitive processes and the way we approach wayfinding tasks.

2 Related Work

In this section, we explore the existing body of literature from three aspects: initially, we review the theoretical foundations underlying wayfinding. Following that, we move on to relevant empirical research aimed at deepening our understanding of wayfinding behaviors and the factors affecting them, in particular familiarity. Lastly, we examine studies that utilize eye-tracking and its measures to explain different aspects of cognitive processing.

2.1 Theories of Wayfinding

Exploring wayfinding from a theoretical aspect has been approached from various disciplines, each contributing to our understanding of how humans navigate in and interpret their environments. Lynch [32] was among the first to suggest that the way we see and remember our surroundings helps us navigate. He argued that these mental images – comprised of paths, landmarks, nodes, districts, and edges – serve not just as a practical tool for navigation but also form a deep emotional connection between individuals and their environments. This foundational idea suggested that the way we perceive and remember our surroundings plays a critical role in how we navigate, combining sensory inputs with memories to create a mental map that guides us through the physical world.

Building on Lynch's conceptual framework, Downs and Stea [12] later introduced a more structured approach to wayfinding, breaking it down into four measurable stages: *Orientation/ self-localization* (figuring out where you are), *Route Selection/ Route Planning* (choosing a path to the destination), *Route Control/ monitoring* (keeping track of the journey), and *Recognition/ Goal Recognition* (recognizing the destination). Their model also emphasized the role of environmental cues and personal experiences in shaping mental maps, offering insights into the cognitive processes that undergo spatial navigation. This approach provided a clear framework for understanding wayfinding, highlighting how individuals interact with their environments by processing information, making decisions, and adjusting their routes based on ongoing feedback from their surroundings.

Researchers like Passini [38] and Golledge [20] expanded the discussion to include the problem-solving nature of wayfinding and the role of cognitive maps – mental representations of spatial environments that guide us in unfamiliar settings. These developments have enriched our understanding of wayfinding, depicting it as a dynamic interaction between cognitive processes, memory, and the physical act of moving through spaces, ultimately splitting wayfinding into three processes: *decision-making*, *decision execution*, and *information processing*. The discussion on wayfinding took a further step with the introduction of Klippel's concept of wayfinding choremes [29] – primitive conceptual elements that form the basis of spatial understanding and map construction. This idea relates cognitive processes with the physical representation of space, focusing on how individuals conceptualize and navigate through environments by identifying structural and functional aspects of the landscape. From the psychological perspective, Montello [36] defines wayfinding and locomotion as two main components of navigation. Wayfinding involves knowing where to go and how to get there, requiring goal orientation and decision-making, and is largely coordinated beyond our immediate sensory reach. He explores the psychological, environmental, and technological factors impacting wayfinding, including orientation, attention, environmental differentiation, visual access, layout complexity, and the design of information displays in navigation systems.

Among the numerous insightful theories, Downs and Stea's theory [12] stands out not only as a widely recognized ground in wayfinding research but also as it provides a detailed, step-by-step breakdown of the wayfinding process. Despite these stages being ongoing and overlapping throughout the wayfinding process, their distinct separation makes the model particularly valuable for empirical testing. Its structured approach allows researchers to examine each aspect of wayfinding in detail, making this model particularly suited for empirical investigation.

2.2 Analyzing Wayfinding: Empirical Insights and Influencing Factors

Numerous empirical studies have explored various aspects of wayfinding. One example is Kiefer et al. [28]'s investigation into the self-localization process. They employed mobile eye-tracking technology to analyze how individuals identify their location on a map by aligning visible landmarks with map symbols in an urban environment. The results indicated that effective self-localization is associated with increased visual focus on relevant map symbols and an increased shift of attention between these symbols and their actual environmental counterparts. To better understand what information wayfinders need to orient themselves using Augmented Reality (AR) navigation tools, Yang et al. [53] investigated how different combinations of AR information affect pedestrian self-localization efficiency and success rates in augmented environments. Using eye-tracking technology, the study found that among AR navigation aids, directional guidance is the most comprehensible and landmarks significantly improve safety, while road names offer little benefit and may worsen cognitive load.

Dalton [9] reported on an experiment recording route choice decisions at road junctions, illustrating that such decisions can be quantified as the sum of choices made throughout a journey. Statistical analysis suggested a tendency among participants to maintain linearity in their routes while minimizing angular differences between route choices and the perceived direction of the wayfinding goal. Wiener et al. [51] presented three experiments to explore how people navigate and plan routes in a virtual setting. Their research demonstrated that when planning routes, humans favor connections between regions rather than just connections between individual places, take into account the location of several targets, and evaluate the complexity of different potential routes.

Recently, Hegarty et al. [21] explored the different strategies humans and animals use for wayfinding in familiar environments, distinguishing between routine routes (response strategy) and innovative routes derived from the use of spatial knowledge of the environment's layout (place strategy). Utilizing Marr's multilevel theoretical model [34] – which includes the computational, representational/algorithmic, and implementation levels – they examined the process through which intelligent beings manage complex navigational tasks. They highlighted individual, sex, and age differences in strategy preference, with men and older adults showing tendencies towards place and response strategies, respectively.

Several studies have focused on the classification of familiarity, utilizing behavioral correlates to distinguish familiar from unfamiliar settings. Alinaghi et al. [1] used eye-tracking to predict, through different machine learning models, the turn decisions of wayfinders at intersections by analyzing gaze and environmental features as well as participants' familiarity with the surroundings. Their research also distinguishes between the gaze behavior of familiar and unfamiliar navigators and finds that familiar individuals exhibit more confident behavior at no-turn intersections, while unfamiliar individuals exhibit more searching behavior [4]. Savage et al. [45] employed Bayesian techniques to assess place familiarity by integrating Foursquare visit data with Facebook profiles and GPS trajectories. Liao et al. [30] applied a Random Forest approach to classify familiarity through gaze behavior in real-world navigation, with data from 38 participants walking routes in familiar and unfamiliar environments.

While these studies offer valuable insights into various aspects of wayfinding, there is a noticeable gap in the literature when it comes to a thorough exploration of all wayfinding stages. Our work aims to complement these studies by providing a detailed examination of the complete wayfinding process, an area not fully covered in existing research.

2.3 Visual Attention and Cognitive Processes

When analyzing the four stages of wayfinding through gaze behavior and visual attention analysis, it becomes crucial to decode the specific information conveyed by each gaze pattern and metric. This understanding helps us to interpret the gaze behavior observed in our study. Therefore, in this subsection, we gather findings from fields such as psychology, neuroscience, and eye-tracking research to understand the meaning behind different gaze patterns.

Gaze events such as fixations and saccades are widely recognized for their reflection of cognitive processes, scene comprehension, and visual search (see Chapters 11 to 13 of [25]). Key metrics related to fixation include duration or dwell time (the period when the eye remains relatively still), frequency (the number of fixations divided by time), and dispersion (the maximum horizontal and vertical distance covered by the gaze positions in a fixation as defined by [44]), offering insights into cognitive engagement and scene perception. Irwin et al. [26] provided a comprehensive review of significant contributions to the field, illustrating how fixation distribution and duration play critical roles in the perception and representation of real-world scenes and sentence production. Schwedes et al. [47] further demonstrate the influence of memory on fixation duration and rate, especially in the context of early memory effects where familiar stimuli attract longer second fixations than unfamiliar ones. Nakayama et al. [37] looked into the impact of task difficulty and eye-movement frequency on oculomotor performance, using indicators like pupil size, blink rate, and eye movements. They suggested that oculomotor responses are finely tuned to the complexity of tasks, revealing the sensitivity of gaze behavior to cognitive demands.

The main saccadic features are saccade duration (the time taken to move between two fixations), saccade amplitude (the distance traveled by a saccade), saccade rate (the number of saccades per second), and saccade direction or orientation (the direction of the saccadic movement). In exploring these features, Foulsham et al. [15] reported a preference for horizontal over vertical movements, highlighting the influence of visual stimuli on saccadic direction. Further investigations into cognitive states through saccadic behavior, such as those by Xin et al. [52] in healthcare training, Vrij et al. [49] in deception, and Schleicher et al. [46] in sleepiness, enrich our understanding of how gaze patterns, in particular saccadic behavior, correlate with cognitive processes and psychological states. These studies collectively emphasize the multifaceted nature of gaze behavior as a window into human cognition, task difficulty, and psychological conditions, underscoring the value of integrating gaze metrics into the broader landscape of behavioral research.

3 Wayfinding Stages: Analysis of Gaze and Visual Attention

This section first explains the methods used for data collection and the initial processing techniques used to prepare the data[1] for our in-depth analysis. Between August 2021 and May 2023, we conducted extensive data collection in Vienna to study different aspects of human outdoor wayfinding behavior, particularly concentrating on the four stages of wayfinding introduced by Downs and Stea [12]. Our focus was on understanding the emergence and interrelationships between these stages and analyzing the transitions between them. Following the two-step data collection (online and on-site), we thoroughly analyzed the collected data, relying on relevant literature and pursuing the core objective of our study: to investigate patterns of visual attention and gaze behavior throughout the wayfinding process and to examine the effects of spatial familiarity on these patterns. This section continues by providing a detailed overview of the analytical strategies employed to extract key features of the two main aspects of visual engagement: gaze behavior, i.e., fixation and saccade patterns independent of the content or semantics of the observed objects, and visual attention, i.e., examination of the content or semantics of the observed objects.

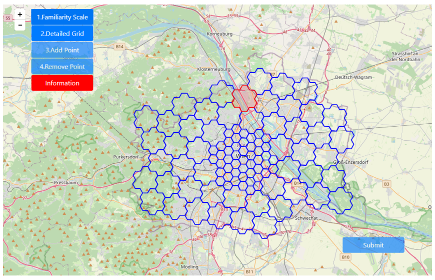
3.1 Data Collection and Preparation

Participant recruitment was conducted through a snowball sampling method, initiating with personal contacts and extending the invitation through their networks. In the online step, 84 individuals registered for the study, with 67 participants attending the on-site part. However, due to sensor failures, data from 11 participants were lost. Ultimately,

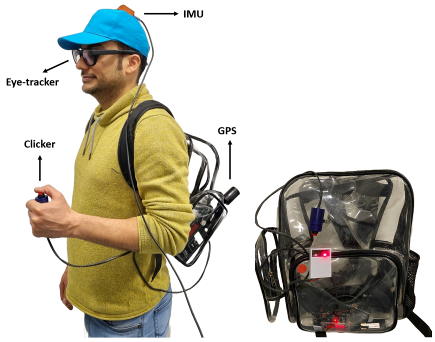
[1] The processed data can be accessed via `https://geoinfo.geo.tuwien.ac.at/resources/`

56 individuals ($mean_{age}$ = 31.1 years, std_{age} = 5.9), consisting of 22 females and 34 males, successfully completed both study phases and form the study population of this paper. After the experiment, a 400 EUR-prize lottery was conducted as an appreciation for participation.

The **online step** involved collecting demographic information, preferences for navigation aids, responses to the Santa Barbara Sense of Direction (SBSOD) questionnaire [22], and the BigFive personality test [39], all submitted via an online registration form. Participants also rated their familiarity with various areas in the city on a 5-point scale, ranging from 1 (completely unfamiliar) to 5 (fully familiar, denoting areas of residence, work, or frequent visitation in which they can easily find their way relying on their spatial abilities). These areas were defined by two levels of hexagonal tessellations (see Figure 1a): inner city parts with cells spanning 1 km^2 and outer parts with cells spanning 7 km^2 (with the option of being subdivided into seven smaller cells). Participants assigned a familiarity score to each cell. This dual-level approach aimed to ensure a detailed familiarity assessment of the inner city and increase the chances of finding both familiar and unfamiliar areas for all participants within this city section, avoiding the impact of diverse urban designs. Despite ratings being collected on a 5-point scale, familiarity was simplified to a binary measure for analysis in this paper.



(a) The tessellation design allowed participants to assess their familiarity with the study city. It featured larger cells dividable into smaller ones, ensuring a consistent detailed rating for inner city areas while allowing personalized refinement for outer areas.



(b) Participant equipped with the sensors for the study, including an eye-tracker, an IMU attached to the cap, a GPS receiver on the back-pack, and a handheld clicker connected to an LED light attached to the backpack.

Figure 1 Spatial familiarity assessment as part of the online step (1a), and sensor equipment (1b) for the on-site step of the study.

For the **on-site step**, participants were randomly assigned to start from either the most familiar or the most unfamiliar location to overcome the learning effect. These locations were determined based on the participants' ratings from the online step: The cell with the highest rating was selected as the familiar location, while the cell with the lowest rating was chosen as the unfamiliar location. If multiple cells shared the same highest or lowest rating, selection prioritized environmental and urban similarities, using the district as a reference based on [42] which suggests that considering the building age districts 1, 3, 4, 5, 6, 7, 8, 9 in Vienna are similar. During the on-site step, participants were equipped with several sensors, including PupilLabs Invisible eye-tracking glasses set, recording gaze movement at 200 Hz (see Figure 1b that shows the complete setup of the sensors). They also received a customized Google Maps application on an 11-inch $2560 * 1600$ LCD display tablet, which disabled real-time location tracking and search functionality and included a route-drawing tool. The map defaulted to a zoomed-in view of the participant's current cell. All participants had a short training session on how to use this map.

Participants were assigned various tasks: **First**, they were instructed to identify and mark their current location on the map which was zoomed by default to the selected cell where the person was located. This task served as the ground truth for the self-localization stage. **Second**, they were shown a destination within a 7-10 minute walking distance on the map and asked to draw a preferred route, establishing the basis for the route planning stage. The destinations were pre-selected from local points of interest with the Open Street Maps (OSM) amenity tag of sustenance, which includes bars, cafes, restaurants, etc. The selection of origin-destination pairs ensured that the connecting route would include at least two turning junctions, avoiding straight routes. This choice aligned with findings from [4], highlighting varied wayfinding behaviors at turning versus non-turning intersections. **Third**, participants proceeded toward the destination, exhibiting their natural walking behavior, with this segment serving as ground truth for the monitoring stage. Participants were informed that map use was not restricted during this and all other stages. **The final task** involved pressing a button (Figure 1b) upon first sighting and recognizing the destination as such, marking the goal recognition stage. The on- and off-sets of all tasks were recorded by the experimenter. Reaching the destination would end the first trial. To account for learning effects, a cognitive reset technique was used before the tasks on the second and third routes began. Participants were asked to read a written text aloud while being guided by the experimenter to a new location within 2-3 minutes of walking. This technique served as a cognitive reset before participants were led to a new location to start a second route and repeat the same tasks. Using this technique, we were able to run three routes per participant in both locations, resulting in six routes per participant (in some cases, one route was omitted due to experimental issues). Once data collection was complete, the gaze recordings, including scene videos and gaze positions, were extracted for each route and the on- and off-set of each stage was marked on them.

According to Downs and Stea [12], the four stages mentioned do not necessarily occur sequentially; rather, they can recur, and the transitions between them remain unclear. However, from an empirical perspective, the collected ground truth data for each stage, while not capturing every instance of each stage (as, for example, route planning and self-localization can occur mentally without physical representation during the whole navigation task), represents the cleanest manifestation of that stage in the context of the experiment. Additionally, what we define as monitoring, starting after route planning and continuing until the individual sees and recognizes the destination, may include other stages as well. Given these considerations, for the purposes of this paper's analysis, we regard the collected ground truth as the available and measurable representative of these stages.

3.2 Content-Independent Analysis: Gaze Behavior

Measures such as fixations and saccades can be extracted independently of the visual content. These content-independent measures are crucial in analyzing gaze behavior. To compute these, we implemented the Identification by Dispersion-Threshold (IDT) detection algorithm by [44]. The parameters, i.e., the gaze-dispersion threshold and the temporal threshold were set to $0.02deg$ and $100ms$ respectively, aligning with recommendations from [17, 18].

Considering the mobile aspect of our study, which allowed participants to move their heads and bodies freely while walking, it was necessary to adjust the saccade calculations for these movements, as the literature reports that saccades are significantly influenced by head movements [13, 2]. We adopted the method proposed by [2], which uses image processing to estimate rotation angles from head movements. In this method, video frames containing fixations that form a saccade are stitched together, applying the Random Sample

Consensus (RANSAC) algorithm to estimate a homography matrix for the stitching process. This matrix, which replicates head movements, is applied to each fixation in the saccade calculation, resulting in corrected fixation coordinates in the stitched frame's reference system. Saccades are then recalculated with these adjusted coordinates. This technique, though computationally demanding, is essential for accurate measurements. Once accurate saccade and fixation data were obtained, four metrics - the duration and frequency of both fixations and saccades - were computed separately for each trial and each wayfinding stage.

3.3 Content-Dependent Analysis: Visual Attention

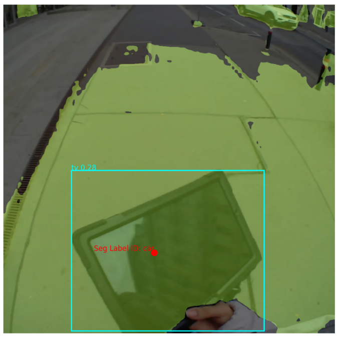
Content-dependent measures, which consider visual content or Areas of Interest (AOIs), provide crucial insights into the semantics of objects viewed by people. These measures enable the definition and detection of events like AOI hits and transitions, transforming recorded data into formats like strings, transition matrices, and time-based proportion graphs [25]. The simplest measure, a *hit* identifies the fixated object, such as "car" if the object fixated upon is a car. *Transitions*, or *gaze shifts* from one object to another (e.g., "car" to "building"), are often presented as matrices (if the from-to order matters), or as a count of changes in hit (if the order is irrelevant), indicating, for example, the number of transitions from "car" to "building" and vice versa. Calculating all these measures requires fixation annotations with semantic information, a very labor-intensive task often done manually, thus impractical for large-scale studies like ours.

We utilized a deep learning-based method, called MYFix, proposed by [3], for automatic fixation annotation. This method provides a late fusion of two pre-trained foundation models with distinct approaches: Yolo [41] for object detection and Mask2Former [6] for semantic segmentation. Mask2Former, pre-trained on the Cityscapes dataset [8] comprising labeled urban scenes from European cities and covering 30 unique classes, aligns well with our study's focus on urban elements. The Yolo model version 8, pre-trained on the MS COCO dataset [31], detects 80 classes of everyday objects and is particularly useful for identifying the *tablet* used as a navigation aid in our study. MYFix includes a mask coverage check to differentiate closely situated or overlapping objects at varying depths which commonly appear in real-world scenes. Inputs to the model are fixation locations along with scene video frame numbers in a CSV file, and the scene video. Frames corresponding to the fixations are processed through Yolo and Mask2Former, with the output being a CSV file listing detected labels by these two models, along with various confidence measures recommended for a final label selection. The accuracy of this method evaluated against two human-annotated datasets is reported to be approximately 81% and 89% for a controlled data collection and a real-behavior dataset collected in a highly complex urban environment.

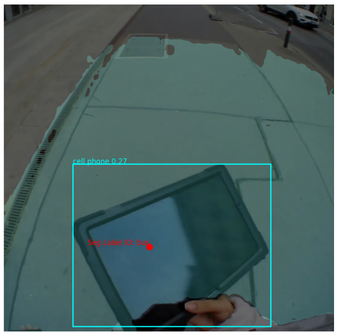
MYFix outputs separate predictions from each of the two models, and the article [3] provides general guidelines on selecting a definitive label for each fixation. Our decision-making process was tailored to these guidelines and our analysis's specific goals. Our method involved: choosing the matching label where both models agreed, choosing the prediction of one model if the other failed (commonly Yolo), and as suggested by this article, we decided to use Yolo's predictions of *laptop*, *cell phone*, *tv*, or *monitor* as proxies for our *tablet* device, since Yolo is not specifically trained to identify tablets and often categorizes them as one of these objects[2]. Finally, for all other frames, we applied Yolo's confidence and mask coverage metrics, setting a 70% threshold for label determination when the two models' predictions disagreed. Figure 2 depicts some exported frames from MYFix.

[2] This approach was feasible because the tablet was the only electronic device used by our participants. This method would not be effective if participants used a second device, such as a cell phone.

Although no method for automatic annotation is perfect and inaccuracies in model predictions are to be expected, these methods remain the only practical solution for analyzing extensive datasets, such as the one presented in this work. To minimize the effects of such inaccuracies and in line with the objectives of our research, we deliberately limited the annotations by focusing on three categories of objects based on their significant semantic relevance to the wayfinding task: *building* (as a proxy for potential landmarks), *street sign* (as an explicit environmental navigation aid), and *tablet* (as the digital navigation aid used in the study). Additionally, we grouped all other elements in the environment into a single category named *other*. This category contains all moving objects (like people and vehicles) and all static objects other than buildings.



(a) The tablet being detected as a *TV* by Yolo.



(b) The tablet being detected as a *cell phone* by Yolo.

(c) The tablet being detected as a *laptop* by Yolo.

(d) The building detected by Mask2Former only.

Figure 2 Exported frame samples from the automatic fixation annotation system suggested by [3] are presented, highlighting the primary objects of our study's interest.

4 Results

In this section, we present the results of our study from two perspectives: gaze behavior and visual attention. Gaze behavior refers to the analysis of participants' gaze movement patterns, focusing solely on the sequence and duration of gaze events recorded. On the other hand, visual attention involves content-dependent analysis, examining the specific objects observed by the participants during the wayfinding process. Both aspects of the participants' interactions with their environment are critical for understanding the cognitive processes during different wayfinding stages. Before this comparison, we first simply compared the duration of each stage between the two familiar and unfamiliar groups. The Mann-Whitney

U test [33] revealed significant differences in all stages: self-localization ($p < 0.001$), route planning ($p = 0.012$), monitoring ($p < 0.001$), and goal recognition ($p < 0.001$). These differences indicate shorter durations of self-localization, route planning, and monitoring for familiar participants, suggesting their quicker completion of these stages. Conversely, the duration of goal recognition is longer for familiar individuals, suggesting they identify the destination earlier. To mitigate the impact of any duration difference, the following measures are all normalized based on the duration of each route and the duration of each stage in the whole dataset.

4.1 Gaze Behavior

Our analysis of gaze behavior included the computation of fixation and saccade frequencies and durations for each wayfinding stage, as shown in Figures 3 and 4. Figure 3 shows fixation behavior, which serves as a proxy for information processing. It presents notched boxplots of fixation frequency (A) and mean fixation duration (B), enabling comparisons of the median and dispersion (interquartile range) across participants who walked in familiar and unfamiliar environments, and across the four stages of wayfinding. These measures have been normalized by the duration of each stage to account for variance in stage lengths. In Figure 4, we illustrate saccadic behavior, which reflects the patterns of visual search. Notched boxplots of saccade frequency (A) and mean saccade duration (B) are presented to show the differences between familiar and unfamiliar groups, as well as to highlight distinct behavioral patterns within the stages of wayfinding. In all boxplots, non-overlapping notches (indicated by $*$ in the plots) indicate significantly different medians with 95% confidence.

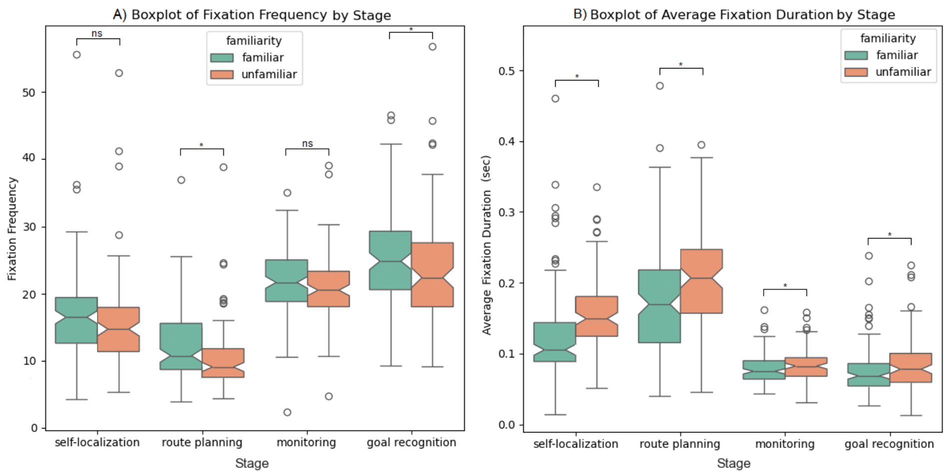


Figure 3 Fixation behavior across different wayfinding stages. Plot A presents the fixation frequency, while plot B shows the fixation duration across various stages of the task. Values are normalized based on stage duration. An asterisk denotes a significant difference in the median with 95% confidence. Shorter yet more frequent fixations may indicate heightened visual search compared to concentrated focus. Further interpretations of this plot can be found in Section 5.

4.2 Visual Attention Behavior

Extending our analysis beyond gaze metrics (content-independent), we examined visual attention by looking at which objects in the environment participants focused their attention on. This helps us to understand how they distributed their visual attention to different

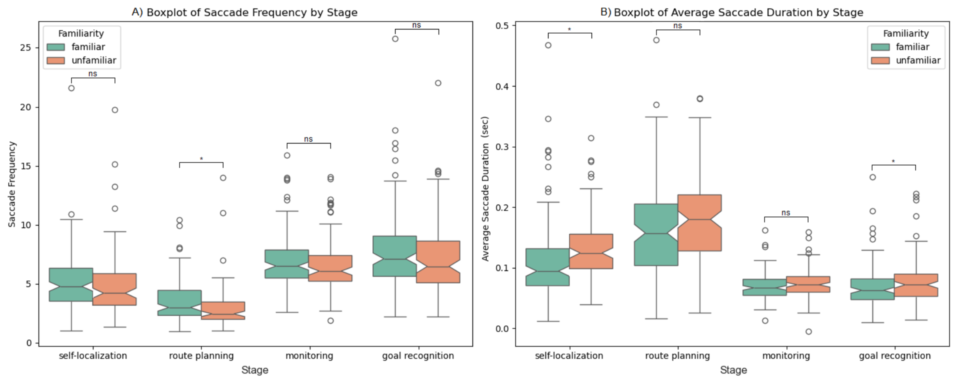


Figure 4 Saccadic behavior across different wayfinding stages. Plot *A* illustrates the variability in saccade frequency, and plot *B* depicts the variation in saccade duration. Values are normalized based on stage duration. An asterisk denotes a significant difference in the median with 95% confidence. Longer and less frequent saccades may suggest increased visual focus or a cognitively demanding task. For additional insights, refer to Section 5.

elements. Figure 5 presents the proportion of attention dedicated to the three primary objects – *tablets*, *street signs*, and *buildings* – and the category of *other*. This classification helps clarify where participants directed their focus during the wayfinding task. Complementing this, Figure 6 displays the average normalized transitions/shifts in attention between the tablet and all other environmental stimuli, offering insights into the dynamics of using an aid across different stages. The normalization allows for relative comparison across the stages, avoiding misleading interpretations based on absolute values.

5 Discussion

In this section, we discuss our results by examining them from two angles: first, how gaze behavior and visual attention evolve during wayfinding at different stages, regardless of whether participants are familiar with the environment or not, and second, how familiarity affects these behaviors. Our analysis of each stage shows how participants switch between using an aid (here provided by a tablet device) and relying on environmental cues as they move through the different stages. Additionally, by analyzing the differences in visual strategies between familiar and unfamiliar individuals, we aim to uncover the cognitive adaptations that familiarity may induce. These two perspectives provide a comprehensive understanding of the visual and cognitive strategies used in wayfinding, shedding light on the impact of prior knowledge on how people interact with and perceive their surroundings. Before exploring the details of gaze behavior and visual attention differences, we summarize various related interpretations from psychology, neuroscience, eye-tracking research, etc., to understand the significance of various patterns. This helps us better interpret the gaze behavior observed in our study:

- Fixation Behavior Interpretations: Fixation duration is linked to cognitive processing (longer durations indicating more intense processing as reported in [26]), scene perception [40], expertise (leading to shorter durations and more fixations as reported by [16]), and stress (higher stress leading to longer and more frequent fixation [24]). The number of fixations is also reported to be associated with semantic importance (more fixations on more important objects [27]), search efficiency (fewer fixations indicating higher efficiency [19]), and memory build-up (less need for fixations once objects are memorized [48]).

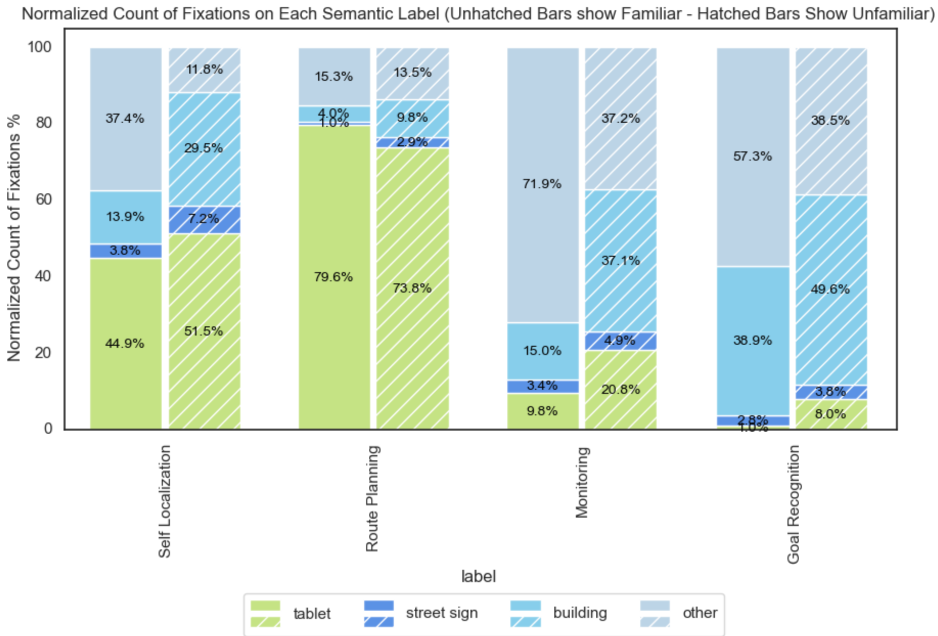


Figure 5 The stacked bar chart presents the distribution of visual attention across different semantic categories within the wayfinding task. Each segment's height indicates the proportion of gaze fixation on the corresponding objects, normalized by the duration of the stage. Unhatched bars represent the familiar group, while hatched bars indicate the unfamiliar group, highlighting differences in their focus during the stages of wayfinding.

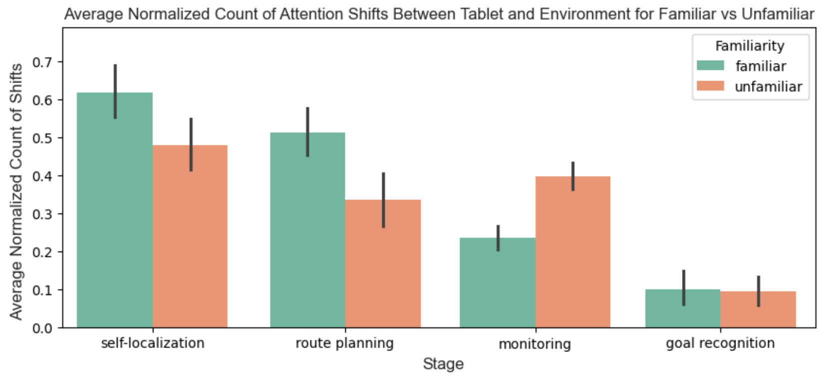


Figure 6 This bar chart illustrates the mean frequency of participants' attention shifts from the tablet to their environment and vice versa, normalized over time and averaged across individuals within each familiarity group. Higher values indicate a greater rate of attention switching, with familiar participants generally showing more frequent shifts in the stages of self-localization and route planning, while unfamiliar participants exhibit this tendency more in the monitoring stage.

- Saccade Behavior Interpretations: Saccadic duration is reported to be longer for more challenging tasks [52]. It is reported to be a sign of decreased processing capacity and higher cognitive loads [37]. It also reflects the nature of inspection (shorter saccades in more careful inspections versus overview scans as shown by [5]). Saccadic frequency is also influenced by mental workload, cognitive activation level (the state of cognitive alertness, engagement, and readiness to process information), and fatigue, decreasing with increased task difficulty or mental workload [46].

With these interpretations in mind, we proceed to discuss our specific findings in the subsequent subsections.

5.1 Wayfinding Stages

Regardless of familiarity, there are distinct visual engagement patterns during the four stages of wayfinding. Examining the plots we see distinct behavioral patterns between the stages, however, it is essential to understand the inherent differences between the tasks in every stage. The given tasks in self-localization and route planning (i.e., *"mark your location on the map"* and *"draw a preferred route to the destination"*) are well-defined and specific, unlike those in monitoring (which is typically the longest stage) and goal recognition, which encompass a variety of undefined tasks. In addition, it should be noted that participants' movement radius was relatively small during the first two stages compared to the later ones, as they were actively engaged with the map. This lack of motion suggests a static visual field, allowing for potentially quicker scanning as the range of visual information remains constant ($\approx 360°$ around the standing location). However, during the subsequent stages, as participants begin to walk, the visual environment continuously changes, introducing a dynamic flow of visual stimuli to navigate and process.

The goal recognition and monitoring stages, which include walking towards the destination, share similarities concerning visual attention and gaze behavior. Goal recognition is defined as the period from the point in space where the destination is first seen and recognized until it is reached. During this walk, one might expect to see more free viewing as the task's goal (recognizing the destination) has been achieved, as opposed to task-dependent gaze behavior. Meanwhile, monitoring can be seen as a combination of various tasks: staying oriented and localized on the route, keeping the planned route in mind, and setting intermediate goals. Therefore, gaze behavior in this stage is likely a mix of constant information searching and processing.

5.1.1 Gaze Behavior: Information Processing and Visual Search

Cognitive task complexity influences gaze behavior: route planning demands concentrated focus (fewer, longer fixations and saccades), while goal recognition involves quick, frequent visual searches (more and quicker fixations and saccades). The data from Figures 3 and 4 reveal a distinct pattern: in the self-localization and route planning stages, there are fewer but longer fixations and saccades. This trend changes in the monitoring and goal recognition stages, where fixations and saccades become more frequent but shorter. Notably, this pattern reaches its peak during the goal recognition stage and declines in the route planning stage. This suggests that drawing a route on the tablet demands more focused attention compared to the more dynamic visual search involved in confirming goal recognition. Furthermore, it can be inferred that once self-localization is completed, a basic memory of the surrounding space is established, leading to a reduction in fixation frequency to its lowest point in the route planning stage.

Gaze behavior shifts from information processing in early stages to active information seeking in later stages, reflecting the cognitive demands and stimuli density of the environment. Visual search in the monitoring stage is similar to goal recognition but is less pronounced, possibly due to the involvement of various tasks that require different cognitive loads. Overall, gaze behavior in the self-localization and route planning stages seems to be more about collecting and processing information, while the monitoring and goal recognition stages are characterized by the seeking of information and cues. Additionally, a combination of high saccade and fixation frequencies could point to an environment filled with stimuli, prompting frequent glances around, as observed in the monitoring and goal recognition stages. In contrast, lower frequencies in both metrics might indicate an environment with fewer stimuli, leading to less need for constant eye movements. Moreover, when the goal is reached, shorter saccades are observed, which could be attributed to a decreased cognitive activation level, as the task is perceived as completed (Section 12.7.1 in [25]).

5.1.2 Visual Attention and Attention Shifts

Participants' attention shifts from the navigation tool to environmental cues through wayfinding stages, reflecting a move from digital aid reliance to environmental engagement. Figure 5 demonstrates how participants' focus on various elements is distributed through different wayfinding stages. There is a notable transition in attention from the *tablet* to *buildings* and *other environmental cues* as individuals progress from the self-localization and route planning stages to monitoring and goal recognition. This shift likely represents a move from initial reliance on navigation tools to a deeper engagement with the surrounding environment as the task advances. In addition, during the self-localization and route planning stages, the tablet is given greater semantic importance as dictated by the requirements of the task, resulting in more attention being paid to it.

Although *street signs* receive less attention overall, their significance as an environmental aid during self-localization and monitoring is notable, particularly considering their less frequent presence in the environment and the fact that their information can more quickly get absorbed. The *other* category, encompassing a diverse range of environmental features such as roads, sidewalks, vegetation, people, and vehicles, maintains a consistent level of attention throughout the monitoring and goal recognition stages. This indicates a steady awareness of the general environment, reflecting the participants' ongoing engagement with both static and dynamic elements around them.

In general, Figure 6 shows a trend where attention shifts are more pronounced during the initial stages (self-localization and route planning) and then decrease as participants move to later stages (monitoring and goal recognition). This trend suggests a common pattern of reliance on the tablet during the early stages of wayfinding, for self-localization and planning, with a gradual shift to environmental cues as the participant progresses through the stages. Such a trend could imply a learning curve where participants initially seek guidance and confirmation but gradually become more adapted to environmental cues as they proceed.

5.1.3 Interplay of Visual Attention and Gaze Behavior

Self-localization and route planning demand focused attention, seen in fewer, longer gaze events, and more transitions between tablet and environment. In later stages, an inverted pattern, together with increased reliance on environmental cues, indicates information seeking and visual search. A high number of transitions between the tablet and the environment (Figure 6), which indicates frequent shifts in visual attention, coincides with relatively low but prolonged fixation, as shown in Figure 3. This would indicate a more concentrated or focused attention strategy for information processing. Conversely, fewer transitions between the tablet and the environment may correspond with more frequent fixations on a higher variety of objects, indicating a more dynamic visual search pattern. The two initial stages, show more visual shifts as they might quickly alternate between the tablet and environmental cues to orient themselves and plan their route. This would be consistent with a higher fixation count on the tablet, as seen in Figure 5. In later stages, we observe fewer visual shifts which could indicate that wayfinders rely more on environmental cues they have learned to recognize, as reflected by a higher fixation count on objects like buildings in Figure 6.

5.2 Influence of Familiarity on Gaze and Visual Attention

We examined the differences between conditions where participants were familiar with the environment and those where they were not. Analyzing the four figures in section 4, it becomes apparent that **spatial familiarity affects gaze behavior, visual attention, and the reliance on navigational aids during wayfinding.**

5.2.1 Gaze Behavior: Information Processing and Visual Search

In all wayfinding stages, unfamiliar participants show fewer but longer fixations and saccades, reflecting higher cognitive load due to increased visual search in the absence of a mental representation. Conversely, familiar participants exhibit more frequent and shorter gaze events, reflecting efficient information processing. Figure 3 provides insight into the overall patterns of attentional engagement across different stages of wayfinding between familiar and unfamiliar groups. Familiar participants represent more (A) but shorter (B) fixations compared to unfamiliar ones. First and foremost, this can be an indicator of expertise or prior knowledge [16], which in our case is translated into familiarity. Additionally, it suggests that the cognitive workload of processing the information and perceiving the environment is higher for unfamiliar participants as they are more focused (less but longer fixations) which might reflect a need for more time to process or understand the information they encounter. This behavior is more pronounced in the self-localization and route-planning stages when they are actually gathering new information. The longer fixations observed in the unfamiliar group may also result from increased stress levels, likely due to their lack of experience in the environment. In contrast, familiar participants show more and shorter fixations, suggesting faster processing or recognition of necessary information which could be due to the fact that they know the environment and are then only looking for cues.

Similar patterns are visible in the saccadic behavior. Differences between familiar and unfamiliar participants in Figure 4 could reflect distinct cognitive strategies between the two groups. Familiar participants have shorter and more frequent saccades, indicative of efficient information processing based on prior knowledge or the fact that the task is less challenging for them. The observation that unfamiliar participants exhibit longer saccades could also suggest a more overview scan of the environment as the amount of information to be processed and structured is probably much higher for them than for the familiar individuals, who already have a mental representation and therefore use a more efficient search strategy based on more careful inspection.

5.2.2 Visual Attention and Attention Shifts

Familiar wayfinders, despite shifting their attention to the navigation device as much as or even more than their unfamiliar counterparts, depend less on the device and focus more on the environment, indicating a more efficient use of both information. Although gaze behavior, as discussed in the previous subsection, provides valuable insights, it cannot present a complete picture of visual engagement. Figures 3 and 4 illustrate that, although gaze behavior differs in self-localization and route planning between familiar and unfamiliar groups, the fixation and saccade patterns during monitoring and goal recognition are quite similar (not significantly different). Yet, the visual attention plot in Figure 5 demonstrates differences in how each group allocates their attention between environmental elements (i.e., all colors together except green) and the tablet (i.e., green color), suggesting variations in search strategies and information processing during these stages. One can see that during goal recognition, the familiar group spends significantly less time fixating on the tablet (1.0% vs. 8.0% for the unfamiliar group), while showing a similar amount of shifts (see Figure 6), suggesting a more efficient information-gathering strategy.

The unfamiliar group consistently prioritizes buildings as landmarks in all wayfinding stages, whereas the familiar group engages more extensively with broader environmental features. Buildings, assumed to serve as landmarks [43], seem to be more important for the unfamiliar group in all wayfinding stages, while the familiar group shows greater awareness, focusing on a wider range of environmental characteristics beyond just buildings. This can indicate deeper engagement with their surroundings. Another observation is that, although the self-localization and route planning tasks are explicitly structured to require significant interaction with the tablet, there are still notable differences between the familiar and unfamiliar groups. As shown in Figure 5, in the self-localization phase, the unfamiliar group split their attention almost evenly between the tablet ($\approx 51\%$) and the rest of the environmental objects ($\approx 48\%$), out of which buildings accounted for about 61% and street signs for 14% of their attention. In contrast, the familiar group devoted less than 45% of their attention to the tablet and focused more on environmental objects other than buildings. Their frequent shifts to the tablet (Figure 6) indicate an efficient approach to correlating environmental cues with map information. This is in line with [28] reporting that more effective self-localization is associated with increased visual focus on the map and increased shift of attention between the map and the environment. The unfamiliar group, however, appears to require more in-depth cross-referencing for self-localization, as reflected in their focused tablet use and fewer environmental transitions. This pattern is also observed in route planning, where familiar participants probably rely more on their mental representation of the environment.

Street signs, despite their infrequent appearance and low visual presence, play a crucial role as navigation aids, for both groups but particularly for unfamiliar wayfinders. The attention paid to street signs by this group underlines their usefulness for wayfinding and particularly self-localization. This observation contradicts the finding by [53] that street names in augmented reality environments increase cognitive load without contributing significantly to orientation.

During the monitoring stage, unfamiliar participants exhibit an increased number of attention shifts, possibly indicating higher uncertainty or a need for more frequent cross-referencing. In contrast, both groups show fewer shifts in the goal recognition stage, suggesting a decreased reliance on the tablet as they approach their destination and possibly rely more on their recognition of the goal location or other environmental landmarks.

5.2.3 Interplay of Visual Attention and Gaze Behavior

Familiarity significantly influences visual attention, gaze behavior, and the approach to using navigational aids (e.g., tablets or street signs), leading to distinct information search and processing strategies. Participants familiar with the environment tend to rely less on the tablet after locating themselves and planning their route, possibly due to better knowledge or a more accurate mental representation. The more and shorter fixations showing efficient search behavior could indicate that familiar wayfinders perform more efficient environmental scans and more efficient cross-referencing with the tablet, relying on their mental representation for deeper engagement with their surroundings.

On the other hand, among all the environmental objects, unfamiliar participants focus more attention on buildings and street signs in all stages compared to familiar participants, peaking at monitoring and goal recognition. This could be to compensate for their lack of knowledge using landmarks and explicit environmental cues such as street names. Their approach is characterized by fewer but longer fixations and longer saccades, indicating a more cautious and detailed inspection behavior rather than a very efficient one. Overall, the results highlight the effects of familiarity with the environment on wayfinding behavior and the information processing and visual search strategies used by familiar and unfamiliar wayfinders.

5.3 Limitations

Acknowledging the constraints of our research is essential for a thorough understanding of its extent and implications. Although an outdoor experiment offers valuable insights, it introduces inherent biases that could be partly mitigated with a larger sample. However, the challenges of recruiting volunteer participants meeting specific criteria –in our case such as residency in Vienna for at least three months (to ensure a level of familiarity) and absence of corrective lenses with dioptre ± 3.5– pose significant challenges. Additionally, despite efforts to select familiar and unfamiliar locations with similar environmental and urban design characteristics, controlling for all external variables is a complex task with potential impacts on the results. Moreover, navigating city streets involves numerous visual distractions. Although participants were instructed to maintain their normal navigation behavior while focusing on the task and avoiding unnecessary interactions, the presence of distractions remains a consideration.

Since conducting such large-scale experiments is costly (in terms of money, time, effort, etc.), we used a cognitive reset technique to effectively collect more data from each participant. However, it is not entirely clear whether this method completely neutralizes the influence of participants' memories on subsequent trials while self-localizing. Fortunately, in our study, the time spent on subsequent self-localization tasks did not noticeably decrease, suggesting that the reset worked to some extent. We also automated the process of labeling video data, which is efficient but could be error-prone. Since it is not possible to manually check every labeling, we randomly checked 10% of the data to ensure its reliability. In addition, cases where participants needed assistance from the experimenter or made mistakes during monitoring should be investigated in more detail to understand their impact on the data.

6 Conclusion and Future Work

In this study, we investigated the gaze behavior and visual attention of wayfinders moving through different familiar/unfamiliar environments, following the theoretical stages of way-finding proposed by Downs and Stea [12]. By observing 56 participants as they walked three familiar and three unfamiliar routes, we were able to uncover different gaze patterns that correspond to the stages of wayfinding and are influenced by both familiarity and the nature of the tasks. The differences are evident not only in fixation and saccade behavior but also in the specific visual content that participants engaged with and their patterns of attention shifting. Our findings open up possibilities for further research: a detailed analysis of the use of the navigational aid (here a digital map) to understand what participants focus on within the map and how this affects wayfinding, as well as an investigation of the *other* category in visual attention to identify environmental elements that attract the attention of familiar wayfinders. Another question would be to investigate the relationship between attention to specific objects and the ability to learn and memorize routes for future unassisted navigation. Furthermore, the observed frequent shifts in attention during self-localization and route planning raise the question of their correlation with cognitive load and perceived difficulty of navigation. In addition, the complexity of the monitoring stage presents a puzzle. Across both familiarity levels, this stage showed a wide-ranging distribution of attention, which could indicate a more exploratory approach or a composition of subtasks within the stage itself. A more detailed, second-by-second analysis could reveal these underlying tasks. Now that unique patterns of vision and gaze have been identified for each stage, this knowledge can be used to improve the recognition of wayfinding stages, which not only enriches our understanding of spatial cognition during navigation but could also contribute to the development of more intuitive navigation systems. These systems could provide stage-specific assistance customized to the immediate needs of the wayfinder, promoting a more efficient and user-friendly navigation experience.

References

1. N. Alinaghi, M. Kattenbeck, A. Golab, and I. Giannopoulos. Will you take this turn? gaze-based turning activity recognition during navigation. In *Proc. of the 11th Intl. Conf. on Geographic Information Science (GIScience 2021)-Part II*, 2021.
2. Negar Alinaghi and Ioannis Giannopoulos. Consider the head movements! saccade computation in mobile eye-tracking. In *2022 Symposium on Eye Tracking Research and Applications*, pages 1–7, 2022.
3. Negar Alinaghi, Samuel Hollendonner, and Ioannis Giannopoulos. Myfix: Automated fixation annotation of eye-tracking videos. *Sensors*, 24(9), 2024. doi:10.3390/s24092666.
4. Negar Alinaghi, Markus Kattenbeck, and Ioannis Giannopoulos. I can tell by your eyes! continuous gaze-based turn-activity prediction reveals spatial familiarity. In *15th International Conference on Spatial Information Theory (COSIT 2022)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.
5. Guy Thomas Buswell. *How people look at pictures: a study of the psychology and perception in art.* Univ. Chicago Press, 1935.
6. Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022.
7. Eric Chown, Stephen Kaplan, and David Kortenkamp. Prototypes, location, and associative networks (plan): Towards a unified theory of cognitive mapping. *Cognitive Science*, 19(1):1–51, 1995.
8. Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.
9. Ruth Conroy Dalton. The secret is to follow your nose: Route path selection and angularity. *Environment and Behavior*, 35(1):107–131, 2003.
10. Weihua Dong, Hua Liao, Bing Liu, Zhicheng Zhan, Huiping Liu, Liqiu Meng, and Yu Liu. Comparing pedestrians' gaze behavior in desktop and in real environments. *Cartography and Geographic Information Science*, 47(5):432–451, 2020.
11. Weihua Dong, Tong Qin, Tianyu Yang, Hua Liao, Bing Liu, Liqiu Meng, and Yu Liu. Wayfinding behavior and spatial knowledge acquisition: Are they the same in virtual reality and in real-world environments? *Annals of the American Association of Geographers*, 112(1):226–246, 2022.

12 M. R. Downs and D. Stea. *The World In The Head*. Harper & Row Series in Geography. Harper & Row, 1977.

13 Karen M Evans, Robert A Jacobs, John A Tarduno, and Jeff B Pelz. Collecting and analyzing eye tracking data in outdoor environments. *Journal of Eye Movement Research*, 5(2):6, 2012.

14 Anna Charisse Farr, Tristan Kleinschmidt, Prasad Yarlagadda, and Kerrie Mengersen. Wayfinding: A simple concept, a complex process. *Transport Reviews*, 32(6):715–743, 2012.

15 Tom Foulsham and Alan Kingstone. Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision research*, 50(8):779–795, 2010.

16 Andreas Gegenfurtner, Erno Lehtinen, and Roger Säljö. Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational psychology review*, 23:523–552, 2011.

17 I. Giannopoulos, P. Kiefer, and M. Raubal. Gazenav: gaze-based pedestrian navigation. In *Proc of MobileHCI 2015*, pages 337–346, 2015.

18 Ioannis Giannopoulos. *Supporting Wayfinding Through Mobile Gaze-Based Interaction*. PhD thesis, ETH Zurich, 2016.

19 Joseph H Goldberg, Mark J Stimson, Marion Lewenstein, Neil Scott, and Anna M Wichansky. Eye tracking in web search tasks: design implications. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 51–58, 2002.

20 Reginald G Golledge. Human wayfinding and cognitive maps. *Colonization of unfamiliar landscapes: the archaeology of adaptation*, 25, 2003.

21 Mary Hegarty, Chuanxiuyue He, Alexander P Boone, Shuying Yu, Emily G Jacobs, and Elizabeth R Chrastil. Understanding differences in wayfinding strategies. *Topics in Cognitive Science*, 15(1):102–119, 2023.

22 Mary Hegarty, Anthony E Richardson, Daniel R Montello, Kristin Lovelace, and Ilavanil Subbiah. Development of a self-report measure of environmental spatial ability. *Intelligence*, 30(5):425–447, 2002.

23 Qiaosong Hei, Weihua Dong, and Bowen Shi. Detecting dynamic visual attention in augmented reality aided navigation environment based on a multi-feature integration fully convolutional network. *Cartography and Geographic Information Science*, 50(1):63–78, 2023.

24 Nadja Herten, Tobias Otto, and Oliver T Wolf. The role of eye fixation in memory enhancement under stress–an eye tracking study. *Neurobiology of learning and memory*, 140:134–144, 2017.

25 Kenneth Holmqvist, Marcus Nyström, Richard Andersson, Richard Dewhurst, Halszka Jarodzka, and Joost Van de Weijer. *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford, 2011.

26 David E Irwin. Fixation location and fixation duration as indices of cognitive processing. In *The interface of language, vision, and action*, pages 105–133. Psychology Press, 2013.

27 Robert JK Jacob and Keith S Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In *The mind's eye*, pages 573–605. Elsevier, 2003.

28 Peter Kiefer, Ioannis Giannopoulos, and Martin Raubal. Where am i? investigating map matching during self-localization with mobile eye tracking in an urban environment. *Transactions in GIS*, 18(5):660–686, 2014.

29 Alexander Klippel. Wayfinding choremes. In *Spatial Information Theory. Foundations of Geographic Information Science: International Conference, COSIT 2003, Kartause Ittingen, Switzerland, September 24-28, 2003. Proceedings 6*, pages 301–315. Springer, 2003.

30 H. Liao, W. Zhao, C. Zhang, W. Dong, and H. Huang. Detecting individuals' spatial familiarity with urban environments using eye movement data. *CEUS*, 93:101758, 2022.

31 Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. Microsoft coco: Common objects in context, 2015. `arXiv:1405.0312`.

32 Kevin Lynch. *The image of the city*. MIT press, 1964.

33 Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pages 50–60, 1947.

34 David Marr. *Vision: A computational investigation into the human representation and processing of visual information.* MIT press, 2010.

35 Daniel R. Montello. Spatial cognition. In Neil J. Smelser and Paul B. Baltes, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages 14771–14775. Elsevier, Oxford, 2001.

36 Daniel R Montello. Navigation. In Priti Shah and Akira Miyake, editors, *The Cambridge Handbook of Visuospatial Thinking*, pages 257–294. Cambridge University Press, 2005. `doi:10.1017/CBO9780511610448.008`.

37 Minoru Nakayama, Koji Takahashi, and Yasutaka Shimizu. The act of task difficulty and eye-movement frequency for the'oculo-motor indices'. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 37–42, 2002.

38 Romedi Passini. Wayfinding: A conceptual framework. *Urban Ecology*, 5(1):17–31, 1981.

39 Beatrice Rammstedt and Oliver P John. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212, 2007.

40 Keith Rayner and Alexander Pollatsek. Eye movements and scene perception. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 46(3):342, 1992.

41 Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

42 Ferdinand Reimer, Ulrich Kral, Emre Can Sönmez, Friedrich Hauer, Severin Hohensinner, Hannah Wolfinger, Klara Stuppacher, Andreas Danzinger, Ingeborg Hengl, Lupina Prospero, et al. Data description of "building age map, vienna, around 1920". *Data in Brief*, 41:107864, 2022.

43 Kai-Florian Richter and Stephan Winter. Landmarks. *Springer Cham Heidelberg New York Dordrecht London. doi*, 10:978–3, 2014.

44 Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78, 2000.

45 Norma Saiph Savage, Wendy Chun, Norma Elva Chavez, and Tobias Höllerer. Seems familiar: An algorithm for inferring spatial familiarity automatically. *Computer Science Department, University of California, Santa Barbara, accessed Feb*, 5, 2013.

46 Robert Schleicher, Niels Galley, Susanne Briest, and Lars Galley. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, 51(7):982–1010, 2008.

47 Charlotte Schwedes and Dirk Wentura. Through the eyes to memory: Fixation durations as an early indirect index of concealed knowledge. *Memory & cognition*, 44:1244–1258, 2016.

48 Benjamin W Tatler, Iain D Gilchrist, and Michael F Land. Visual memory for objects in natural scenes: From fixations to object files. *The Quarterly Journal of Experimental Psychology Section A*, 58(5):931–960, 2005.

49 Aldert Vrij, João Oliveira, Annie Hammond, and Howard Ehrlichman. Saccadic eye movement rate as a cue to deceit. *Journal of Applied Research in Memory and Cognition*, 4(1):15–19, 2015.

50 J. M Wiener, S. J Büchner, and C. Hölscher. Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Computation*, 9(2):152–165, 2009.

51 Jan M Wiener, Alexander Schnee, and Hanspeter A Mallot. Use and interaction of navigation strategies in regionalized environments. *Journal of Environmental Psychology*, 24(4):475–493, 2004.

52 Liu Xin, Zheng Bin, Duan Xiaoqin, He Wenjing, Li Yuandong, Zhao Jinyu, Zhao Chen, and Wang Lin. Detecting task difficulty of learners in colonoscopy: Evidence from eye-tracking. *Journal of Eye Movement Research*, 14(2), 2021.

53 Fan Yang, Zhixiang Fang, and Fangli Guan. What do we actually need during self-localization in an augmented environment? In *International symposium on web and wireless geographical information systems*, pages 24–32. Springer, 2020.

54 Z. Zhou, R. Weibel, and H. Huang. Familiarity-dependent computational modelling of indoor landmark selection for route communication: a ranking approach. *International J.of Geographical Information Science*, pages 1–33, 2021.