

Predicting Mobile Map Interactions with a Context-Enriched Sequential Fusion Model

From Unobtrusive Data Collection to Explainable Models Across
Intrinsic, Behavioral, and Extrinsic Contexts

Dissertation

zur Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von
Donatella Zingaro

aus
Italien

Promotionskommission

Prof. Dr. Sara I. Fabrikant (Vorsitz und Leitung)
Dr. Tamasch Reichenbacher (Leitung)
Prof. Dr. Ross S. Purves
Prof. Dr. Johannes Schoening

Zürich, 2026

Donatella Zingaro

*Predicting Mobile Map Interactions with a Context-Enriched Sequential Fusion Model -
From Unobtrusive Data Collection to Explainable Models Across Intrinsic, Behavioral, and
Extrinsic Contexts*

PhD Defense, 3rd December 2025

Supervisors: Prof. Dr. Sara Irina Fabrikant, Dr. Tumasch Reichenbacher

Committee Members: Prof. Dr. Sara Irina Fabrikant, Dr. Tumasch Reichenbacher, Prof.
Dr. Ross Stuart Purves, Prof. Dr. Johannes Schoening

University of Zurich

Geographic Information Visualization and Analysis (GIVA) Group

Department of Geography

Faculty of Science

Winterthurerstrasse 190

8057 Zurich

Switzerland

Abstract

Despite the ubiquity of mobile maps in everyday life over the past 30 years, interdisciplinary research in cartography, human–computer interaction, and cognitive science has yet to fully elucidate when and why people use them in real-world. The present thesis addresses this gap by analyzing rich contextual data obtained at the moment an individual unlocks their phone to open and interact with a mobile map application. This research thus integrates three complementary contextual levels: behavioral (phone usage patterns, such as taps and app switching); intrinsic (individual characteristics, such as sense of direction, spatial anxiety, and demographics); and extrinsic (properties of the surrounding street network, i.e., urban forms).

We conducted two in-the-wild data collection campaigns designed according to the parsimony principle (Occam’s Razor); that is, identifying the best explanations constructed using the smallest possible set of elements. We implemented this approach by employing tappigraphy, a well-established neuroscience method used to capture the temporal variation of cognitive states by analyzing a user’s taps on the touchscreen. For this project, we developed MapOnTap, an application that runs in the background and collects tap sequences only when the user is actively using the smartphone and the corresponding application in the foreground. Adhering to the parsimony of information, no additional information was recorded beyond what was strictly necessary, and participants retained full control over data collection. Building on this basic data collection procedure, we explored whether the parsimony principle could still be respected while gradually adding contextual features. This involved adjusting the timing of recordings and the quantity of information required across two data campaigns to determine which set of information could be used most effectively to understand how people use mobile maps in the real world. The two main data campaigns we implemented were MapOnTap I (MoT I) and MapOnTapII (MoT II). MoT I collected tappigraphy and GPS traces over two years and did not limit the duration for which participants contributed data. MoT II was conducted over three months, with each participant contributing data for two weeks. For this campaign, tappigraphy and GPS traces were combined with two standardized questionnaires to investigate links between interaction traces and individual traits (sense of direction and spatial anxiety) within the geographic context.

Exploratory analyses of MoT I and MoT II used simple features, such as tap counts for specific categories of mobile map apps and distance-from-home measures, to assess the granularity and quality of the collected data. Analyses of additional features, such as map session duration and the frequency of application switching during phone sessions, revealed associations with individual traits and gender. These features were used to examine the relationship between the moment-to-moment interactions captured by tappigraphy and individual differences in sense of direction and spatial anxiety. Results showed that

women aged 25–34 exhibited higher levels of spatial anxiety and more frequent application switching than men of the same age when using mobile map applications. Conversely, men in this age group spent more time using mobile map applications and reported a better sense of direction.

Focusing on the MoT II data, which are contextually richer, we derived more complex across-session features from tappigraphy. These included the inter-map interval (IMI; a measure of the regularity of map use over time) and the session initiation rate (SIR; the probability of opening a map app within a phone session). These indicators were used to examine variations in participants’ engagement with map applications according to their individual traits and gender. This analysis did not yield significant results, which is not unexpected given that our study spanned only two weeks, and such measures typically require longer observation windows (months or years) to stabilize.

Consequently, we shifted our focus to more fine-grained within-session features, such as taps per second (reflecting the frequency of interaction), the coefficient of variation of inter-tap intervals (CV-ITI, reflecting tap rhythm regularity), and transitions between app categories per minute (the rate of switching across different app categories). Using dimensionality reduction and density-based spatial clustering of applications with noise (DBSCAN) clustering, we identified six recurring interaction modes that together represented approximately 69% of the 3667 valid navigation sessions. We then identified the most recurrent mode for each participant (the cluster appearing most frequently during their phone sessions) and tested whether it predicted variations in individual traits (sense of direction, spatial anxiety) or gender. Again, no substantial effects emerged, given the short observation window and the limited sample size.

Finally, we integrated into the analysis the three contextual levels: behavioral (interaction patterns), intrinsic (traits), and extrinsic (urban form). For each navigation session, we extracted street network measures for the local area surrounding the GPS trajectory associated with that session, along with the two behavioral indicators, tap rate and tapping regularity (CV-ITI). Mixed effects and robust regression analyses revealed that spatial anxiety was the most consistent individual predictor, with higher anxiety being associated with fewer taps and less regular rhythms. The street network structure was also significant: irregular layouts (high orientation entropy) and dense local connectivity (high degree centrality) predicted lower tap rates, while greater accessibility (high closeness centrality) was associated with steadier interaction rhythms, suggesting that more legible environments support more consistent engagement. Trait–environment interactions further revealed that participants’ sense of direction had a stronger influence on tap rate in irregular layouts, as those with a good sense of direction maintained higher tap rates and more consistent interaction rhythms compared to participants with a poorer sense of direction. In contrast, spatial anxiety amplified the effect of network topology on irregular tapping, with more anxious participants showing greater variability and less consistent tapping patterns than those with lower spatial anxiety.

Building upon these findings, we developed a context-aware long short-term memory (LSTM) fusion model to predict the probability that the user’s next interaction would be with a map app. The model combined two tensors: one capturing 10-minute windows of recent tap sequences, and the other representing individual traits, gender, and urban network features. The LSTM sequence model alone performed above chance (accuracy = 0.77, AUC = 0.74), detecting 57% of real map events at the chosen threshold ($p = 0.8$), but with very low precision (22%). The low precision reflects the highly imbalanced nature

of the data, in which map app interactions were relatively rare. To address this, we added contextual information, first as an eight-dimensional feature set and then in compressed form using an autoencoder (AE), which reduced the inputs to three latent dimensions. The additional context improved performance (accuracy = 0.83, recall = 0.71, F1 score = 0.47, AUC = 0.86, and precision = 36%). The AE-based model, however, performed slightly worse overall (accuracy = 0.82, recall = 0.64, F1 score = 0.40, AUC = 0.84) and exhibited very low precision (32%). A post hoc analysis on the AE dimensions was used to determine which elements drove the prediction towards map and non-map classification, revealing that urban complexity and higher spatial anxiety tended to reduce immediate map use, while greater confidence (better sense of direction and lower anxiety) increased it.

This thesis represents the first attempt to explain when and why individuals use mobile maps, employing real-world navigation data and a combination of behavioral traces, individual traits, and urban form. Despite the unsuccessful endeavor to build a predictive model of near-term map use, analyses of model explainability demonstrated that engagement is jointly shaped by three levels of context — behavioral, intrinsic, and extrinsic. While the available data were not sufficient to establish a robust computational model, insights from behavioral traces and their interactions with individual traits and urban form support the feasibility of this approach. To improve predictive modeling, future work should use larger, more diverse samples and extend data collection over longer periods. The current dataset, collected across two data campaigns, included participants recruited primarily from university networks and early-career adults based in Europe, which limits the generalization of the findings.

Finally, the present findings resonate with three decades of mobile map design research, highlighting the need to move beyond a focus on mere efficiency, which has traditionally been built on a narrow view of extrinsic factors. Future map design should account for users' cognitive and emotional states, adapting timing, complexity, and presentation to minimize cognitive and emotional strain and better support everyday navigation.

Keywords: mobile maps, smartphone interaction, spatial anxiety, sense of direction, urban form, context-fusion, LSTM

Acknowledgments

There are so many people who I would like to thank for their support along this journey. I will begin with those who matter most in my life: my *mum* and *dad*. I owe you endless gratitude for everything you have given to Andrea and me. I have been lucky enough to know what selfless love truly means: a love that asks for nothing in return, and which keeps giving simply because it exists. You would give your lives for us, and you have always put our happiness before your own, ensuring that we would have a better future and the freedom to pursue our dreams. You are the true giants on whose shoulders I wish to stand. You taught me to accept myself, to appreciate what I have without taking it for granted, and to respect others. Your kindness and good nature have nurtured my trust in the world and made me fall in love with life. If I know how to love today, it is because you taught me, thank you.

To the members of my committee: *Tumasch*, MapOnTap was not an easy project, but together we managed to give it shape and direction. Thank you for your help and support, and for the trust that you have consistently placed in me. Whenever frustration convinced me that I had failed, you helped me step back, see the path from the beginning, and recognize the milestones I had achieved. I am infinitely grateful; you pushed me to always give my best. *Sara*, your experience and rich scientific background have been an inexhaustible source of learning. I am truly lucky to have learned from you, not only in research, but also in the art of writing and shaping ideas. *Ross*, your guidance has meant a lot to me. I am grateful for your genuine interest and thoughtful support. A PhD is never easy, but having someone who truly cares and offers help when it matters makes all the difference, a rare quality that I deeply admire and aspire to pass on. *Johannes*, your energy and creativity have been a constant source of inspiration. Thank you for welcoming me to St. Gallen and for sharing your enthusiasm and endlessly innovative ideas. Your ability to see problems from new perspectives enriched me and taught me to value research as a space of freedom and imagination.

Andrea, I suspect that thirty years ago, you were disappointed to learn that your little sister would not be the perfect football partner. You were wrong; I turned out to be quite sporty. Today, looking ahead, I feel extremely lucky to still have you by my side. None of us was ready to lose you, and I am deeply grateful that you are here. I am also grateful for your family: *Tania*, who has stood by your side for more than 16 years with a quiet strength that I deeply admire; *Daphne*, whose warmth and boundless creativity I adore; and *Victoria*, a clever, curious whirlwind who reminds me so much of myself. Thanks to all of you, my life is filled with joy, even when I am not in Modena.

Aunt Nunzia, you have been a second mother, a friend, and someone I can turn to and trust. I treasure the childhood holidays we spent together, as well as the countless times we watched *Pride and Prejudice* on the BBC, followed by *Downton Abbey*, our own little

romantic ritual. Thank you for those moments, for your affection, and for setting such a good example.

To my colleagues at GIVA: You are more than just colleagues; you are my friends. *Armand*, you are like a brother to me. We had a lot of fun, and I know you are always there when it matters. You taught me not to overthink, and to remember that there is time and space for everything and everyone in life. *Alex*, these years have been filled with good music and good food. Here's to a thousand more karaoke nights, with little JJ joining the chorus now. *Qi*, my officemate, a pure and generous soul. We supported each other in countless ways. Thank you for the trust you have always placed in me. I will carry with me our moments of comfort and fun. I will miss having you nearby every day. *Corrado*, I admire you because you remind me that life must be lived to the fullest. Despite everything, you always do that and, your strength is a rare example that I will always carry with me. *Enea*, you are one of the most brilliant people I have ever met. It was a privilege to know you throughout the project, and especially in these last months. *Sergio*, thank you, our conversations have taught me so much, especially about metal music and movies.

Tiziana, together, we represent the Italian Renaissance: big responsibilities! Jokes aside, your calm, gentle presence has meant a lot. *Peter*, you and *Matilda* (*who bears a striking resemblance to Karen*) form a remarkable family. I appreciate your time and the consistently engaging discussions we have.

Nina, being part of DSI was special, and you are one of the main reasons this journey was so enriching. Thank you for the care and dedication you bring to everything you do. To all Ph.D. students in the Excellence Program. It has been a joy to meet and learn from such intelligent and inspiring people. *Ibra* and *Sandra*, in particular, you will always remain in my heart. We were the perfect trio, able to turn even the hardest days into moments of laughter. With you, I found attentive, sincere support and a rare, authentic bond. And, of course, I cannot forget to thank *Ioana* for the best pizza I have ever had!

Stef and *Arko*, you have not only provided me with the opportunity to work in your labs, but have also imparted so much knowledge. Your openness, curiosity, and research methodologies continue to inspire me. I am grateful for your hospitality and for sharing your time, ideas, and expertise, which have significantly enhanced my knowledge.

To *Francesca*, my best friend, with whom I have a deep connection. No matter where we are or what stage of life we are in, whenever we meet, our thoughts align, as if not a single day had passed. I do not think there is anything more beautiful than that. Thank you for always being there, even when I am far away.

To the Gaina Group: *Saverio*, *Luca*, *Albo*, *Ventu*, *Dalla*, *Paga*, *Verru*, *Manto*, *Mulas*, *Malo*, *Chiara*, and *Simona*. I learned that Modena Est truly has its 'Good Fellas', an indispensable part of Lollo's life, and now mine too. You are an extraordinary group of friends who grew up together and have always stood by one another. Thank you for making me feel part of your story. You are the reason we miss Modena every weekend and why every excuse to come back and be together becomes a good one.

To my friends from Modena who moved to Switzerland: I am so glad our paths crossed. *Memi*, the most capable and ambitious woman I know. I feel lucky to have met you. I have learned so much from you, and continue to do so every day. *Fede* is just as brilliant, and together you make a beautiful couple. You always bring the fun, thank you for that. *Paso*, now in Ticino, was the wildest addition during those desperate moments when in Wetzikon, I really miss these moments.

Melissa, Elisa, Valerio, Gabriele, and Alex L. I shared with you some of the most beautiful experiences, hiking, swimming, singing, watching football, and countless happy dinners. Thank you for being there and for the moments that made life away from home so much easier. *Alexej*, you were always caring and made sure that every event was perfectly organized and fun (although I wish I had never tried Berliner Luft). To *Sascha* and *Marlon*, thank you for your inspiring determination and ambition. *Daria*, thank you for your invaluable support. To *Dani* and *Josh*, the good lads. *Simon*, thank you for the coffee breaks that often turned out to be the best part of the day. *Leon, Bingjie, Max, Mona and Daniela, Anna and Victor*, thank you for the time we shared and the memories I will carry with me. To *Micha*, despite our football rivalries, I will always smile at the memory of your piece on ‘Inter’s Five Pears’, the best meme ever.

To those I was lucky to supervise, *Ayana, Eunice, and Oliva*, thank you for trusting me during such an important stage of your Master’s program. I learned a lot from each of you.

To *Mario*, I have great respect for you. You are someone who made his own way from nothing, went far, and gave everything to his children out of love. Your story is a valuable lesson and constant motivation to work hard. Thank you for all the Saturday dinners together. *Grazia*, your kindness and warmth have always made me feel safe. I admire everything you have done and continue to do tirelessly to make everyone around you happy. *Nunzia*, Lollo’s grandmother: I have never seen anyone carry 91 years with such grit, strength, and grace. You are wonderful. Thank you for the love you have always shown me, and for sharing the stories of your life. To *Leoparda*, the funny and cuddly one: One little snort is enough to get my attention and make me smile. You will always be the best dog in the world for me.

Last but certainly not least, *Lollo*, my sweetest love. Hannah Arendt wrote that sometimes, very rarely, a world is born between two people. I do not know how rare that is. I have no statistics. But I know it was born between us. Our microcosm, the place we retreat to when the world collapses, is our couch and the soft light of the lamp beside it. In moments when everything seemed to fall apart, you were always the calm, the warmth, and scent that dissolve every irrational fear. I know that, no matter what happens in my life, you will be on my side. This knowledge makes me one of the happiest women in the world, every day. For that, I thank you with all my heart.

My doctoral program was an intense and valuable journey, full of challenges, setbacks, and unexpected achievements. I have learned from each of you, and I will carry these moments with me as I look ahead to my future. Thank you.

Contents

Abstract	I
Acknowledgments	IV
List of Figures	XI
List of Tables	XVI
1. Introduction	1
1.1. Motivation and Problem Statement	3
1.2. Research Gaps	6
1.3. Research Objective	8
1.4. Research Questions and Approach	9
2. Related Work	10
2.1. Theoretical Foundations of Mobile Map Design	10
2.1.1. Retrieval and Filtering Spatial Information	11
2.1.2. Visual Representation of Spatial Information	12
2.1.3. Interaction as a Core Design Element	14
2.1.4. Prediction in Mobile Map Design	16
2.1.5. From Theory to Operationalization: Toward Context-Aware Map Design	18
2.2. Challenges in Context-Aware Map Design	19
2.2.1. Data Collection for Ecological Validity in Mobile Map Use	20
2.2.2. Fragmented Perspectives: Intrinsic, Behavioral, Extrinsic	23
2.2.3. Prediction of Mobile Map Use	35
2.3. Summary	39

3. Methodology	41
3.1. Map App Interaction Data Collection	42
3.1.1. The MapOnTap Application	43
3.1.2. Data Collection Campaigns	46
3.2. Map App Interaction Data Integration	48
3.2.1. Data Pre-Processing	48
3.2.2. Feature Extraction	50
3.2.3. Data Analysis	54
3.3. Map App Interaction Data Prediction	58
3.3.1. Model Architecture	60
3.3.2. Explainability Analysis of Fusion Head Predictions in Everyday Mobile Map Use	64
4. Results	66
4.1. Descriptive Results of Mobile Map App Interaction Data Collection	67
4.1.1. MoT I Descriptive Overview	67
4.1.2. MoT II: Descriptive Overview	71
4.2. From Exploration to Integration: Contextualizing Mobile Map App Interactions	76
4.2.1. Exploratory Analyses of Mobile Map Interaction Across Contexts . .	77
4.2.2. Core Analyses: Integrating Behavioral, Intrinsic, and Extrinsic Features	84
4.3. Modeling Next Mobile Map Interaction with Context-Aware LSTM Fusion Model	96
4.3.1. Overall Predictive Performance	97
4.3.2. Contribution of the Context Branch: Autoencoder vs. Full Vector .	98
4.3.3. Interpretability of Context Features	102
5. Discussion	107
5.1. Achieving Ecological Validity in Mobile Map Research	107
5.1.1. Effect of Minimal Enrichment: Evidence from MoT I and MoT II Data Collection	108
5.1.2. Implications	110

5.1.3. Limitations	110
5.2. Mobile Map Interactions are Jointly Shaped by Who Navigates and Where They Navigate	111
5.2.1. Across-Session Mobile Map Interaction Metrics: Stability Over Time	111
5.2.2. Effects of Individual Traits on Mobile Map Interactions	112
5.2.3. Effects of Urban Forms and Individual Traits on Mobile Map Interactions	112
5.2.4. Implications	113
5.2.5. Limitations	114
5.3. Feasibility of Explainable and Enriched Next Mobile Map Interaction Prediction	115
5.3.1. Person–Place Context in Mobile Map Interaction Prediction	115
5.3.2. Latent Dimensions in Person–Place Mobile Map Interaction Prediction	116
5.3.3. Implications	117
5.3.4. Limitations	118
5.4. Broader Implications of the Research	118
6. Conclusions and Outlook	120
6.1. Main Findings	120
6.1.1. Parsimonious and Ecologically Valid Data Capture is Possible, but Comes With Trade-Offs	121
6.1.2. Mobile Map Interactions are Jointly Shaped by Who Navigates and Where They Navigate	121
6.1.3. Context-Aware Prediction is Possible and Interpretable	121
6.2. Design Recommendations	122
6.3. Contributions	123
6.4. Outlook	125
Bibliography	128
A. Appendix	144
A.1. Questionnaires	144
A.2. Classification Categories according to the Google Play Store	152

A.3. Applications Used for Navigation-only Filtered Dataset	153
A.4. Supplementary Descriptive Statistics (MoT II, Raw Data)	153
A.5. GPS Data Alignment Supplementary Information	154
A.6. Urban Form: Additional Analyses	155
A.7. Additional Notes for the Exploratory Analyses Results	157
A.8. Additional Notes for the Main Analyses	160
A.9. Context-Aware LSTM Fusion Model	163

List of Figures

1.1.	<i>An everyday navigation scenario: while the mobile map application efficiently recalculates routes, it remains oblivious to the user’s growing disorientation.</i>	3
1.2.	<i>Research objectives.</i>	8
2.1.	<i>The representation of a city layout within the brain symbolizes how environmental knowledge is internalized to support navigation. The labeled nodes highlight four core cognitive functions involved in this process: visual perception (extracting environmental cues), spatial working memory (maintaining and updating spatial representations), spatial reasoning (inferring spatial relationships and planning routes), and decision-making (selecting navigation strategies and adapting to environmental changes).</i>	24
3.1.	<i>Overview of the project workflow, aligning each thesis research question (RQ1–RQ3) with its corresponding methodological stage: data collection, integration, and prediction.</i>	41
3.2.	<i>MapOnTap Android application interface. Left: the privacy dialog shown when the app is launched for the first time. It informs participants about the types of data collected (e.g., touchscreen events with a timestamp, app name, GPS location, and device state) and links to a detailed privacy policy. Right: An example of recorded location entries (city, timestamp, and altitude) logged during GPS-enabled sessions. These interface elements demonstrate the app’s commitment to transparency and unobtrusive design for in-situ behavioral data collection.</i>	43
3.3.	<i>Structure of Tap Data Collection Across Phone Sessions. Schematic representation of the MapOnTap data logging process. Touchscreen interactions (taps) are recorded during each phone session, defined as the period between unlocking and locking the device. The app remains active in the background, passively collecting timestamped tap events for every session over the entire recording period, spanning both MapOnTap campaigns (MoT I: 2019–2022; MoT II: 2023), during which participants contributed continuous data for a minimum of two weeks.</i>	44

3.4.	<i>Map App Interaction Data Integration Pipeline. Raw data from taps, GPS, and questionnaires underwent pre-processing, synchronization, and feature extraction across three contextual domains: behavioral, intrinsic, and extrinsic. Analyses included clustering (grouping interaction patterns), archetypal analysis (identifying representative behavior profiles), correlation (quantifying relationships), and regression modeling (predicting interactions based on contextual variables).</i>	48
3.5.	<i>Architecture of the context-aware LSTM fusion model. The sequence branch (blue on the left) encodes recent app interaction patterns, while the context branch (orange on the right) encodes urban form and individual traits either directly (8D, Without AE) or via a latent 3D autoencoder (With AE). The fusion head (green and red) then integrates these two sources to make the final prediction.</i>	61
4.1.	<i>Top 10 Most Frequently Used Map-Related Apps Across Participants. Google Maps was by far the most widely used app, whereas others were adopted by comparatively fewer participants. The bars indicate the number of participants who used each app at least once during the study period.</i>	70
4.2.	<i>Top 10 Most Frequently Used Navigation-Relevant Apps in the MoT II Dataset. Counts represent the number of participants using each app at least once during the study period. Google Maps dominated usage, while apps such as DB Navigator (official Deutsche Bahn rail planner) and FAIRTIQ (Swiss public transport ticketing) illustrate country-specific adoption.</i>	73
4.3.	<i>Kernel Density Estimates (KDEs) of Taps per Session as a Function of Distance From Home in MoT I. Each panel shows the interaction patterns for one participant, comparing map-apps (Maps and Navigation + Travel and Local; blue) and other apps (orange). The three recurring interaction profiles illustrated are: (a) near-home-dominant, with peak activity at 0–3.5 km; (b) bimodal, with two peaks at 0–3.5 km and 100–200 km; and (c) travel-dominant, with greater activity at longer distances.</i>	78
4.4.	<i>MoT I: Unsupervised Profiles of Distance-Related Tapping Behavior. Panels (a–d) present the results from the hierarchical clustering analysis (HCA) and archetypal analysis (AA). The HCA dendrogram (a) and mean cluster profiles (b) indicate two main groups: participants whose tapping was concentrated near home and those whose activity extended into travel distances. AA (c, d) further decomposed behavior into four archetypes with two dominant signatures: a compact near-home profile (A_2) and a more dispersed travel-oriented profile (A_3). Together, both approaches highlight a consistent distinction between near-home and travel-related interaction patterns.</i>	81

- 4.5. *Gender-Specific Correlations in the 25–34 Age Band (MoT II).* Among men (top panel), more time spent using map apps per session was positively associated with sense of direction ($r_s = .48$, $p = .046$, 95% CI [.02, .77]). Among women (bottom panel), more frequent app switching per session was positively associated with spatial anxiety ($r = .51$, $p = .021$, 95% CI [.09, .78]). These findings suggest that navigation app use patterns may relate differently to spatial traits across genders. 83
- 4.6. *Relationships Between Inter-Map Intervals (IMI) and Trait Measures.* Scatter plots show Santa Barbara Sense of Direction (SBSOD) scores (left) and Spatial Anxiety (SA) scores (right) plotted against $\log_{10}$ -transformed 25th percentile IMI values. Correlations: SBSOD, $r = .15$, $p = .285$; SA, $r = -.24$, $p = .083$. Lines represent least-squares fits; shaded areas indicate 95% confidence intervals. 87
- 4.7. *Relationships Between Session Initiation Rate (SIR) and Trait Measures.* Scatter plots show Santa Barbara Sense of Direction (SBSOD) scores (left) and Spatial Anxiety (SA) scores (right) plotted against SIR, defined as the inverse of the mean $\log_{10}$ -transformed inter-map interval. Correlations: SBSOD, $r = -.27$, $p = .050$; SA, $r = .05$, $p = .730$. Lines represent least-squares fits; shaded areas indicate 95% confidence intervals. 87
- 4.8. *Relationships Between Session Initiation Rate (SIR) and Santa Barbara Sense of Direction (SBSOD) Scores Stratified by Gender.* Scatter plots show SIR plotted against SBSOD for women (left) and men (right). Correlations: women, $r = -.46$, $p = .024$; men, $r = .15$, $p = .413$. Lines represent least-squares fits; shaded areas indicate 95% confidence intervals. 88
- 4.9. *UMAP Embedding of Session-Level Metrics (CV-ITI, Taps/s, Transition-s/min) with DBSCAN Clusters.* The two-dimensional projection shows six distinct clusters (colored regions) covering 69% of valid sessions, while the remaining sessions were classified as noise (label -1, not shown). One large cluster (Cluster 0) emerged, while the others were smaller and more compact, clearly separated in the embedding space. Axes represent UMAP components in arbitrary units. 90
- 4.10. *Standardized Regression Coefficients and 95% Confidence Intervals From the Robust Models Predicting Interaction Dynamics.* Panel (a) shows predictors of interaction frequency (tap rate), where spatial anxiety, orientation entropy, and degree centrality were significant negative predictors. This indicates that fewer taps occur in more anxious individuals and in more complex or connected environments. Panel (b) shows predictors of interaction regularity (CV-ITI), where spatial anxiety, sense of direction, and male gender were linked to greater irregularity, while closeness centrality predicted more regular interactions. Gray bars represent 95% confidence intervals. 94
- 4.11. *Moderation Effects on Tap Rate.* Panel (a) shows that the influence of sense of direction on tap rate decreased with higher spatial anxiety. Panel (b) shows that sense of direction predicted tap rate more strongly in environments with higher orientation entropy (greater street network complexity). 95

- 4.12. *Moderation Effects on Interaction Regularity (CV-ITI). Panel (c) shows that spatial anxiety amplified the effect of network circuitry on interaction irregularity. Panel (d) shows that sense of direction was more strongly linked to CV-ITI in environments with higher closeness centrality.* 95
- 4.13. *Confusion Matrix for the Selected Augmented LSTM Model at $p = 0.8$. The figure shows that the model correctly identified most non-map taps (true negatives) and more than half of the real map taps (true positives), although many map taps were still missed (false negatives). Model configuration: time-step masking = 0.20, Gaussian noise $\sigma = 0.15$, feature dropout = 0.00, hidden size = 16.* 98
- 4.14. *Receiver Operating Characteristic (ROC) Curve for the Selected Augmented LSTM Model Evaluated on the Held-Out Test Set. The curve plots the true positive rate (TPR) against the false positive rate (FPR) across all classification thresholds. The area under the curve (AUC) is 0.743, showing that the model distinguishes map from non-map taps better than chance, although the performance is moderate. Model configuration: time-step masking = 0.20, Gaussian noise $\sigma = 0.15$, feature dropout = 0.00, hidden size = 16.* 99
- 4.15. *Pearson Correlation Coefficients Between Autoencoder (AE) Latent Dimensions and the Original Contextual Features. The three latent dimensions showed distinct and interpretable patterns: AE latent 1 was negatively correlated with orientation entropy and positively correlated with closeness and spatial anxiety; AE latent 2 was negatively correlated with gender and spatial anxiety but positively associated with sense of direction; AE latent 3 was strongly negatively correlated with betweenness and degree, and moderately associated with closeness.* 100
- 4.16. *Confusion Matrices on the Development Set for the Fusion Head Using (a) the Full Eight-Dimensional Standardized Context Vector (Without AE) and (b) the 3D Autoencoder Embedding (With AE). Counts are raw predictions; map is treated as the positive class. The With AE configuration produced fewer false positives for map predictions than the Without AE configuration, with comparable false negative counts.* 102
- 4.17. *SHAP Summary Plots for Context Inputs in the Fusion Head (With AE configuration). The horizontal axis indicates SHAP values, where positive values increase the probability of predicting a map tap and negative values decrease it. Color encodes the feature value (blue = low, red = high). Features are ordered by average absolute SHAP value, with the most influential at the top. Panel (a) shows all samples, highlighting that AE latent 1 and AE latent 2 carried most of the predictive weight. Panels (b) and (c) show class-specific impacts: AE latent 2 was the strongest positive driver of map predictions, while AE latent 1 more often reinforced non-map predictions.* 103

- 4.18. *Positive and Negative SHAP Contributions for the With AE Fusion Model on the Development Set. Bars show the share of the total SHAP impact, separated into positive contributions (towards predicting map) and negative contributions (towards predicting non-map). Panel (a) shows the aggregation of all samples, where AE latent 1 and AE latent 2 accounted for most of the model's influence, while AE latent 3 played only a minor role. Panels (b) and (c) split the analysis by class: AE latent 2 was the strongest positive driver for map predictions, whereas AE latent 1 provided the largest negative evidence for non-map predictions.* 104
- 4.19. *Map Recall Across the Three AE Ablation Variants (Full AE, No Traits: Urban only, and No Urban: Traits only) on the Development Set. The figure shows that recall was highest for the Full AE (0.570) and Traits-only (0.567) models, and slightly lower for the Urban-only model (0.548), suggesting that traits were particularly important for sustaining recall of map events. Values above the bars correspond to recall.* 105
- A.1. *Distribution of participants by country. The dataset includes participants from 11 countries: Switzerland (23), Italy (9), France (3), Austria (3), United Kingdom (2), United States (1), Germany (1), Hungary (1), Pakistan (1), Poland (1), and the Netherlands (1).* 155
- A.2. *SHAP summary plots for context inputs in the fusion head (Without AE configuration). Features are ordered by mean absolute SHAP value, with the most influential at the top.* 168
- A.3. *Positive versus negative SHAP contributions for context features in the fusion head (Without AE configuration).* 169
- A.4. *MAP recall across the three 8D ablation variants on the development set: Full model, No Traits (Urban only), and No Urban (Traits only). Recall was highest for the Full model (0.65), followed by the Urban-only model (0.55), with the Traits-only model slightly lower (0.53). This pattern indicates that urban features contributed more to recall of map events than trait features. .* 171

List of Tables

3.1.	<i>Overview of data types collected during the two MapOnTap campaigns (MoT I and MoT II). Note: SBSOD, SA, and demographic data were collected using external questionnaires, not through the MapOnTap app.</i>	46
3.2.	<i>Overview of tappigraphy features extracted within phone sessions and across participants. Session-based features capture the detailed interaction dynamics within individual phone sessions. Participant-based features summarize users' overall interaction patterns throughout the entire recording period, facilitating comparisons across individuals with different navigation behaviors.</i>	51
3.3.	<i>Participant-Based Intrinsic Features Derived From Questionnaires.</i>	52
3.4.	<i>Urban Form Features Derived From Localized Street Networks</i>	53
4.1.	<i>Participant Coverage in the MoT I Campaign</i>	67
4.2.	<i>MoT I Tappigraphy Statistics by Phone Session</i>	68
4.3.	<i>MoT I Tappigraphy Statistics Aggregated by Participants</i>	68
4.4.	<i>Within-Participant Mean Taps per Session Across Map-Related App Categories. Participants engaged in more intensive interaction with Travel & Local (TL) apps than with Maps & Navigation apps (MN), as reflected in the higher median and mean values.</i>	69
4.5.	<i>GPS Usage Statistics Across Participants. Participants exhibited highly variable GPS engagement, with some recording only a handful of sessions and others exceeding 10,000. On average, sessions were short in duration, but the total number of GPS points per participant was substantial.</i>	71
4.6.	<i>Participant Coverage in the MoT II Campaign. Although participants typically contributed about 16–18 days of tappigraphy and GPS data, coverage varied substantially across the sample, with some providing only minimal data and others extending their data contribution to several months.</i>	71

4.7. <i>Tappigraphy Statistics at the Session Level in the MoT II Navigation-Filtered Cohort. Navigation sessions were generally short (see Table 4.7, with a median duration of approximately 51 seconds (less than 1 minute), while the longest sessions extended to over 4 hours. Map-only sessions were much shorter (median ≈ 12 seconds, maximum ≈ 26 minutes) and accounted for 19.49% of all navigation sessions.</i>	73
4.8. <i>Tappigraphy Statistics Aggregated by Participant in the MoT II Navigation-Filtered Cohort. Participants varied widely in their use of map apps: while some engaged with only a few sessions, others accumulated hundreds of them. Map-only sessions constituted a notable but highly variable share of participants' overall activity.</i>	74
4.9. <i>GPS Statistics in the MoT II Navigation-Filtered Cohort. Participants exhibited highly variable GPS engagement, with some recording only a few sessions and others exceeding 2000. On average, sessions were short in duration, but the total number of GPS points per participant was substantial.</i>	74
4.10. <i>Descriptive Statistics for Individual Difference Measures in the MoT II Cohort. Sense of direction scores were moderately high on average, while spatial anxiety scores indicated generally moderate levels across participants.</i>	75
4.11. <i>Descriptive Statistics for SBSOD and Spatial Anxiety (SA) Scores by Gender. Women reported a poorer sense of direction and higher spatial anxiety than men.</i>	75
4.12. <i>Descriptive Statistics for Participant-Level Variables ($N = 55$). The table shows that IMI values were highly skewed, ranging from only a few seconds to several days, while SIR values were more stable across participants. Trait scores also showed variability, with SBSOD spanning most of the 1–7 scale and SA ranging from low to moderately high values.</i>	86
4.13. <i>Descriptive Statistics for Session-Level Interaction Metrics ($N = 3667$ Sessions). Tapping regularity (CV-ITI) exhibited substantial variability across sessions. In contrast, interaction frequency (taps per second) and switching behavior frequency (transitions per minute) were relatively low. Only a few sessions exhibited much higher values.</i>	90
4.14. <i>Descriptive Statistics of Spatial Anxiety and Sense of Direction Scores for Participants Grouped into DBSCAN Clusters (Noise Cluster Excluded). Values show the number of sessions (n), means (M), and standard deviations (SDs) for each cluster. The table illustrates the slight variations in average trait levels across clusters, with overlapping distributions.</i>	91

4.15. <i>Classification Threshold Analysis of the Test Set for the Augmented LSTM Model. The table shows how changing the decision threshold (p) shifts the balance between recall and precision. Lower thresholds capture almost all map taps (high recall) but produce many false alarms (low precision), while higher thresholds increase precision but miss more true events. The operating point chosen for this study was $p = 0.8$, which offered the best trade-off between sensitivity and specificity. Model configuration: time-step masking = 0.20, Gaussian noise $\sigma = 0.15$, feature dropout = 0.00, hidden size = 16.</i>	97
4.16. <i>Overall Performance of the Fusion Head on the Development Set. The model using the full eight context features (Without AE) achieved higher accuracy, precision, recall, F1, and ROC-AUC than the autoencoder-compressed version (With AE), indicating that the uncompressed context provided stronger predictive power.</i>	101
4.17. <i>Per-Class Performance (Map vs. Non-Map) on the Development Set. The autoencoder version (With AE) showed higher precision for the map class, but at the cost of much lower recall compared to the full feature model (Without AE). For the non-map class, both models performed similarly well, with very high precision and recall.</i>	101
A.1. <i>Application Categories in the Google Play Store ($n = 33$).</i>	152
A.2. <i>Mobile Map Applications used when filtering MoT II Dataset for Navigation only</i>	153
A.3. <i>Participant-Level GPS Descriptive Statistics (filtered sample, $N = 61$ participants)</i>	154
A.4. <i>Participant-Level Tappigraphy Descriptive Statistics (MoT II, $N = 61$ participants)</i>	154
A.5. <i>GPS Data Descriptive Statistics Before and After Alignment ($N = 60$ participants)</i>	155
A.6. <i>SMOD (2020, 1 km) classification of participant session locations.</i>	156
A.7. <i>Sensitivity of urban form metrics to buffer size. Descriptive statistics are reported for 200 m, 500 m, and 1000 m buffers.</i>	157
A.8. <i>Correlation Coefficients between Santa Barbara Sense of Direction scores (SBSD), Spatial Anxiety (SA) scores, and the extracted variable linked to interactivity behaviour. Significant correlations are highlighted in bold, with asterisks denoting p-value thresholds ($*p < 0.05$).</i>	158
A.9. <i>Correlation Coefficients for the Male Dataframe, between Santa Barbara Sense of Direction scores (SBSOD), Spatial Anxiety (SA) scores, and the extracted variable linked to interactivity behavior. Significant correlations are highlighted in bold, with asterisks denoting p-value thresholds ($*p < 0.05$).</i>	159

A.10.	<i>Correlation Coefficients for the Female Dataframe, between Santa Barbara Sense of Direction scores (SBSD), Spatial Anxiety (SA) scores, and the extracted variable linked to interactivity behaviour. Significant correlations are highlighted in bold, with asterisks denoting p-value thresholds (*$p < 0.05$).</i>	160
A.11.	<i>Kruskal–Wallis tests across DBSCAN clusters (clusters 0–5; noise excluded).</i>	161
A.12.	<i>Hidden-size comparison (oversampling only; threshold $p = 0.5$). Test set $N = 5,290$; positives = 494, negatives = 4,796.</i>	166
A.13.	<i>Overall development set performance for fusion head ablations (With-AE) .</i>	169
A.14.	<i>Per-class development set performance for fusion head ablations (With-AE)</i>	170
A.15.	<i>Overall model performance in ablation analyses (8D, Without-AE)</i>	170
A.16.	<i>Per-class model performance in ablation analyses (8D, Without-AE)</i>	170

1

CHAP -
TER

Introduction

“The highest form of understanding we can achieve are laughter and human compassion.”

Richard P. Feynman

Imagine having to attend a job interview in an unfamiliar urban area. Like most people today, you would probably instinctively reach for your smartphone. Mobile navigation apps have become so ubiquitous and embedded in our daily lives and mobility habits (Savino et al., 2019) that it feels entirely natural to open a map app, enter the address of the office and start the journey.

Navigating the first few hundred meters is easy; however, you soon find yourself in the historic district, paved with narrow and uneven streets, where every intersection looks the same. It is often more challenging to navigate such areas, as irregular street layouts provide fewer stable reference points to rely on, making the surroundings harder to interpret and understand (Lynch, 1960; Montello, 2005). You have never had a strong sense of direction and, like others with weaker orientation skills, you tend to rely more heavily on external aids, such as maps or GPS (He & Hegarty, 2020). To stay oriented, you repeatedly check your progress, tapping the screen to reopen the app, in between flicking through emails or queuing up music to motivate yourself for the interview. Such frequent checking and app switching represents a typical habit observed in smartphone users (Deng et al., 2019; Oulasvirta et al., 2012), with maps rarely being used in isolation.

At one intersection, you realize that you have taken a wrong turn. The application’s instant recalculation of the route shows you how to proceed, and this feels reassuring; the mistake disappears, and your journey can resume. However, the application does not fundamentally understand what is happening. It does not recognize that your repeated taps reveal orientation difficulties or that your constant app switching reflects growing spatial anxiety. If the application could detect such signs and adapt its guidance — for example, by simplifying the map display to emphasize perceptually salient and task-relevant information (Fabrikant et al., 2010; Hegarty et al., 2010; Reichenbacher, 2009) or by highlighting key landmarks to help you stay oriented (Kapaj et al., 2024) — it could support attention, and alleviate emotional strain. This would help you consolidate confidence in your spatial abilities. But this does not happen. Instead, with every instant recalculation, the application provides reassurance that may, over time, encourage greater dependence on the mobile map and foster a habitual reliance on automated guidance (Dahmani & Bohbot, 2020).

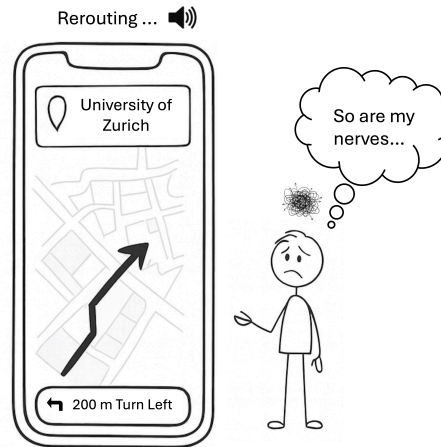
Now, imagine relying on this type of navigation support everyday, not only for unfamiliar routes but also for your daily travel to work, shopping, or visiting friends. Over time, using mobile map applications in this way has measurable consequences. Evidence from lab and field studies indicates that, while useful for immediate orientation, frequent reliance on such applications can weaken spatial learning and environmental knowledge acquisition (Chrastil & Warren, 2012; Dahmani & Bohbot, 2020; Gardony et al., 2015; Ishikawa et al., 2008; Parush et al., 2007; Ruginski et al., 2019). While individuals successfully reach their destinations when using these tools, they often learn less about the surrounding environment (Brügger et al., 2019; Dahmani & Bohbot, 2020; Hejtmánek et al., 2018; Ishikawa et al., 2008). The extent of this effect depends not only on individual differences, but also on the complexity of the environment. For those with a weak sense of direction or higher levels of spatial anxiety, these traits will shape how you interact with the application and determine how much spatial knowledge you retain (He & Hegarty, 2020). When these traits are combined with a confusing urban layout and irregular streets, the challenge becomes even greater (Coutrot et al., 2022; Montello, 2005).

The above scenario highlights a deeper scientific issue that mobile map applications are not designed to adapt to user interaction behavior. Consequently, these applications disregard various factors that influence how people interact with navigation applications, such as individual differences in spatial abilities, emotional traits, and the influence of the urban environment layout.

Current adaptive features in mobile map applications mostly depend on explicit user input, for instance, manually selecting a preferred map style, mode of travel, or set of points of interest. Although such settings reflect conscious preferences, they capture only a narrow slice of user needs during navigation. In reality, navigation behavior is deeply context-dependent and changes moment to moment (Bartling et al., 2021). As users could alternate between confidence and uncertainty along a route, their requirements for guidance, visual detail, or reassurance also fluctuate. Static preference settings, by contrast, remain fixed and cannot anticipate these shifting demands (Bartling et al., 2022).

What these applications are missing is a *predictive layer* that can interpret daily mobile map interactions as implicit proxies for a user’s spatial abilities and emotional states. This layer could relate these interaction patterns to the environment in which navigation occurs. Unlike explicit preferences, which involve users stating what they want, implicit proxies can be inferred directly from how individuals engage with everyday navigation applications (Ballatore & Bertolotto, 2015). Implicit proxies can capture behavioral patterns that users may not consciously recognize or be able to express through the mobile map interface such as feelings of disorientation, rising spatial anxiety, or increased cognitive effort during navigation. However, these behavioral signals gain meaning only when interpreted alongside extrinsic contextual data, such as GPS position, movement dynamics, or environmental complexity. While GPS and sensor data describe where and when navigation occurs, behavioral signals reveal how users engage with the system within those contexts. Integrating these complementary layers creates a more comprehensive foundation for understanding real-world navigation and paves the way for personalization by linking real-world environment, user behavior, and internal states within a single predictive framework. Without the inclusion of such anticipatory design, these systems are limited in their capacity to provide personalized support.

Figure 1.1.: *An everyday navigation scenario: while the mobile map application efficiently recalculates routes, it remains oblivious to the user’s growing disorientation.*



1.1 Motivation and Problem Statement

Over the past 30 years, interdisciplinary research spanning the fields of mobile cartography, geographic information science (GIScience), human–computer interaction (HCI), and cognitive science has established a robust theoretical foundation for the design of mobile maps that support everyday navigation. This body of work emphasizes the need for the **retrieval**, **filtering**, **visualization**, **interaction**, and **prediction** of geographic information to go beyond the mere static display of spatial information (i.e., GPS coordinates).

Early research quickly recognized that, unlike with desktop-based GIS, it was neither possible nor useful for mobile maps to retrieve *all* available geographic for display due to the limited bandwidth of mobile networks, small screen size, as well as limited user attention span (Meng & Reichenbacher, 2005; Raper et al., 2007a; Reichenbacher, 2007). Instead, these applications must be selective, **retrieving** and **filtering** only the information relevant to the user’s current goals, tasks, and situational needs (Crease & Reichenbacher, 2011; De Sabbata, 2013; Emmanouilidis et al., 2013; Raper, 2007; Reichenbacher, 2001, 2007, 2009), preventing them from being overwhelmed with unnecessary detail (De Sabbata, 2013), thereby serving as an attention guiding mechanism that directs users to the most useful information for their current task (Swienty et al., 2008).

Beyond retrieval and filtering, researchers acknowledged that maps are not simply neutral containers of information; they are symbolic artifacts that directly influence perception, memory, and reasoning (Montello et al., 2018). Their **visualization** must therefore be cognitively aligned, for example, by incorporating visual salience and congruent representations (e.g., realistic landmarks) to guide attention and support spatial learning (Cheng et al., 2022; Dahmani & Bohbot, 2020; Fabrikant et al., 2008, 2010; Kapaj et al., 2024; Ruginski et al., 2022).

Mobile devices changed the way people engage with maps, utilizing touch gestures such as tapping, pinching, and zooming, as well as multi-modal inputs such as voice and auditory cues, all of which are experienced on screens much smaller than those of desktop computers (Griffin et al., 2017; Roth et al., 2024). These conditions highlighted that mobile interaction design could not simply shrink desktop conventions to fit a phone screen, but required its own vocabulary of interaction tailored to the mobile context (Roth et al., 2024). Therefore, **interaction** must be designed not merely as a technical channel between user and system, but as part of a dynamic exchange involving the user, the environment, and the process of navigation itself (Griffin et al., 2017). From this perspective, interactions with mobile maps represent more than commands to be processed; they can also reveal something about the underlying individual differences among users (Savino et al., 2021).

Building on this insight, mobile maps have evolved from systems that merely react to the explicit inputs from users toward systems that **predict** these inputs (Huang et al., 2018). Map applications already include predictive features, such as suggesting destinations or nearby points of interest, but they remain focused on efficiency (i.e., minimizing travel time) while ignoring user cognitive and emotional needs, including spatial learning, environmental understanding, and emotional factors like spatial anxiety (He & Hegarty, 2020; Ishikawa et al., 2008; Lyons et al., 2018; Parush et al., 2007). Researchers have long argued that navigation support should extend beyond reactive information delivery to account for users' goals, activities, and situational constraints (Raubal, 2001; Winter, 2001; Worboys, 2001). They emphasize the importance of anticipating people's space-time needs, that is, the information or assistance a person requires at a specific location and moment, given their ongoing task and movement context, rather than only responding simple to requests (Raubal et al., 2004). More recently, map personalization research has built directly on this principle, proposing that interaction patterns themselves can serve as implicit indicators of user intent, preferences and cognitive states and that maps should adapt both their content and their presentation according to these patterns in real time (Ballatore & Bertolotto, 2015).

To enable this anticipatory design, mobile maps must draw on a broader and richer concept of **context**, one that is not limited to geographic information but encompass any information that characterizes a user's situation (Dey, 2001). This includes the user's intentions, behavior, and cognitive and emotional states during interaction with the application, which change according to their goals (De Sabbata, 2013; Emmanouilidis et al., 2013; Raubal, 2001).

Operationalizing this multifaceted view of context within an integrated framework remains a central challenge in mobile map design. Most applications rely on one-size-fits-all representations that treat all users identically, failing to account for their individual goals, traits, or environments. However, researchers have recently been calling for context-aware applications that can adjust to different user characteristics in real time. These applications should adapt to users current activity, such as searching for a specific address, exploring an unfamiliar area, or looking for nearby services, as well as their cognitive and emotional states, and the level of automation in the navigation process, that is, how much control or decision-making is delegated to the system versus retained by the user (Fabrikant, 2023; Griffin et al., 2017, 2023; Kiefer et al., 2017; Roth et al., 2023, 2024).

These research agendas have been shaped by broader developments in mobile application personalization, including context-aware recommender systems, targeted advertising, and personalized content delivery. These developments have established the potential for mobile

maps that could dynamically adapt to user information. These navigation tools would not only suggest destinations, routes, or points of interest, but also leverage contextual clues and implicit behavioral signals derived from user interactions with the mobile map to proactively predict map content, structure, and interaction mode (Ballatore & Bertolotto, 2015).

Bartling et al. (2023) present a conceptual framework that summarizes and integrates fundamental contributions from previous theories and the latest research developments. The authors propose a three-level cartographic context taxonomy that systematizes the multiple dimensions that influence daily mobile map use. These dimensions, presented below, should be considered when designing mobile navigation applications (Bartling et al., 2023):

Extrinsic context: environmental and technological factors (e.g., brightness, noise, GPS quality, device type).

Intrinsic context: cognitive, emotional, and motivational traits (e.g. spatial anxiety, sense of orientation, cognitive load).

Behavioral context: interaction patterns that can be observed in real time, including touches, pauses, and recurring gestures.

Despite being firmly anchored in a solid theoretical foundation, the implementation of these principles in mobile navigation application design remains largely unexplored.

Specifically, the three contextual dimensions—extrinsic, intrinsic, and behavioral—have not been integrated within a single computational framework capable of predicting user needs and reactions in real time. This gap primarily stems from the fragmentation of empirical and applied research, which continues to explore these dimensions separately rather than examining them as interconnected elements.

For example, most predictive models used in applied research applications focus almost exclusively on the extrinsic context (Huang et al., 2018). These models exploit easily accessible environmental data, such as GPS coordinates, traffic conditions, and points of interest (POIs), through smartphone technology (Huang et al., 2018; Schiller & Voisard, 2004). While this information facilitates the development of efficient point-to-point navigation algorithms, it neglects the cognitive and emotional dimensions fundamental to personalization. Additionally, behavioral context data are collected and used, especially in commercial applications, but are rarely shared with academic researchers (Ballatore & Bertolotto, 2015).

Empirical research focused on the intrinsic context, however, has demonstrated that the prolonged use of navigation applications can impair users’ spatial memory, environmental learning, and orientation abilities (McKinlay, 2016; Ruginski et al., 2022). Individual differences, such as spatial anxiety, sense of orientation, and gender, modulate the impact of these technologies, influencing the degree of a user’s reliance on GPS and the navigation strategies they adopt (He & Hegarty, 2020; Lyons et al., 2018).

Despite its important results, this line of research is primarily carried out in controlled and simplified contexts (both indoors and outdoors), making it difficult to generalize its findings to the real, complex environments in which these applications are used. Furthermore, these

studies fail to consider how the complexity of urban morphology also modulates interactions with mobile map applications, focusing only on the cognitive aspects of navigation (Coutrot et al., 2022; Watts et al., 2015; Xu et al., 2022). This is despite factors such as road network configuration having a profound influence on the cognitive processes involved in spatial learning and the mental representation of space (Coutrot et al., 2022; Lynch, 1960; Montello, 2002). However, these factors are rarely considered in studies of spatial cognition and individual differences, limiting the ecological validity of the results.

Finally, research on the behavioral context remains limited. Most HCI studies only address interaction indirectly, typically by examining the outcomes of specific navigation application design elements. For example, Lin et al. (2017) showed how mismatched visual and audio instructions, as well as errors in route prediction or map visualization, can result in users making critical navigation mistakes with potentially life-threatening consequences. This is largely due to users' over-reliance on automated guidance even in the presence of conflicting environmental cues.

However, this research seldom involves **systematically collecting and modeling detailed behavioral interaction data** during real-world navigation. This is primarily because obtaining granular data on how users tap, pan, zoom, and pause within commercial map apps is both technically challenging and subject to strict privacy constraints.

A rare exception is the study by Savino et al. (2021), in which the authors developed a web-based wrapper app capable of unobtrusively recording users' interactions with Google Maps in detail during everyday navigation. This approach enabled the capture and analysis of the types of interaction users frequently adopt when deploying map applications for real-world navigation (e.g., exploration, zoom), revealing distinct patterns of map use. However, this study's scope was restricted to browser-based usage rather than full-featured mobile apps, limiting the extent to which the findings could be generalized. Consequently, a critical gap remains regarding the empirical study of behavioral context in mobile map use, especially in ecologically valid, naturalistic settings.

In summary, the persistent segregation of research on extrinsic, intrinsic, and behavioral contexts remains a major limitation for advancing mobile map design. Before adaptive, personalized navigation systems can be realized, the field must implement the key steps of capturing, integrating, and modeling everyday, real-world mobile map navigation.

1.2 Research Gaps

There are three key gaps in the research that hinder advances in mobile map design. Each reflects a missing link in the collection, integration, and modeling of **extrinsic**, **intrinsic**, and **behavioral** contexts.

While existing applied research on mobile map application design makes extensive use of extrinsic signals such as GPS or traffic conditions, it provides little evidence on how these external factors interact with intrinsic traits and behavioral patterns during everyday navigation. The absence of **ecologically valid datasets** that capture all three dimensions together in-situ prevents the creation of an accurate representation of real-world navigation.

Gap 1: Lack of Ecologically Valid Data across Contexts

The lack of ecologically valid datasets that combine extrinsic, intrinsic, and behavioral dimensions of real-world navigation leave theories of user-centered map design disconnected from practice.

Addressing this first gap requires experimental designs that move beyond controlled laboratory or scripted field tasks. Such designs should capture users' everyday interactions with mobile maps in situ, combining interaction logs with measures of spatial skills and emotional traits. Only this type of ecological grounding can enable the meaningful testing of theories of user-centered cartography against real-world navigation.

While research has made progress in understanding the roles of individual abilities, environmental structure, and mobile map use from a cognitive science perspective, these aspects have rarely been examined in direct relation to interaction with mobile map applications on smartphones. Therefore, these strands of work remain largely isolated. The lack of integration across these contextual levels aligns with a second gap:

Gap 2: Lack of Integration Across Contexts

Extrinsic, intrinsic, and behavioral dimensions are rarely incorporated within a unified framework, making it difficult to explain how the surrounding environment, individual traits, and user-application interaction jointly shape real-world navigation.

Addressing this second gap necessitates the development of frameworks that move beyond isolated analyses and explicitly model the interplay among extrinsic, intrinsic, and behavioral contexts. Only by integrating these dimensions into shared representations can research begin to explain how they interact dynamically during everyday map use.

The third gap follows directly from the first two. Even when partial data exist and fragmented insights are available, current systems focus on **explicit inputs** or **efficiency-driven predictions** such as estimated time-of-arrival (ETA) or the next destination. While current applications can predict *where* a user wants to go, they rarely anticipate *how* an individual navigates, *what state* they are in or *what kind of support* they might require along the way. Missing from these tools is a predictive modeling layer that combines extrinsic, intrinsic, and behavioral contexts to anticipate map interaction patterns.

Gap 3: Lack of Predictive Model of Map Interaction

Predictive models that integrate extrinsic, intrinsic, and behavioral contexts to anticipate how users interact with maps are largely absent, limiting progress in adaptive design.

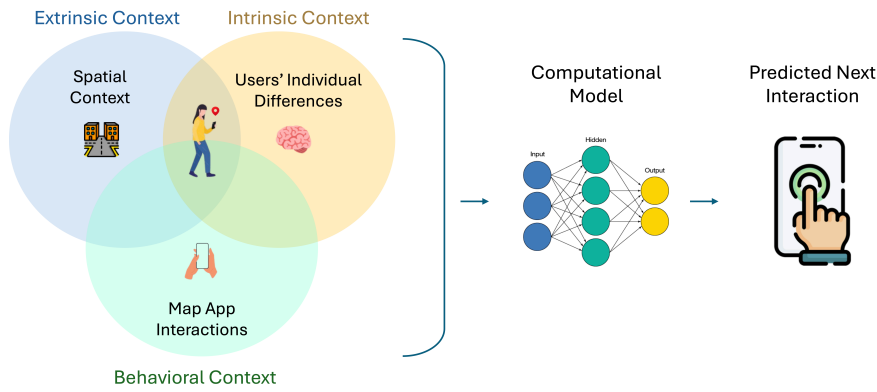
Tackling this final gap demands computational approaches that interpret user interaction patterns and context in mobile map applications. The goal is to transition from static tools to dynamic systems that adapt in real time to not only where the user is, but also how they are navigating, what their goal is likely to be, and why they need to use the map in that moment. By modeling behavioral patterns over time, these approaches can provide information about when and how the interface should change, for example, by adjusting the level of detail, the type of feedback produced, or the route guidance mode. This would

enable mobile maps to respond more effectively to individual users' spatial abilities and emotional states, as well as the specific demands of navigation tasks.

1.3 Research Objective

The goal of this thesis is to develop a computational model that predicts a user's future interaction with mobile map applications during navigation, considering their individual differences, patterns of prior mobile map application use, and the spatial context in which the interaction occurs.

Figure 1.2.: *Research objectives.*



This model is informed by the integration of three key dimensions (as illustrated in Figure 1.2):

1. **Map App Interaction:** Understanding users' historical interaction patterns with mobile map applications, including the rhythm, frequency, and style of their interactions over time.
2. **Individual Differences:** Incorporating user-specific characteristics, including sociodemographic factors, spatial abilities, and emotional responses, that influence how individuals interpret and engage with mobile map applications used for navigation.
3. **Spatial Context:** Situating individual differences and interaction behavior within the user's spatial surroundings, taking into account characteristics of the built environment, the route structure, and the geographic characteristics of the navigation setting.

1.4 Research Questions and Approach

To structure this research and achieve its objective, this dissertation is organized around the following research questions.

- **RQ1:** How can mobile map interaction data be collected in situ to capture everyday navigation behavior in real-world contexts?
- **RQ2:** How can mobile map interactions be meaningfully linked to individual differences and spatial context?
- **RQ3:** How can a user's individual traits, prior usage behavior, and spatial context be used to predict their future interactions with mobile map applications during navigation?

To address these questions, this dissertation employs a three-part methodological strategy for both collecting real-world data and carrying out predictive modeling.

The first part focuses on developing an ecologically valid data collection framework to capture individuals' interactions with mobile map applications in everyday settings. This includes gathering behavioral data on how users engage with mobile maps and environmental information such as location and movement, based on the principles of ecological momentary assessment (EMA) and mobile sensing (Montag & Baumeister, 2023).

The second part focuses on the translation of raw behavioral data from smartphone use and its integration with intrinsic (individual-level) and extrinsic (environmental) contextual features. This integrative step is essential for transforming isolated data streams into a unified computational model that reflects the interactions among users' behavior, cognitive traits, and their surrounding urban environment.

The third part addresses the challenge of predicting the future interactions between users and the mobile map interface. Patterns are identified in the data using a sequential modeling approach, enabling the system to anticipate future behaviors based on user traits, spatial context, and past interaction history.

Combined, these methodological steps facilitate the development of a computational model that integrates behavioral, intrinsic, and extrinsic contextual data from real-world navigation. This model provides a predictive framework for understanding how people interact with mobile maps in everyday contexts, establishes a solid foundation for advancing the design of context-aware mobile map applications.

2

CHAP -
TER

Related Work

*“No social stability without
individual stability.”*
*Aldous Huxley, Brave New
World*

This chapter, organized into two parts, reviews the theoretical and empirical foundations of mobile map application design.

The first part synthesizes 30 years of research in mobile cartography, geographic information science (GIScience), HCI, and cognitive science, illustrating how it has shaped the theory of mobile map design 2.1. This review is structured around the following four themes: (i) **retrieval and filtering** (Section 2.1.1), (ii) **visualization** (Section 2.1.2), (iii) **interaction** (Section 2.1.3), and (iv) **prediction** (Section 2.1.4). Collectively, these themes demonstrate that the design of mobile maps is not simply a matter of processing signals such as GPS coordinates or street layouts (*extrinsic context*); it also involves users’ cognitive abilities and emotional states (*intrinsic context*), as well as the behavioral traces of their interactions with the mobile application (*behavioral context*). Research programs have built on this solid conceptual foundation for user-centered mobile mapping over the years, exploring how emerging technologies could help bring this theory to applied practice (Section 2.1.5).

The second part examines why these themes are only partially realized in everyday applications (see Section 2.2). It sets out the three areas that must be *operationalized* to achieve context-aware adaptation: ecologically valid data (Section 2.2.1), the integration of research on intrinsic–behavioral–extrinsic contexts (Section 2.2.2), and user-centered predictive support (see Section 2.2.3), highlighting where current applications fall short.

The chapter concludes by outlining the elements that are still required to realize these goals: an integrated **context-aware** approach that *fuses* intrinsic (user), behavioral (interaction), and extrinsic (environmental) *signals* to anticipate the use of everyday mobile maps in real time (see Section 2.3).

2.1 Theoretical Foundations of Mobile Map Design

Understanding the conceptual foundations established by previous research is essential for identifying existing gaps. Initially, scholars recognized that designing for a phone was fundamentally different from designing for a desktop GIS. By the late 1990s, desktop computers could display virtually unlimited layers of spatial data, but this was neither

possible nor desirable on a mobile screen. Small displays, limited bandwidth, and the need to navigate while walking down a busy street rendered comprehensive data dumps useless. The challenge was not in displaying as much information as possible, but in deciding which information mattered at a given moment. This realization resulted in researchers turning their attention to the first major theme of mobile map design: **retrieving and filtering** spatial information.

2.1.1 Retrieval and Filtering Spatial Information

When map services began to migrate from desktops to mobile devices in the early 2000s, researchers quickly identified the emergence of new design problems: information overload, visual clutter, ambiguous or inaccurate representations, and inadequate management of contextual relevance (Reichenbacher, 2001, 2005). All of these issues could hinder users' decision-making processes while on the move (Crease & Reichenbacher, 2011; De Sabbata, 2013; Reichenbacher, 2004b, 2009).

A primary limitation associated with these problems concerns how context is defined and operationalized. Depending on the discipline in which it is used, the concept of context is elusive and has no single definition (Hong et al., 2009). However, from an informatics perspective, and specifically related to location-based services (LBS), one of the most influential characterizations remains that proposed by (Dey, 2001). According to this definition, context is **any information that can be used to characterize the situation of an entity, where the entity can be a user, a device, a location, or an object** (Dey, 2001). Many researchers have built upon this definition to emphasize the importance of operationalizing context for information retrieval in LBS, particularly for GPS navigation applications.

Many scholars have highlighted that knowing a user's exact location is insufficient to define their context. Understanding the context in terms of activity, intentions, cognitive abilities, and emotional state is essential to provide truly relevant, usable, and adaptive information during navigation (Reichenbacher, 2007).

Researchers have criticized the tendency of existing systems to implement a narrow operationalization of context, equating it with observable inputs such as GPS coordinates or user-provided tags (Emmanouilidis et al., 2013). This limits the context to the geographical aspect alone. Scholars argued that this "geography-only" definition of context was too restrictive, as it ignored the richer dimensions of activity, mobility, and preference, which shape the user's actual informational needs (Reichenbacher, 2009; Schmidt et al., 1999). These issues sparked a significant theoretical shift centered on geographic relevance (GR), which extended the principles of textual information retrieval (IR) to the spatial domain (Raper, 2007; Reichenbacher, 2007; Reichenbacher & De Sabbata, 2011). According to this perspective, the relevance of a geographic entity (for example, a POI on a map) is not fixed; instead, it is dynamically constructed based on the context in which the individual is utilizing the map (Reichenbacher & De Sabbata, 2011). This theory departs from the semantic focus of IR, introducing an operational definition of context as a central design dimension.

Thus, we speak of not only the geographical context (the physical and spatial characteristics of the environment) but also the context of the user. The latter includes the individual's objectives, activities (Kaptelinin & Nardi, 2009), mobility (Hägerstrand, 1970; H. J. Miller & Bridwell, 2009), and preferences. The concept of GR was then developed based on

these factors (De Sabbata, 2013). Over the years, a redefinition of spatial information retrieval as a context-sensitive and user-aware process has emerged from these theoretical contributions.

The GR concept has introduced a significant conceptual shift, moving the focus from filtering based solely on the user's location to a more nuanced understanding of what makes a geographic object useful in a given situation. From this perspective, relevance is not intrinsic to geographic information but is dynamically constructed in relation to the user's goals, constraints, and activities in the context in which they find themselves (De Sabbata, 2013).

However, while this approach has significantly advanced our understanding of how to retrieve and filter spatial information, it is important to recognize that relevant information is insufficient if the user does not perceive and interpret it correctly. For example, Swienty et al. (2008) partially demonstrated how the usefulness of geographic information depends on cognitively appropriate visualization. Their attention-guiding geo-visualization framework combines GR with human visual processing principles to reduce cognitive load, improve attention guidance, and support decision making (Swienty et al., 2008). This supports the notion that filtering by relevance alone is inadequate in mobile contexts characterized by time constraints and perceptual limitations (Cheng et al., 2022; Roth et al., 2017). To adapt to users, spatial information retrieval must integrate the principles of cartographic visualization. This is a key aspect of supporting perception, memory, and decision-making in real-time navigation contexts (Fabrikant et al., 2010; Montello et al., 2018).

These limitations highlight the idea that identifying contextually relevant information is not enough to effectively support navigation; how this information is perceived, interpreted, and used by individuals must also be understood. This implies that it is necessary to integrate the principles of geographic relevance with those of cartographic visualization and cognitive science, an approach that will be explored in more detail in the following section on cognitive geovisualization.

2.1.2 Visual Representation of Spatial Information

Alongside the evolution of geographic information retrieval models, a second fundamental theoretical strand has emerged: the visual representation of spatial information in mobile maps. This field is rooted in the classical cartographic tradition, which has, for centuries, addressed the problem of how to effectively visualize geographic space to support orientation, planning, and spatial reasoning.

Cartography has undergone a profound transformation with the advent of digital maps. The map has evolved from being a static, generalized representation designed for experts to a dynamic, interactive interface that changes based on the location of its users (Meng & Reichenbacher, 2005). This transition has radically changed the entire cycle of accessing, exploring, and interpreting spatial information. Designing maps for mobile environments presents new challenges. Digital maps are now often used in highly variable conditions on small screens, and they involve rapid interactions with users engaged in cognitively demanding tasks while on the move. In this context, the following critical issues have emerged (Montello et al., 2018):

Visual oversimplification or excessive reduction of detail and graphic variability can hinder the formation of stable mental representations.

Information fragmentation linked to the use of multiple layers or successive navigation steps impedes overview and orientation.

Excessive automation, such as automatic rerouting or the selection of predefined routes, reduces user agency and can inhibit spatial learning processes.

Since the early 2000s, the cartographic community has sought to address these challenges and has begun to define principles of adaptive and context-aware design for mobile cartography. An early contribution in this direction was provided by (Meng & Reichenbacher, 2005), who proposed a conceptual framework for the design of adaptive mobile maps, based on integrating three main dimensions of context: the user, the current activity, and the physical and technological environment. This model anticipated many of the subsequent developments in the field of context-aware cartography, laying the foundations for personalized and situated interaction (Meng & Reichenbacher, 2005; Reichenbacher, 2005).

This vision is part of a broader perspective on cartography that considers the map not only as an information tool but also as a symbolic cognitive artifact designed to support the perception, memory, reasoning, and communication of geographic information (Montello et al., 2018). As such, maps directly influence the way individuals think, learn, and orient themselves in space, which is particularly critical during real-time navigation (Montello, 2005).

According to this perspective, cartography is a cognitive design process in which every graphic choice (e.g., colors, symbols, layout, landmark salience) must be calibrated to the user's task, the situational context, and the cognitive goals. The technical limitations of mobile maps (small screens, touch input) and the dynamic conditions under which interaction takes place increase the stringency of these constraints (Fabrikant et al., 2010; Montello et al., 2018).

From a cognitive point of view, an effective map must align with the user's perceptual mechanisms, attentive strategies, and short-term spatial memory (Montello, 2018). A key technique is the so-called "pop-out" effect, which demonstrates how visual design can selectively direct user attention to dense and cognitively demanding information displays (Fabrikant et al., 2010). Furthermore, visual advantages, such as congruence between graphical representations and user mental models, can play a critical role in spatial inference (Fabrikant et al., 2008). For example, when reasoning about temporal change, static maps employing small multiples (i.e., side-by-side frames of the same area over time) can outperform animated maps by reducing cognitive load and allowing the user to control the pace and sequence of comparison (Fabrikant et al., 2008).

Empirical studies provide support for this point. For example, showing a moderate number of salient landmarks (i.e., five) during virtual reality (VR) navigation improves route memory and landmark recognition without intensifying cognitive load (Cheng et al., 2022). However, increasing the count to seven created a greater neurocognitive load with no additional learning benefits (Cheng et al., 2022). Other studies using real-world outdoor navigation tasks found that realistic 3D visualizations of landmarks improved directional knowledge and increased visual attention to the environment compared with abstract landmarks, especially among users with lower spatial ability (Kapaj et al., 2024).

These studies demonstrate that visualization is never neutral; design choices shape what information users perceive, encode, and retain. Visualization on mobile devices is associated with interaction, depending on the results of user actions and commands performed in

the map application. In turn, these actions influence what the user sees and remembers. Therefore, interaction **emerges as the next essential element of mobile map design**, linking how information is presented with how individuals use it when navigating in the real world.

2.1.3 Interaction as a Core Design Element

Maps have been static throughout most of their history; once printed, their content and form were fixed. Interactivity was introduced in the first wave of digital cartography, although this was limited to predefined interactions. On a desktop GIS, users could click on menus and toolbars to toggle layers or query attributes. This approach treated maps less as dynamic tools and more as databases with a graphical shell (Schiller & Voisard, 2004).

By the late 1990s, geovisualization researchers had recognized that this was a missed opportunity, arguing that interaction should be elevated to a central design principle rather than remaining a peripheral complement to cartographic representation (MacEachren & Kraak, 2001). They proposed the conception of maps as dynamic interfaces that support exploration, reasoning, and learning, rather than as static representations of information (Fabrikant & Lobben, 2009). The advent of mobile devices made this theoretical shift inevitable. In practice, smartphones have embedded interactivity in everyday navigation. Unlike physical maps or desktop GIS, digital maps on smartphones allow users to manipulate them directly by pinching and swiping with their fingers on a touchscreen, tilting or rotating the device using motion sensors, or issuing spoken commands through voice input (Roth et al., 2024).

However, the first generation of mobile map applications often recycled desktop conventions, compressing the same menus and toolbars onto smaller screens (Meng & Reichenbacher, 2005) or failing to account for the other limitations of mobile devices, including unstable connectivity and the fact that users' attention would be divided between the device and their surroundings (Griffin et al., 2017; Roth et al., 2024). The gap between new mobile capabilities and traditional desktop design habits revealed the limitations of simply shrinking existing systems onto smaller screens. Touch screens, sensors, and multimodal interfaces created new ways of engaging with maps that were unprecedented in cartographic traditions (Roth et al., 2024).

Designing for these new modalities required more than technical adaptation; it demanded a reconceptualization of interaction as not merely a peripheral function but as an integral part of the ongoing exchange between the user, their environment, and the task at hand (Griffin et al., 2017; Roth et al., 2023, 2024). Over time, the recognition of the need for this reevaluation matured into a structured line of research.

The first step in this new agenda was to establish conceptual clarity. For effective design and evaluation, researchers required a consistent description of how individuals actually use maps. Early work responded to this requirement by decomposing map use into its fundamental operations, such as selecting and identifying features, querying and filtering data, navigating the map (panning and zooming), or measuring and overlaying spatial information (Nyerges et al., 1995). Building on this foundation, subsequent research moved from broad task categories to more formalized **taxonomies of interaction operators**. These operators describe the basic low-level actions performed by users engaging with a map interface, such as **panning** to change the visible extent, **zooming** to alter scale,

filtering to refine visible features, or **reprojecting** to shift between coordinate systems (Roth, 2013). Defined in this way, the operators function as reusable **design primitives**, enabling the comparison and analysis of different map interfaces and facilitating their construction from a common set of components.

More recently, this framework has been refined with a **mobile first orientation**, proposing that operators should feel intuitive on touchscreens and multimodal devices (e.g., zooming using pinch gestures or map rotation by tilting the device) rather than replicating the mouse-and-menu conventions of desktop GIS (Roth et al., 2024). Simultaneously, the concept of map use context emphasized the inextricable interconnection between interaction and the conditions of use. The user, the environment, the activity, and the map jointly shape how the interaction unfolds (Griffin et al., 2017).

Despite these theoretical advances, **empirical evidence on how people actually interact with mobile map applications in real-world conditions remains limited** (Roth et al., 2017). Interactive prototypes developed by academic research never progress beyond experimental settings or achieve sustained uptake, whereas commercial platforms such as Google Maps and Apple Maps dominate everyday practice. This imbalance indicates that current theories of interaction design are only partially grounded in ecological reality and that our knowledge of **usability, adoption, and contextual constraints** remains incomplete. A community-wide research agenda has thus outlined several persistent methodological gaps, including the need for **greater transparency and consistency in reporting study designs**, more **purposeful participant sampling** that reflects relevant user characteristics rather than defaulting to convenience samples, and the **integration of field and mixed-method studies** to capture interaction in naturalistic settings. This agenda also advocates for an evaluation that extends beyond narrow usability metrics, such as task completion time or error rates, to include broader assessments of **utility, user experience, and affective or cultural dimensions** of map use, which are often decisive for long-term adoption (Roth et al., 2017).

The limited empirical evidence available is suggestive of both the potential and the limitations of current approaches. For example, a study that unobtrusively recorded the use of mobile maps for everyday navigation revealed different sequences of actions, such as panning, zooming, and reopening the application, that correspond to distinct map use strategies. While these findings demonstrate that interaction traces can reveal underlying behaviors in real-world contexts, the results were limited to a single platform. Consequently, the generalizability of such strategies across the wider ecosystem of navigation tools remains unestablished (Savino et al., 2021).

While the scarcity of systematic studies is troubling for theory, its implications extend far beyond academic concerns. Failures of interaction design not only risk creating minor inconveniences but can also present genuine threats to safety. Well-documented cases of “death by GPS” illustrate how design shortcomings can have serious consequences (Lin et al., 2017). For example, ambiguous turn-by-turn instructions can confuse drivers at complex intersections. Insufficient treatment of road attributes, such as failing to distinguish paved roads from unpaved tracks or ignoring clearance restrictions, can lead users to unsuitable routes. Map-matching errors, whereby the system misaligns a user’s position or planned route with a segment of the network, can be particularly dangerous when they direct vehicles onto dirt tracks, seasonal roads, or paths that are officially closed. In such situations, subtle interaction failures can escalate into life-threatening situations (Lin et al., 2017). These cases establish that interaction design is not merely a superficial

matter of usability or interface polish; rather, it is a safety-critical concern requiring systematic empirical investigation and the establishment of robust design principles.

A broader conceptual lens is needed to address this challenge. Interaction should be viewed in two ways: first, as the set of operations through which users control a system, and second, as the behavioral traces that reveal insights into their emotional and cognitive states. This frame of reference highlights interaction as the richest yet least exploited layer of mobile map design. Treating mobile maps as solely functional risks making them cognitively impoverished, even if they are technically efficient. They might deliver a destination, but would be unable to support spatial understanding, learning, or recovery when difficulties arise. This consideration leads to the final important element of mobile map application design: **prediction**. Rather than being regarded as a series of isolated interactions, navigation must be considered an intentional, adaptive process that unfolds in space and time and can be anticipated. This is discussed in the following section.

2.1.4 Prediction in Mobile Map Design

While interactions tell us what users *do* with maps, prediction asks what a user will *need next* and why. Navigation is not a random sequence of clicks or gestures but rather an intentional, goal-directed activity that unfolds spatiotemporally under cognitive and environmental constraints (Montello, 2005).

This perspective broadens the challenge for mobile map design beyond simply displaying information or responding to explicit inputs. A predictive system should instead **anticipate when, why, and how** support will be needed as navigation progresses. A first step toward achieving this anticipation is a shared, formally defined vocabulary for navigation regarding what exists, who acts, and how the elements of navigation relate (Winter, 2001). Ontological theories provide this structure by providing explicit categories, such as entities (e.g., streets, junctions/decision points, and landmarks), actors (e.g., users and devices), and the relations among them (e.g., constraints and proximity). This formalization enables the mapping of raw traces, such as GPS points, onto meaningful states and the validation of predictions against real-world conditions instead of internal data structures alone. For example, an application could define what counts as a street, a decision point, a landmark, or an interaction state (Winter, 2001). For this vocabulary to be effective, it must resonate with the way individuals naturally reason about space. Research on naive geography has shown that people conceptualize their environment in terms of concrete entities, such as streets, landmarks, and routes, rather than as the abstract geometric primitives favored by formal GIS representations (Smith & Mark, 1998). Activity semantics are built upon this, formalizing the ways in which task descriptors and situational states shape behavior. It offers a vocabulary for context-dependent instructions, such as “turn left at the church” or “head home” (Kuhn, 2001). However, these categories are rarely clear-cut. Qualitative spatial reasoning research has demonstrated that concepts such as “near” and “far” are inherently vague and must be modeled with uncertainty, reflecting the bounded rationality of spatial reasoning (Frank, 1992). Formal treatments of vagueness, such as proximity relations, extend this idea by offering methods for capturing indeterminate states, such as hesitation or confidence, as seen in (Worboys, 2001). Encoding such “folk” concepts into structured vocabularies could allow for the interpretation of low-level traces, such as GPS points, as higher-level behaviors, such as exploring, reorienting, or navigating the way home.

Research has highlighted that once such vocabularies are in place, they can not only interpret traces but also enable the simulation of navigation decision-making processes. Agent-based way-finding models have built upon this foundation, translating conceptual categories into dynamic models of decision-making. Ontologies provide the terms for entities, tasks, and states, and simulations operationalize how navigation proceeds under constraints such as limited memory, selective attention, and time pressure within environments that impose particular advantages and restrictions (Raubal, 2001). This perspective was later extended to incorporate a user-centered time geography of mobile navigation, which framed navigation as an activity that progresses in space and time. The feasibility of this activity depends on spatial layout, temporal budgets, activity sequences, and contextual constraints that determine the range of possible user actions within a given spatio-temporal context (Raubal et al., 2004). This work explained the observable choices emerging from the structured interactions among the user, the task, and the environment, establishing that anticipation is possible once observations can be mapped to these spatio-temporal and contextual states, that is, representations capturing where the user is, when an action occurs, and under which environmental and task constraints it takes place.

However, identifying the kinds of observations that can be captured in practice, especially on mobile devices, and translating them into actionable signals remains a challenge. Some scholars have proposed that personalization should be based not only on **explicit inputs**, such as queries or stated preferences but also on **implicit interaction traces** (Ballatore & Bertolotto, 2015). The way in which individuals navigate digital maps, using actions such as zooming, panning, and toggling layers, creates a behavioral record that can be analyzed to infer their tasks and interests, suggesting that map content could potentially be automatically tailored to better match inferred needs through prototype examples (Weakliam et al., 2008). For instance, repeated zooming could reflect a search for finer detail, or frequent toggling of layers may signal overload. Although this work is still largely theoretical and exploratory, it marks a shift from providing conceptual accounts of anticipation to making operational proposals for prediction, that is, designs for systems that can read interactions as evidence of user needs and adjust the map accordingly (Weakliam et al., 2008; Wilson et al., 2010).

These efforts converged in a multilevel framework for digital map personalization based on user modeling and behavioral prediction (Ballatore & Bertolotto, 2015). This framework defined **map personalization** not as a static function of location or search history, but as a process of dynamically inferring user intent, analyzing interaction patterns (e.g., taps, clicks), and continuously updating behavioral models. Experimental systems, such as CoMPASS and RecoMap, demonstrated the feasibility of this approach, showing that preferences and objectives could be inferred from implicit behaviors such as proximity to POIs, repeated gestures, or temporal tap sequences (Doyle et al., 2010). These studies highlight the need for prediction to operate as a continuous loop involving observing behavior, inferring the task, and adapting the representation to provide a unique map for each individual at each moment (Ballatore & Bertolotto, 2015). This could entail highlighting a landmark just before a turn, simplifying visual detail when time is limited, or suggesting an alternative route when conditions change (Ballatore & Bertolotto, 2015).

However, this vision remains unrealized in the design of mobile mapping applications. Although studies have provided valuable conceptual guidelines, demonstrating that interactions can be interpreted as indications of intent and that maps can adapt accordingly, these ideas have been primarily confined to prototypes. Limited data availability, high computational requirements, and narrow behavioral signals that exclude cognitive dimensions have

restricted the scope of research. This persistent gap between theory and implementation has shaped the trajectory of more recent research agendas. Although adaptation has long been a guiding principle in mobile map design theory, only recently have researchers argued that its implementation is technically feasible. Advances in mobile sensing, improved data recording and storage capacity, and scalable modeling techniques, such as deep learning, now provide the basis for transitioning from static to dynamic models of mobile map use.

As discussed in the following section, current research programs are shifting their focus from abstract principles to practical design implications that enable real-time map adaptation. These efforts conceptualize adaptation as the real-time, dynamic integration of geographic and cognitive-behavioral dimensions tailored to the unique context of each user.

2.1.5 From Theory to Operationalization: Toward Context-Aware Map Design

A new generation of research agendas in mobile cartography has emerged in recent years. These programs explicitly address the challenges of renovating mobile map applications so that they can adapt to diverse user needs (Bartling et al., 2022; Griffin et al., 2017; Huang et al., 2018). The growing interest in this area has been driven in part by technological advances in user profiling, which have facilitated the application of adaptive principles—originally developed in domains such as advertising and recommender systems—to geographic interfaces. A key theoretical contribution was made by (Griffin et al., 2017), who investigated how and why context affects map use. Rather than prescribing a fixed solution, they proposed a research agenda that foregrounds variations in user–map interaction across dimensions such as intent, cognitive load, task type, and physical setting. This paved the way for theories that aim to transform the rich principles of context-aware cartography into adaptive systems that leverage a deeper integration of conceptual models, empirical findings, and computational approaches in order to sense, interpret, and respond to context as it changes. One of the central contributions of this work is the distinction it makes between design context, the assumptions that guide map creation (e.g., audience, intended use, visual abstraction), and use context, that is, the actual situation in which a map is consulted (e.g., walking in an unfamiliar city, under time pressure). They argued that ignoring this distinction reduces usability and compromises decision-making effectiveness. Building on this, (Bartling et al., 2021) emphasized that the most important determinant of user experience is activity type, rather than static user characteristics. They advocated for context-based map adaptation that is grounded in situational needs, challenging the one-size-fits-all logic of standard personalization approaches. In parallel, (Kiefer et al., 2017) introduced the notion of adaptive map interfaces that adjust their design or content based on the state or context of the user. Crucially, their findings show that adaptation must remain under user control: individuals prefer predictable interactions and are likely to resist unsolicited changes. Pushing this trajectory further, (Fabrikant, 2023) proposed a cutting-edge model of neuro-adaptive geographic displays. These systems monitor the user’s cognitive state using physiological signals (e.g., eye movements, brain activity, or heart rate) to dynamically adjust information complexity, salience, and detail in real time. This aims to optimize attention, memory, and decision-making, particularly for users with cognitive vulnerabilities or those under high situational stress.

Collectively, these contributions mark a clear shift, indicating that to build adaptive mobile maps, researchers must go beyond static spatial representations to embrace a richer,

user-centered notion of context—one that incorporates cognitive, emotional, behavioral, and environmental dimensions.

This conceptual richness has, however, resulted in a form of operational fragmentation. Each line of research tends to address a specific dimension (e.g., design context, interface control, or cognitive load), but few efforts have integrated these layers into a unified framework. To bridge this gap, (Bartling et al., 2023) proposed a three-level cartographic context taxonomy, which serves as a foundational reference for the holistic modeling of user–map interaction. The taxonomy distinguishes between three interdependent levels of context, each capturing a distinct component of mobile map use:

Intrinsic context includes the user’s cognitive, emotional, and motivational factors (e.g., spatial anxiety, sense of orientation, perceived cognitive load, intentionality);

Behavioral context refers to interaction patterns that are observable in real time (e.g., gesture frequency and type, pauses, hesitations, recurrent use of specific functions);

Extrinsic context concerns the environmental and technological conditions in which the map is used (e.g., background noise, screen brightness, GPS signal quality, device characteristics).

To design mobile map applications that are context-aware, extrinsic, intrinsic, and behavioral signals must be integrated into a unified framework that guides the user step by step and adjusts support in real time (Bartling et al., 2023). Only this kind of integration can produce responsive maps that adapt to the continually fluctuating conditions of way-finding; static snapshots of the user or environment are inadequate for this task. However, this vision is far from being fully implemented.

Three obstacles in particular stand out, the first of which is **data collection**. There are significant challenges to obtaining ecologically valid traces across all three levels. Such data are inherently sensitive, and their handling requires careful protection of privacy; however, they must also be rich enough to yield meaningful insights for research. The second obstacle is **integration**. Extrinsic, intrinsic, and behavioral contexts are typically studied in isolation, producing fragmented insights rather than a unified account of map use. The third hurdle is **prediction**. While predictive approaches exist, they mainly focus on extrinsic signals such as location, trajectory, or ETA. Although some prototypes have incorporated implicit behavioral traces, none combine extrinsic, intrinsic, and behavioral dimensions into a comprehensive framework to anticipate user needs. The following section outlines how research has approached these contextual dimensions, showing where progress has been made, where results diverge, and what is still needed for the provision of a predictive, integrated account of mobile map adaptation.

2.2 Challenges in Context-Aware Map Design

The theory of user-centered mobile map design is well-established, as acknowledged in the first part of this literature review (Section 2.1). However, the application of this theory to everyday navigation remains limited. This discrepancy can be attributed to the three obstacles mentioned above, which the following subsections address in turn; that is, the challenges involved in (a) collecting ecologically valid data on real-world mobile map use (Section 2.2.1), (b) integrating the three contextual dimensions (extrinsic, intrinsic, and behavioral) to enhance understanding of real-world mobile map use (Section 2.2.2), and (c)

developing predictive systems that leverage this integrated framework to anticipate how mobile maps will be used, where modern prediction models are limited by their reliance on a narrow definition of extrinsic context (Section 2.2.3).

2.2.1 Data Collection for Ecological Validity in Mobile Map Use

To design mobile map applications that respect users' cognitive processes, emotional states, and behavioral patterns as they navigate their environments, it is crucial to first understand how, when, and where individuals use these tools for everyday navigation (Bartling et al., 2023). Such understanding should form the basis of mobile map design. However, existing applications have evolved in the opposite direction: instead of supporting decisions during navigation, they increasingly script them, offering precalculated suggestions, thereby prioritizing externally defined goals such as efficiency or speed. This shift reduces active spatial reasoning and pushes users into a passive role (Parush et al., 2007). Consequently, these applications do not help users refine their internal representations of space; instead, they often provide pre-packaged routes that discourage exploration and planning. User reliance on these functions carries well-documented risks, including impaired landmark learning, reduced allocentric memory, and increased spatial anxiety, leading to long-term overreliance on the mobile map (Dahmani & Bohbot, 2020; He & Hegarty, 2020; Hejtmánek et al., 2018; Montello, 2018; Parush et al., 2007; Ruginski et al., 2022).

A first step toward addressing these risks is to observe navigation as it occurs in daily life. However, the literature is weakest in this area; most research on mobile map use has **limited ecological validity**. The main challenge is obvious; unobtrusively capturing people's moment-to-moment behavior as they actually navigate is technically and practically difficult. **Self-report questionnaires**, therefore, remain the most common tool for studying mobile map use. Such measures can estimate the frequency with which individuals use mobile maps (e.g., weekly or monthly) and for what purposes they are employed (Ishikawa, 2019; Ruginski et al., 2019; Sugimoto et al., 2022). Findings from studies using these measures consistently show that the widespread adoption of GPS-assisted navigation is associated with key spatial cognition variables. Individuals who rely more heavily on such systems tend to report a poorer sense of direction, higher spatial anxiety, and reduced spatial memory performance (He & Hegarty, 2020; Hegarty et al., 2010; Ruginski et al., 2022). However, the ecological validity of such measures is very low. They are retrospective and therefore prone to capturing memory errors, they offer only coarse temporal resolution, and they are vulnerable to social-desirability bias. In practice, this makes them unsuitable for capturing the fine-grained moment-to-moment dynamics of real-world map engagement (Topete et al., 2024). To study navigation under more naturalistic conditions, researchers have adopted two primary methods: simulating outdoor environments in indoor laboratories and measuring behavior in the field.

For instance, studies conducted indoors to understand the relationship between spatial skills and mobile maps have largely focused on **virtual reality (VR)**, simulating navigation tasks under controlled conditions. VR configurations allow researchers to rigorously manipulate variables as well as avoid the complexities of real-world environments while investigating individual differences in spatial ability and spatial knowledge acquisition during navigation (Cheng et al., 2022; Dahmani & Bohbot, 2020; Gardony et al., 2015; Hejtmánek et al., 2018). Interestingly, studies have found that performance on virtual tasks correlates with real-world navigational ability in complex cities, bolstering VR's ecological credentials (Coutrot et al.,

2019). Methodological extensions add embodied elements, such as immersive VR with head-mounted displays (head movement), combined with locomotion interfaces such as foot-pedal systems (walking). When integrated with mobile neuroimaging, these extensions enable the linking of behavioral measures with neural activity to explore the cognitive and neural mechanisms of navigation more directly (Cheng et al., 2022; Jeung et al., 2023). Although these studies have validated the use of VR for spatial navigation research, it still has significant limitations. VR cannot fully replicate the complexity of real-world navigation; the sensory richness and physical movement involved are still difficult to replicate in a virtual environment. Furthermore, VR’s reliance on simplified stimulus–response strategies and controlled, pre-designed tasks limits the acquisition of spatial knowledge compared to the more dynamic and fragmented use of devices and technologies in real-world experiences (Chrastil & Warren, 2012; Oulasvirta et al., 2012). Research using outdoor, real-world navigation studies, however, offers distinct advantages over VR in its ability to provide insights with higher ecological validity. These studies are better at capturing natural human behavior than those performed in controlled laboratory environments. Hence, recent technological advances such as **mobile electroencephalogram (EEG)** and **eye-tracking (ET)** and combining them with mobile GPS navigation applications have bridged the gap between laboratory studies and real-world environments, facilitating more detailed cognitive and visual data collection during real-world navigation (Hilton et al., 2023; Kapaj et al., 2024). However, these techniques have significant limitations. The inherently dynamic nature of real-world environments introduces variability that complicates data quality and consistency. Mobile EEG, for example, is highly sensitive to motion artifacts, making it challenging to capture clean neural data during active navigation. Similarly, ET data can be affected by fluctuating lighting conditions and environmental distractions, further reducing data reliability. Additionally, the synchronization of data streams from multiple devices, such as EEG, ET and GPS, adds another layer of complexity, requiring extensive post-processing and often limiting analyses. There are also additional concerns in terms of the participant’s perspective. The Hawthorne effect is a significant issue in studies that use wearable devices (Adair, 1984), whereby individuals alter their behavior due to their awareness of being observed. Participants wearing EEG caps or ET goggles may become overly conscious of their actions, potentially focusing more on landmarks or attempting to navigate more effectively than they would otherwise in their daily life. In addition, the discomfort or fatigue caused by these devices may cause participants to rush or modify their behavior in ways that do not reflect their natural navigation patterns. These limitations highlight the need for methods that can investigate real-world navigation without burdening participants with heavy equipment or artificial tasks. Smartphones offer a promising avenue for such techniques because they are already integrated into everyday life, thereby facilitating data collection with minimal disruption (G. Miller, 2012). Smartphone-based approaches are therefore *less intrusive* than techniques such as mobile EEG or ET, which can alter behavior through physical discomfort or heightened awareness of being observed. Smartphone-based data collection has created new research opportunities through **mobile sensing**, which refers to the analysis of data generated by mobile devices to identify cognitive, emotional, and behavioral patterns on a large scale (Montag & Baumeister, 2023). Smartphones’ integrated sensors and ability to record background data enable the continuous, naturalistic, and minimally invasive collection of various behavioral indicators, such as personality traits and mental health signals, by monitoring the user’s app use, location, text input, and media consumption (Huckvale et al., 2019; Martinez-Martin et al., 2018). This ensures that participants’ daily routines suffer minimal disruption (G. Miller, 2012; Montag & Baumeister, 2023). The logic that

underlies mobile sensing is closely aligned with the principle of **ecological momentary assessment (EMA)**, which emphasizes improving the ecological validity of data by using repeated, in-situ measurements (Shiffman et al., 2008). EMA involves collecting real-time data in participants’ natural environments, typically by employing short self-reports or diary entries completed several times a day. These repeated assessments cover factors such as mood, activity, location, or social context, thereby reducing recall bias and providing a more authentic, fine-grained view of behavior. The widespread use of smartphones has further facilitated EMA by making in-situ data collection seamless, automated, and integrated into daily routines without requiring participants to use additional devices or make any special effort (G. Miller, 2012). However, as we have already acknowledged, mobile sensing raises important concerns. First, it makes the retrieval of interaction data, including mobile map application data, from applications particularly difficult, due to operating system restrictions and the fragmented nature of the data across devices and platforms (Oulasvirta et al., 2012; Savino et al., 2021). Second, it raises ethical issues related to privacy and data security. Passive, continuous data collection via smartphones therefore requires carefully balancing the collection of fine-grained and detailed ecologically valid data with the protection of user rights and expectations. Thus, data collection should be strictly limited to gathering only the type and amount of information that is necessary to produce scientifically meaningful results, and the accumulation of unnecessary or sensitive data should be avoided (Martinez-Martin et al., 2018; Montag & Baumeister, 2023).

A promising methodological strategy for addressing this challenge both prioritizes privacy and minimizes data collection. This strategy is guided by the **principle of parsimony**, favoring the simplest and most adequate explanation over more complex solutions (Burnham & Anderson, 2002; Gori et al., 2024). In this case, parsimony means collecting only the smallest set of behavioral signals necessary for predictive accuracy while avoiding the acquisition of superfluous or sensitive data (Gori et al., 2024). This approach fulfills the ethical requirements of minimizing information demands without sacrificing scientific rigor (Martinez-Martin et al., 2018).

Tappigraphy is one approach that embodies this principle. A technique originally developed in cognitive neuroscience to study behavioral rhythms and sensorimotor functions by passively recording touch events on smartphones, tappigraphy does not record content, unlike many other mobile sensing protocols (Balerna & Ghosh, 2018; Borger et al., 2019; Ceolini & Ghosh, 2023; Ceolini et al., 2022). Instead, it focuses exclusively on the timing and sequence of touch events and the activity of foreground applications (Balerna & Ghosh, 2018). This ensures discreet and non-invasive data collection while maintaining high levels of privacy and ecological validity. Recent work has begun to explore using tappigraphy to study how mobile maps are used in real life (Reichenbacher et al., 2022), reflecting a growing interest in developing unobtrusive methods for capturing navigation behavior. Applied to navigation research, tappigraphy could facilitate the exploration of how individuals with different spatial abilities and/or experiencing diverse emotional states interact with mobile maps in their daily lives. By operating unobtrusively in natural settings, this technique preserves ecological validity while capturing the fine-grained touch dynamics of smartphone use. These interaction traces **could reveal implicit patterns of behavior** that may serve as proxies linking **intrinsic factors** (e.g., spatial abilities or emotional states) to various **extrinsic factors** (e.g., environmental structure) as users engage with mobile maps within their daily routines.

However, tappigraphy has not yet been applied to the study of mobile navigation and most research in this broader field has so far only examined these three contextual factors

separately, resulting in a fragmented understanding of daily navigation. The following section reviews empirical research on each factor (intrinsic, behavioral, and extrinsic), highlighting both the insights each strand provides and the gaps that remain when they are considered in isolation.

2.2.2 Fragmented Perspectives: Intrinsic, Behavioral, Extrinsic

As outlined above, research on the use of mobile maps falls into three categories that partially overlap but are largely separate from one another. The first part of this section examines the literature related to the intrinsic context, focusing on the user's spatial cognition, abilities, and emotional traits. This body of work demonstrates the impact of GPS-assisted navigation on learning, memory, and strategy choice, as well as how these effects vary among individuals. However, this literature tends to treat navigation as a cognitive outcome rather than as an ongoing, device-mediated activity. It rarely captures the detailed interaction patterns that take place on the screen and often overlooks the specific properties of the environments in which navigation occurs.

The second part of this section, the behavioral context, is the least developed in the literature. Such studies would be expected to provide careful documentation of what individuals actually do on their phones while navigating, including touch patterns, pauses, app switching, and session timing. In practice, such research is scarce. The studies that do exist rarely explore individual traits or the environments in which interactions occur. This gap is significant because behavior is the immediate expression of both the person and the place. Without systematic behavioral data, the crucial link between intrinsic traits and extrinsic settings is missing.

The third part of the section shifts focus to research examining the extrinsic context, that is, the structure of street networks, the distribution of landmarks, and the visual complexity of urban form. These studies establish that some environments are more difficult to navigate than others and reveal that the built environment shapes spatial cognition in lasting ways. However, they rarely trace those environmental influences to the specific behaviors of individuals using map applications for everyday navigation.

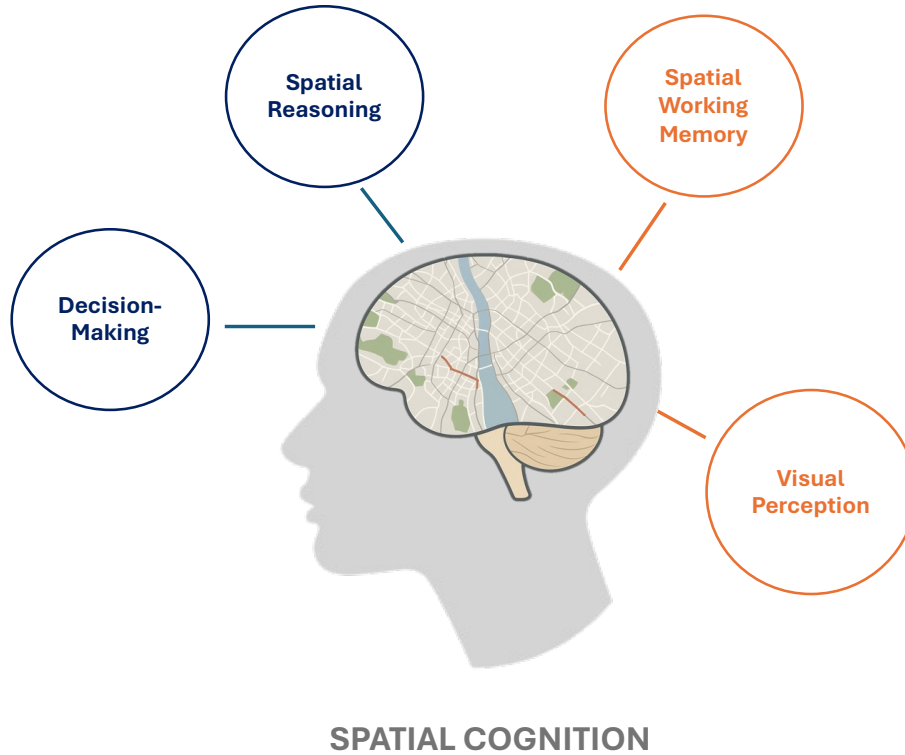
Intrinsic Context: Effects of GPS-Assisted Navigation on Spatial Cognition

Cognitive science views navigation as a complex mental activity that requires the integration of several key functions: visual perception, spatial working memory, spatial reasoning, and decision-making processes (Ekstrom et al., 2018).

While visual perception enables the extraction of information from the environment (landmarks, obstacles, signs), spatial working memory is a short-retention type of memory specifically used to temporarily maintain and manipulate spatial representations that are useful for specific tasks, such as navigation (Ekstrom et al., 2018). Spatial reasoning is the means by which we understand the relationships between places and objects, plan alternative routes, and integrate new information. Finally, decision-making processes support choosing the best route, adapting to the context we find ourselves in, and the management of unexpected events (Ekstrom et al., 2018).

Combined, these functions support what has been classically defined as the construction and updating of cognitive maps (i.e., mental representations of space that support flexible

Figure 2.1.: The representation of a city layout within the brain symbolizes how environmental knowledge is internalized to support navigation. The labeled nodes highlight four core cognitive functions involved in this process: visual perception (extracting environmental cues), spatial working memory (maintaining and updating spatial representations), spatial reasoning (inferring spatial relationships and planning routes), and decision-making (selecting navigation strategies and adapting to environmental changes).



and efficient behavior Tolman, 1948). These abilities have played a fundamental role in the evolution of the human species, enabling survival, migration, the exploration of territory, and cultural development (Tolman, 1948). This process is visually summarized in Figure 2.1. Researchers often group these functions into two core components of navigation: *wayfinding* and *locomotion* (Montello, 2005; Wiener et al., 2009). Successful navigation behavior emerges from the ongoing coordination between wayfinding and locomotion. Wayfinding specifies the direction and overall strategy, whereas locomotion implements this strategy and generates feedback that can be used to adjust and refine plans (Montello, 2005).

A set of **spatial reference systems** underpins both processes. These are mental frameworks that facilitate the understanding, organization, and interpretation of the spatial relationships among objects, places, and ourselves. They are typically classified into three categories: egocentric, allocentric, and (more recently introduced) beacon-based systems (Burgess, 2006; Ekstrom et al., 2018).

The **egocentric system** represents space in relation to the individual's body (e.g., "the Limmat river is on my right") and is essential for locomotion, as it guides immediate action based on what is perceived in real time (Burgess, 2006). The **beacon system**, or target strategy, is based on the presence of a visible object (e.g., a sign, a monument) toward

which we can move directly, without elaborating a complete spatial representation. It is a useful strategy in familiar or structured contexts and can support locomotion and orientation in a rapid but limited way (Ekstrom et al., 2018).

The **allocentric system** encodes spatial relationships between objects or places regardless of the subject’s position (e.g., “the Grossmünster is east of the Lindenhof hill”). Fundamental for wayfinding, it enables the construction of coherent mental maps, models of the environment that can be used to plan routes, deduce directions, and navigate flexibly, especially in large or unfamiliar spaces (Burgess, 2006; Tolman, 1948).

Navigating effectively entails integrating these three reference systems according to the context and objectives, flexibly switching from one to another (Ekstrom et al., 2018).

Empirical research has demonstrated that GPS-assisted navigation disproportionately disrupts the allocentric system by automating key processes such as route planning and localization (Dahmani & Bohbot, 2020). Münzer et al. (2006) offer experimental support for this claim, demonstrating that computer-assisted navigation systems lead users to develop better route knowledge but weaker survey knowledge than paper map users, thus revealing a trade-off in which automation hinders the formation of a more global spatial understanding. In a similar vein, Münzer et al. (2012) replicate this trade-off in a controlled experiment, where navigational assistance enhanced wayfinding efficiency yet diminished allocentric learning, indicating that users found it harder to perform tasks that required a holistic mental representation of the environment. Simultaneously, the use of GPS increasingly reinforces beacon-based and egocentric strategies, supporting the immediate navigation task, but offering limited support for long-term spatial learning. A growing body of research has highlighted how the specific features of mobile map applications could have unintended consequences by offloading cognitive processes and weakening allocentric strategies. Importantly, these consequences are not directly derived from the predictive functions of these tools (e.g., destination inference, ETA prediction, route optimization, and POI classification), which involve algorithms operating on the back-end of the application (see Section 2.2.3 for a detailed overview). The effects on users are mediated by specific interface features and interaction flows. These design elements translate the predictive logic for the user, providing navigation support in a simplified, step-by-step format. In the following, we list the interface elements and their associations with adverse effects.

 **Sequential Instructions:** Both auditory and visual turn-by-turn instructions (e.g., “in 250 meters, turn right”) eliminate the need to understand the overall structure of the route or anticipate upcoming decision points. As a result, users often develop fragmented or shallow mental maps (i.e., they might remember isolated turns of the navigated path, but not how they connect within a broader spatial layout (Münzer et al., 2006; Parush et al., 2007)). Segmented instructions also reduce users’ awareness of the surrounding environment and impair their ability to recall spatial features or adapt to unexpected situations (e.g., incorrect directions, detours), since they have not developed an independent internal representation of the area (Gardony et al., 2013; Ishikawa, 2019). At the front end, this function is the visible manifestation of continuous route-optimization algorithms running in the back end. While the system constantly recalculates the most efficient path based on the user’s real-time position, the front end presents these computations as a sequence of discrete, turn-by-turn instructions.

📍 Real-Time Geolocation and Tracking: The real-time location indicator representing a user's current location anchors their attention to the immediate surroundings and reduces their engagement with the broader spatial context. This encourages egocentric strategies over allocentric processing. Empirical evidence confirms that users who rely on GPS interfaces displaying their real-time location spend significantly more time fixing on the map interface rather than on environmental cues. This leads to a less robust acquisition of spatial knowledge (Hejtmánek et al., 2018). This form of egocentric anchoring aligns with findings of reduced survey knowledge acquisition, compared to map-based planning (Münzer et al., 2012). At the front end, this function provides a visible representation of the continuous ETA updates and route optimization processes operating in the background. The interface displays the user's position in real time while underlying algorithms estimate travel time, monitor deviations, and adjust the route to maintain optimal efficiency.

🔄 Automatic Re-Routing: This feature simplifies daily navigation and has contributed to the popularity of GPS-assisted systems. However, automatic corrections of deviations from the planned route mean that users are not required to reflect on the error, identify its cause, or find alternative paths. This bypasses key spatial learning mechanisms, such as error recovery and strategic inference. Research shows that effective spatial learning involves managing navigational uncertainty, that is the natural ambiguity that arises when one must estimate position, direction, or the next action under incomplete information at the perceptual, representational, and motor levels (Kessler et al., 2024). At the front end, this function represents the visible outcome of continuous ETA update and route optimization processes operating in the back end. When a deviation occurs, the underlying algorithms instantly recompute the optimal path and project the new solution to the interface as an updated route.

🏠 Adaptive Filtering and Predictive Algorithms: Personalized algorithms typically suggest routes or POIs based on the users' past behavior, current context, or inferred preferences (e.g., frequently used transportation modes), thereby optimizing the perceived relevance of geographic information (De Sabata, 2013; Reichenbacher, 2009). While this can reduce cognitive load and facilitate navigation, it also entails significant trade-offs. By favoring familiar routes or the algorithms that prioritize them, these systems can limit the diversity of a user's spatial experiences and restrict their exposure to alternative routes. Users are directed to previously visited, pre-selected routes, excluding viable alternatives. This invisible filtering creates spatial biases and generates a "geofilter bubble", influencing how space is experienced before it can be explored (Low, 2022). At the front end, this function appears as personalized recommendations, such as ranked points of interest (POIs), suggested destinations, or preferred routes. These recommendations visually and interactively express the results of back end filtering and inference algorithms. These algorithms continuously evaluate contextual and behavioral data to predict the user's preferences, translating those computations into map highlights, route suggestions, or search rankings.

The negative effects of GPS use on spatial cognition are not uniform across individuals.

Differences in spatial ability, emotional traits, and gender play a key role in shaping the impact of GPS-assisted navigation on spatial learning.

A significant predictor of an individual's subjective spatial ability is their sense of direction. The instrument most widely used for assessing this is the **Santa Barbara Sense of Direction Scale (SBSOD)**, a self-report questionnaire that asks respondents to rate their confidence and perceived ability in performing tasks such as navigation, route memorization, and mental map construction in unfamiliar environments (Hegarty et al., 2002). While it is a subjective measure, SBSOD scores consistently correlate with objective spatial performance, with high scores being associated with better task outcomes (Hegarty et al., 2002).

Spatial updating is the ability to continuously track one's position in real-time while moving through space;

Route Learning refers to the ability to memorize and accurately retrace a route after limited exposure to it;

Map-based tasks include reconstructing layouts from memory and translating between egocentric (first-person) and allocentric (map-like) perspectives.

Importantly, the study by Gardony et al. (2015), in which participants completed a series of spatial learning tasks in a virtual desktop environment with varying levels of environmental complexity (e.g., number of intersections, the density and distinctiveness of landmarks, and the irregularity of route geometry), provides clear empirical evidence of the link between sense of direction and spatial learning during GPS-assisted navigation (Gardony et al., 2015). The results indicated that individuals with low SBSOD scores exhibited significantly lower performance in key areas. They were less accurate in retracing routes, experienced greater difficulty reconstructing spatial layouts, and were more likely to rely on egocentric strategies. Furthermore, these individuals tended to use navigation aids more passively, treating GPS-style directions as rigid scripts instead of integrating surrounding environmental cues. Conversely, participants with high SBSOD scores demonstrated greater cognitive flexibility, a preference for allocentric strategies, and the formation of more coherent and accurate mental representations of the space (Gardony et al., 2015).

These findings emphasize that people with lower spatial confidence tend to delegate spatial processing to the device, limiting their engagement with the environment and strengthening their cognitive dependence on technological support.

In addition to spatial abilities like sense of direction, emotional traits such as **spatial anxiety (SA)** have also been identified as an important predictor of how individuals interact with GPS systems and how effectively they learn spatial information. SA refers to the feelings of fear, uncertainty, or stress that arise when a person has to perform orientation tasks, especially in unfamiliar environments Lawton, 1994. This form of anxiety is distinct from general anxiety or broader neurotic traits (e.g., emotional instability, depression, etc.) and is typically assessed using validated self-report instruments. One of the first and most cited questionnaires is the Lawton Spatial Anxiety Scale, which assesses anxiety associated with the daily task of finding the right place (i.e., navigating an unfamiliar city or locating a car in a crowded parking structure) Lawton, 1994; Lawton and Kallai, 2002. In this study, participants with higher spatial anxiety were more likely to rely on egocentric landmark-based navigation strategies, while those with lower anxiety tended to use map-based allocentric strategies (Lawton, 1994). The 2002 follow-up study also

showed that spatial anxiety influences how people give directions: those with higher anxiety preferred familiar concrete references (e.g., landmarks), whereas those with lower anxiety were more likely to use abstract or metric descriptions (e.g., cardinal directions, distances) (Lawton & Kallai, 2002). Hund and Minarik (2006) built on this work by exploring the implications of spatial anxiety in the real world with a study in which participants were asked to navigate an unfamiliar campus using a map and were then instructed to reproduce the route from memory. Spatial anxiety was measured using the Lawton scale (Hund & Minarik, 2006). Participants with higher spatial anxiety took significantly longer to complete the route, were less accurate in recalling it, and paused more frequently, despite all participants having access to the same environmental information (Hund & Minarik, 2006). These findings suggest that spatial anxiety impacts spatial memory as well as navigation performance, likely due to increased cognitive load and attentional demands (Hund & Minarik, 2006).

Lyons et al. (2018) later developed the Multidimensional Spatial Anxiety Scale (MSAS) based on this previous work, dividing spatial anxiety into three subdomains: navigation anxiety, mental rotation anxiety, and visualization anxiety (Lyons et al., 2018). In their study, 217 participants completed the MSAS along with a series of corresponding spatial tasks. The results revealed domain-specific associations: participants with high navigation anxiety made more directional errors and took longer to complete virtual wayfinding tasks; those with high mental rotation anxiety were less accurate on 3D rotation tasks; and visualization anxiety was linked to poorer performance on mental imagery tasks (Lyons et al., 2018). These findings empirically validate the hypothesis that spatial anxiety is multifaceted and predictive of varying performance deficits across different spatial abilities.

Finally, as part of a larger spatial thinking assessment, He and Hegarty (2020) introduced a spatial anxiety subscale specific to navigation. Their experiment involved participants performing route planning and spatial updating tasks in virtual environments. Individuals with higher spatial anxiety exhibited poorer performance, particularly on tasks requiring them to learn spatial configurations (He & Hegarty, 2020). They also reported a greater reliance on GPS tools, especially in unfamiliar settings. The study revealed the key mediating role of spatial anxiety in the relationship between spatial ability and GPS use. Individuals with less confidence in their spatial skills were more likely to rely on navigation systems, limiting their engagement with the spatial environment and reducing their opportunities to engage in active learning (He & Hegarty, 2020).

Gender differences in navigation-related traits are well established, particularly in terms of self-reported spatial ability and spatial anxiety. Studies have shown that women generally report a weaker sense of direction and higher levels of spatial anxiety than men (Hegarty et al., 2006; Lawton, 1994). In her 1994 study, Lawton found that spatial anxiety was negatively correlated with spatial ability and positively associated with the tendency to avoid navigation tasks. This suggests that emotional factors play a key role in shaping both navigation strategies and confidence. However, these self-reported gender differences do not always align with actual performance. For example, some research has revealed no substantial gender differences in performance on navigation subtasks such as spatial updating or landmark learning. Instead, observable disparities tend to arise in behavioral tendencies, particularly in the strategies individuals adopt and the extent to which they rely on external aids. Women are more likely to favor egocentric or landmark-based strategies and to depend on turn-by-turn directions or verbal instructions, whereas men more frequently use allocentric or survey-based strategies that involve integrating spatial

relationships across a broader environment (Hegarty et al., 2006; Lawton, 1994; Lawton & Kallai, 2002). This suggests that perceived difficulty, rather than actual ability, could be the primary factor influencing navigation behavior (Hegarty et al., 2006).

The perception of difficulty could also affect how people use navigation tools. For example, individuals with high spatial anxiety—more often reported by women—tend to be more cautious and exhibit increased reliance on external guidance, including GPS-based assistance. However, the relationship between gender and GPS reliance is not straightforward. Despite tending to report higher spatial anxiety and lower spatial self-efficacy than men, women do not necessarily demonstrate a greater dependence on navigation tools. Research suggests that emotional and cognitive traits, rather than biological sex, are the main predictors of tool reliance (He & Hegarty, 2020; Ishikawa, 2019). This interpretation is in line with Münzer et al. (2016), who, in a large-scale validation study, reported no direct gender effects on navigation aid use; instead, they found that reliance on aids was mediated by wayfinding strategy preferences and sense of direction, supporting the view that psychological factors outweigh demographic ones.

The overall literature on intrinsic factors shows that navigation is based on an integrated set of cognitive processes, including visual perception, spatial working memory, spatial reasoning, and decision-making, which work together to support effective wayfinding (Montello, 2005). Furthermore, this body of work suggests that overreliance on GPS-assisted navigation can alter these processes over time (Dahmani & Bohbot, 2020). The valuable support provided by common interface features such as automated route planning and real-time geolocation comes at a cognitive cost. Most of these features are offloaded to allocentric processing in favor of egocentric or beacon-based strategies, leading to predictable consequences for spatial learning and memory (Hejtmánek et al., 2018; Kessler et al., 2024; Parush et al., 2007). Löwen et al. (2019) build on this by examining orientation maps that accentuate local or structural elements. They find that highlighting landmarks facilitates route knowledge, whereas stressing global features enhances survey knowledge, and they propose design strategies to counteract the cognitive offloading associated with GPS use.

Notably, these negative effects are moderated by individual differences. People with lower spatial ability or greater spatial anxiety are more likely to adopt passive, device-dependent behaviors and consequently suffer larger declines in allocentric updating and long-term spatial knowledge (He & Hegarty, 2020).

However, studies on intrinsic context have a significant limitation: they primarily focus on the cognitive dimension and rarely capture the device-mediated interactions that could reflect these cognitive processes. This research rarely records fine-grained interaction traces, such as those based on touchscreens, alongside individual neuropsychological measures. However, these behavioral data are necessary for determining how individual susceptibility appears during everyday use of map applications or identifying specific interaction patterns that reflect negative over-reliance behavior. The following subsection outlines the limited existing research on the behavioral context and explains why, despite its importance, this aspect is seldom studied.

Behavioral Context: How Mobile Map Interfaces Shape Interaction

Research on intrinsic contexts has highlighted the cognitive and behavioral vulnerabilities associated with using mobile maps. It is also important to examine the behavioral context,

as it could reveal the ways in which these vulnerabilities manifest on the device itself, moment by moment. Every interaction with an application leaves behind a behavioral trace. These interactions range from the number of touches and types of gestures performed to the use of functions such as search, zoom, or re-centering, differences in session duration, and switching from one application to another. In mobile map applications, these traces signal the connection between the user and the query result and, ultimately, the navigation decision. If interpreted carefully, such data can reveal different navigation strategies, such as following directions step by step versus scanning the map for information. It can also suggest states of confusion, disorientation, or stress as they develop in real time.

Despite this potential, empirical research on the behavioral context of map interaction remains limited. This stands in stark contrast to the study of general mobile interactions, which has flourished within the field of HCI and produced a rich body of work. Studies have monitored the *temporal rhythms* of app usage, revealing how individuals' engagement with their phones follows daily and weekly patterns influenced by various routines such as commuting, work, and leisure (Bhavnani et al., 2017; Böhmer et al., 2011; Falaki et al., 2010). Other research has identified *groups of users* with distinctive interaction styles (i.e., some check many apps frequently but briefly, while others use fewer apps for longer sessions), highlighting the diversity of mobile utilization across populations (Rahmati & Zhong, 2013; Zhao et al., 2016). A further line of research built *predictive models* of app launches, showing that upcoming app use could be anticipated from prior contexts such as the time of day, location, or recent activity (Leroux et al., 2013). Recent work in mobile sensing has collected continuous app interaction traces, demonstrating associations between everyday usage and broader characteristics, including personality traits and well-being indicators such as stress level and mood (G. Miller, 2012; SettanniMichele et al., 2018; Stachl et al., 2020).

Mobile maps are distinctive apps in that they are tools that introduce unique challenges, such as real-time geolocation, safety risks associated with divided attention, and the tension between following a prescribed route and exploring alternatives. While map applications have been included in large-scale studies of mobile use as one category among many (Böhmer et al., 2011), dedicated research on map interaction itself is scarce.

One notable exception is the work of Savino et al. (2021), who developed a wrapper that incorporates a web view of a map. This wrapper recorded in-app states, such as map center, zoom level, taps, and pan or zoom gestures, without relying on commercial telemetry (Savino et al., 2021). By categorizing these interactions into states such as Search, Place, and Directions, the researchers identified distinct usage modes, recurring action sequences, and measurable differences between exploratory and direction-seeking sessions (Savino et al., 2021). Their work demonstrated the feasibility of capturing detailed map-specific behaviors under real-world conditions and that such data produce recurring, interpretable patterns (Savino et al., 2021). At the same time, this approach faces significant limitations in terms of reproducibility and generalization. Wrapper interfaces are not equivalent to native applications, and web-based solutions are largely obsolete as users are automatically redirected to native apps. This form of data collection is also intrusive, raising privacy concerns and constraining the scope of what can be recorded. Furthermore, the study relied on relatively small samples and focused exclusively on Google Maps, although navigation is currently often distributed across a plurality of applications. The value of this study, therefore, lies in demonstrating the potential for recording detailed map-related behavior in real-world conditions. However, instead of using a wrapper, this recording should be carried out using a **purpose-built application** designed to collect data unobtrusively

in the background while relying on minimally invasive and privacy-preserving sensing techniques.

In summary, research on the behavioral context of map use is still in its infancy. Existing studies that analyze log data and interaction traces rarely focus on specific mobile map applications. When they do, they are seldom linked to the intrinsic dimension. They describe *what* users are doing in an application without explaining *why* this matters regarding navigation abilities or emotional states. Additionally, these traces are frequently disconnected from the environment in analyses, leaving the chain between the user, the interaction, and the environment incomplete.

To complete this chain, we must understand why certain behaviors or states emerge more readily in some situations than in others. This leads directly to the following section, which addresses the extrinsic context, specifically how environmental structure and urban form shape spatial cognition.

Extrinsic Context: How the Environment Shapes Spatial Cognition

To fully understand how spatial cognition is affected by the reliance on GPS-based mobile map applications, it is important to consider how the built environment supports or hinders spatial knowledge acquisition. Human navigation has never occurred in a vacuum; it has always been, and remains, embedded in a physical context, the structure, organization, and complexity of which directly influence how individuals learn and remember a space (Montello, 2005).

It is particularly important to study urban areas because they represent the dominant human habitat and are the most cognitively demanding navigation environments. The United Nations estimates that more than 55% of the world's population was living in cities in 2018, and this percentage is expected to reach almost 70% by 2050 (United Nations Department of Economic and Social Affairs, 2019). This means that for most people, spatial behavior takes place in cities rather than in rural or natural settings.

The navigation of urban environments presents a particular challenge. Their road networks tend to be dense and irregular, with many decision points such as intersections, turns, and dead ends. In addition, visual complexity is high, with a constant flow of people, vehicles, and architectural features competing for attention. This overload requires individuals to maintain their attention, filter out distractions, and update their spatial understanding in real-time. Thus, as well as being critical environments for daily navigation, cities are also ideal testing grounds for studying the interaction between environmental complexity and individual differences in spatial perception. The seminal work of the renowned urbanist Kevin Lynch, published in 1960 under the title “The Image of the City,” laid the foundation for what would become the central questions in cognitive geography (Lynch, 1960).

Lynch's work demonstrated that both the clarity and structure of urban environments play crucial roles in how people form their mental representations of space. Lynch identified five key elements: **paths** (the routes people take); **edges** (limits such as walls or rivers); districts (areas with their own identity); **nodes** (strategic intersections or junctions); and **landmarks** (recognizable visual anchors). Together, these elements contribute to a city's “legibility”, i.e., how easily its structure can be understood, remembered, and navigated (Lynch, 1960).

Building on Lynch’s insights, cognitive geography research has expanded our understanding of how environmental structures shape spatial cognition. Although Lynch originally identified the five elements that make a city “legible” from an urban planning perspective, his work also laid the groundwork for investigating how individuals form internal representations of space (i.e., mental maps) based on their interactions with the built environment (see (Montello, 2018)). This has been further formalized by identifying the following three environmental properties that critically influence effective wayfinding and the strategy an individual adopts (Montello & Sas, 2006):

Differentiation refers to the degree to which environmental elements can be distinguished based on attributes such as shape, color, and architectural style, supporting both recognition and memory encoding;

Visual accessibility refers to how visible destinations and landmarks are from various points to help with orientation and encourage allocentric reasoning;

Layout complexity is defined as the irregularity of the road network, including curves and angles, which affects the strategies employed to build mental representations.

These characteristics are central to how individuals construct, update, and use mental representations of their surroundings. A well-differentiated, visually accessible, and moderately complex environment fosters more accurate and flexible spatial learning, which is particularly important during real-world urban navigation (Montello & Sas, 2006). Conversely, when a city lacks a clear spatial structure or distinctive visual features, individuals struggle to create accurate mental maps of it. As a result, they can resort to employing ineffective strategies, such as trial and error, backtracking, or constant checking of GPS-assisted navigation tools, instead of developing an internal understanding of the environment (Ishikawa et al., 2008; Montello, 2009).

Much of the foundational research related to the aforementioned theories has relied on qualitative, small-scale methods. These methods include map sketching, in which participants draw their internal representation of space (Boschmann & Cubbon, 2014; Kitchin, 1994; Schwering et al., 2014); verbal reports, in which they describe their orientation strategies or spatial experiences (Tversky, 2003); and interviews conducted on site during or after completing orientation tasks (Gifford, 2014; Kitchin, 1994). These methods are widely used in the fields of environmental psychology and cognitive geography and have provided fundamental insights into how individuals learn to mentally construct space (Gifford, 2014; Ishikawa & Montello, 2006).

However, despite their value, such approaches are usually context-specific and difficult to replicate (see Golledge, 2004). They offer only a limited means of objectively linking internal representations to measurable properties of the built environment. Consequently, it is challenging to systematically link individuals’ mental models of space to quantifiable characteristics of urban form or to comprehend how the structure of real environments influences spatial learning and navigation behavior. This limits the generalizability of findings across different contexts and constrains the ability to test theories on a large scale (Montello, 2002). A significant development in this regard occurred when urban analysis was first applied to cognitive research. For example, researchers have begun to use space syntax, primarily used in architectural theory and urban planning (see Penn, 2003), more systematically in cognitive geography and environmental psychology. This change has also been driven by advances in network science, the availability of open geospatial data (such

as OpenStreetMap), and the development of computational tools such as OSMnx (Boeing, 2017).

Such tools facilitate the modeling of urban road networks as graphs in which intersections are represented as nodes and roads are represented as edges. They can also be used to calculate metrics such as betweenness centrality (i.e., the frequency with which a segment is found on the shortest paths), closeness centrality (i.e., the average distance from all other nodes) (Boeing, 2017), and to derive additional metrics such as orientation entropy (i.e., the variability of street angles) (Coutrot et al., 2022).

These methods enable researchers to move beyond qualitative descriptions to systematically test how the built environment shapes cognitive outcomes such as navigation performance, spatial memory, and even cognitive development. For example, Watts et al. (2015) found that people living in more spatially integrated environments (i.e., those with more accessible and interconnected road networks) exhibited significantly improved basic cognitive performance and a slower rate of cognitive decline over a two-year period. Interestingly, these effects persisted even after controlling for physical activity, suggesting that the urban form itself may directly support the cognitive resilience of older populations (Watts et al., 2015).

On a global scale, Coutrot et al. (2022) analyzed the spatial orientation skills of more than 400,000 people in 38 countries using data from the mobile game Sea Hero Quest. They quantified urban complexity using entropy measures derived from OpenStreetMap, based on the road network of each participant’s childhood city. The results revealed a strong association between growing up in a city with irregular and non-grid-like layouts (e.g., London or Prague) and better adult navigation performance (Coutrot et al., 2022). This effect was robust and comparable to a five-year difference in spatial ability, highlighting the lasting impact of early exposure to urban development (Coutrot et al., 2022).

Furthermore, Xu et al. (2022) demonstrated that early exposure to urban environments is associated with structural brain development. The researchers developed the UrbanSat index, which utilizes satellite-based measures of urbanicity and neuroimaging data from over 3800 young adults in China and Europe to quantify aspects such as housing density, infrastructure, and land use during childhood. They found that growing up in dense, stimulus-rich urban areas was associated with measurable changes in brain regions involved in spatial cognition, attention regulation, and executive functioning. These neural changes were developmentally specific: exposure to urban environments in early childhood was associated with alterations in areas of the brain responsible for spatial processing, while exposure during adolescence affected regions linked to decision-making and stress regulation (Xu et al., 2022).

Urban environments do not affect all navigators equally. Their complexity is filtered through individual differences in cognition and emotion.

Recent experimental research has begun to investigate the interactions between internal traits and external environmental load in visually complex, cluttered urban-like settings. Rather than simply adding noise to performance, these individual traits actively shape the cognitive strategies employed and the degree to which the user benefits from or suffers due to various environmental features. An interesting example is provided by studies on the influence of environmental load and internal emotional states on spatial learning (Nori et al., 2023; Piccardi et al., 2011; Topete et al., 2024). In one experiment, participants navigated through a realistic virtual environment enriched with dynamic visual distractions

designed to simulate a high cognitive load. The results revealed a significant interaction: the performance of those with higher levels of spatial or general anxiety deteriorated as environmental complexity increased (Nori et al., 2023). This effect was particularly pronounced for allocentric learning, i.e., the ability to construct a flexible mental map of the environment. Participants with high anxiety struggled with perspective and mental rotation tasks, while those with low anxiety maintained stable performance (Nori et al., 2023). These findings suggest that visual clutter and perceptual overload disproportionately alter spatial learning in individuals predisposed to anxiety by activating latent vulnerabilities. Nori et al. (2023) extended this theory by demonstrating how environmental complexity exacerbates these tendencies. Crucially, when the virtual environment was simplified, the performance gap between participants with high and low anxiety narrowed significantly, implying that designing environments with reduced cognitive load could help level the playing field (Nori et al., 2023). Perceptions of one's own spatial abilities also play a significant role. Using the SBSOD (Hegarty et al., 2002), Nori et al. (2023) found that participants who had low confidence in their spatial abilities tended to use egocentric strategies, memorizing turn sequences instead of constructing allocentric representations (Nori et al., 2023). Although these strategies may be effective in the short term, they hinder flexible navigation and deeper spatial learning, reflecting the concept of cognitive offloading (Parush et al., 2007). This is particularly important in urban environments, where visual density, noise, and spatial irregularity are typical. In such contexts, the interaction between the shape of the city and the characteristics of its users is crucial not only for moment-to-moment navigation but also for developing long-term spatial competence. Recent findings support this notion. For example, Topete et al. (2024) demonstrated that individuals with high spatial anxiety or poor orientation skills were more likely to rely on GPS for turn-by-turn directions. This behavior was context-sensitive, occurring more frequently in unfamiliar areas than in familiar settings (Topete et al., 2024).

To summarize, navigation is always embedded within the built environment, and the characteristics of the urban form influence how people acquire and use spatial knowledge. According to research on cognitive geography, environments with clear spatial organization and distinctive visual characteristics facilitate the formation of mental maps. Conversely, irregular layouts or a high visual load can hinder the construction of accurate mental maps and encourage reliance on external aids, such as GPS (Ishikawa et al., 2008; Lynch, 1960; Montello, 2005; Montello & Sas, 2006). Large-scale studies confirm systematic links between urban complexity and cognition, ranging from population-level variation in orientation skills (Coutrot et al., 2022) to structural brain changes associated with childhood exposure to dense cities (Xu et al., 2022). However, these effects are not equal across the entire population. Traits such as spatial anxiety and sense of direction influence how individuals respond to environmental complexity. Those who are anxious or have low confidence are more likely to adopt rigid strategies or rely on GPS in demanding settings (Lawton, 1994; Nori et al., 2023; Topete et al., 2024). What remains unclear is how the use of mobile maps mediates these environmental pressures. While it has been established that urban form shapes navigation strategies and that reliance on GPS varies according to individual traits, we still lack an integrated account of how environmental structure, cognitive predispositions, and mobile map interaction converge in real-world navigation.

This leads us to the final part of this chapter, highlighting the need for predictive approaches that go beyond describing extrinsic factors alone, instead capturing the combined influence of person, place, and technology to account for when a user will engage with a mobile map application.

2.2.3 Prediction of Mobile Map Use

Most mobile map applications still operationalize ‘context’ narrowly and are built and developed primarily for algorithmic efficiency. Their primary objective is to compute and deliver the shortest or fastest route with minimal delay. Although this optimization is computationally effective, given that the shortest or fastest routes are well-defined problems that algorithms can solve quickly and reliably at scale, it does not account for how people experience space. Landmarks, paths, and spatial structures act as anchors that make places recognizable and memorable. While navigation applications that replace these cues with abstract, step-by-step instructions, such as ‘Turn left in 200 meters’ or ‘Merge onto the highway’, may provide efficient routes, they omit the information individuals rely on to orient themselves, build mental maps, and feel confident while moving through the urban environment. This mismatch is rooted in the early development of mobile map applications, which did not follow cartographic theory or cognitive principles and instead emerged from a wave of technological innovation in the late 1990s and early 2000s when GPS positioning and mobile networks became widely available, giving rise to LBS (Rao & Minakakis, 2003; Raper et al., 2007b).

LBS could take advantage of the new positioning capabilities and wireless connectivity to provide information such as nearby POIs, emergency alerts, or directions tailored to a user’s current location (Raper et al., 2007b). Their development was primarily driven by advances in telecommunications infrastructure and middleware (network protocols), as well as service platforms that enabled mobile operators to manage and distribute location data between devices and applications (Cupper et al., 2006). The real turning point came with the launch of **Apple’s iPhone in 2007** and the early **Android** smartphones, such as the **HTC Dream in 2008**. For the first time, these devices introduced operating systems (i.e., iOS and Android) that offered standardized built-in frameworks specifically designed to manage sensors and connectivity efficiently. This marked a technological breakthrough: smartphones could support real-time location updates, continuous background detection, and intuitive user experience capabilities that earlier mobile devices could not (Krajci & Cummings, 2013; Reid, 2018).

With the introduction of **application programming interfaces (APIs)** (Lee, 2012; Meier, 2012), any authorized application (such as Google Maps) could easily request current location data or receive continuous positioning data without directly managing hardware-specific GPS (Lee, 2012; Meier, 2012; Singhal & Shukla, 2012). Finally, to accelerate innovation and expand the ecosystem beyond in-house applications, major platform providers such as Google and Apple (for Android and iOS) launched **Software Development Kits (SDKs)** in 2008. These SDKs bundle the APIs with additional tools, including code libraries, integrated development environments (IDEs), testing emulators, and documentation. For the first time, this allowed third-party developers to easily create applications that could access location updates and integrate location-based functionality into various services and applications (Lee, 2012; Meier, 2012).

The democratization of development has led to the rapid growth of localization applications in various fields, such as social networks, mobility and transportation, and the tracking of physical activity (Huang et al., 2018). One of the most transformative innovations was turn-by-turn navigation. Enabled by the combination of real-time GPS tracking, dynamic routing algorithms, and interactive map interfaces, these applications have redefined users’ daily access to spatial information and transformed human navigation (Parush et al., 2007).

A new era began in 2009 with the release of the **Google Maps Navigation** application for Android 2.0 mobile devices (Keith, 2009). This new application came with everything one would expect to find in standalone car navigation devices (e.g., Garmin or TomTom), such as 3D views, turn-by-turn voice guidance, and automatic re-routing. However, unlike most navigation systems, which often relied on local storage, Google Maps Navigation was designed from the ground up to take advantage of the mobile internet connection, ensuring access to the most current map and business data in real time (Keith, 2009). This was crucial and introduced a series of advantageous features offered free of charge, eliminating the need for costly GPS devices or specific navigation subscriptions (typically **paid monthly subscriptions** offered by mobile network providers in partnership with GPS software companies).

The launch of Android with Google Maps marked a milestone in the way geospatial data were delivered and monetized. By making Google Maps and its APIs default tools for both users and app developers, Google created a complete, vertically integrated system for collecting data, providing services, and generating revenue (McQuire, 2019). This setup allowed Google to collect detailed high-frequency location data from millions of Android devices, which often ran silently in the background. This constant stream of passive and active data allowed for the detection of real-time traffic and later supported additional features, such as providing pedestrian, cycling, and public transit information, as well as the identification of road changes and patterns of business activity (Google, 2020).

Google’s efficient model for designing navigation systems soon became the industry benchmark. Commercial competitors, such as Apple, had to respond to Google’s rapid expansion and disruptive decision to offer free turn-by-turn navigation. However, Apple’s attempt to gain independence by launching its own map service in 2012 struggled to achieve comparable accuracy and reliability (Apple Inc., 2012). Open-source initiatives also emerged, most notably the OpenStreetMap (OSM) project, which championed transparency and community-driven cartography (Bennett, 2010). However, even applications built on OSM data (e.g., OSMand) often echoed Google’s interface conventions and routing logic. This was the result of pragmatism rather than a matter of design preference: by the early 2010s, Google Maps had established the design conventions for digital maps and the way users learned to interact with them. Academic research on prediction has been shaped in similar ways. Thanks to Google’s global coverage and easily accessible APIs, numerous studies adopted Google Maps as their default experimental platform or relied on its data services for analysis. This research, therefore, often reinforced and advanced technical performance, but it left the **intrinsic** and **behavioral** dimensions of navigation largely unexamined.

Collectively, all these developments rapidly reframed navigation as a computational and optimization problem related to calculating the most efficient way to guide users from point A to point B. This efficiency paradigm prioritizes speed, distance, and traffic minimization over cognitive engagement and environmental exploration, as defined by the classic concept of navigation. At the core of this shift lies an ensemble of predictive components designed to automate key aspects of navigation. Modern mobile map applications draw on real-time data, historical user behavior, and machine learning models to anticipate user needs and automate decision-making. These core predictive components integrated into everyday navigation applications include:

Destination Inference which allows for suggestions or auto-completes likely destinations without requiring the user to enter them manually. This is possible by analyzing a combination of data sources such as GPS traces, search history, and frequently visited locations. Typically, it relies on Markov models to detect sequential location patterns (Ashbrook & Starner, 2003), Bayesian inference for mid-journey prediction (Krumm & Horvitz, 2006), and location clustering used to identify common patterns.

ETA Prediction estimates how long it will take to reach a destination and updates that estimate dynamically as conditions change. Early ETA models depended on static traffic averages derived from loop detector data embedded in roadways (Herrera et al., 2010). Over time, these models have evolved to incorporate real-time traffic data from mobile devices and road sensors, historical traffic patterns, aggregated speed profiles, and travel mode-specific behavior. Today, ETA prediction often uses advanced segment-level models powered by deep neural networks and graphs (GNNs) (Derrow-Pinion et al., 2021).

Route Optimization involves finding the most time-efficient route based on real-time conditions. It combines information about the underlying road network, live traffic data, user-specified preferences (e.g., avoid tolls), travel mode, and environmental factors (e.g., weather, time of day). Early implementations included the classic path-finding algorithms (e.g., Dijkstra’s (Dijkstra, 2022) and A* search (Hart et al., 1968)). The latter later expanded to handle time-dependent routing (Baum et al., 2016). More recent apps leverage adaptive graph-based models using GNNs to dynamically compute optimal paths under changing conditions (Cui et al., 2019).

POI Search and Ranking surfaces and prioritizes relevant POIs based on a combination of user context (such as current location, time of day, or movement speed), behavioral history (such as past searches, visited places, and in-app interactions), and the semantic intent of the user’s query during navigation. Data include GPS trajectories, current and past locations, temporal context (e.g., time of day), search and location histories, and in-app interaction patterns of behavior. The early methods relied on spatiotemporal pattern mining techniques to identify frequent destinations (Zheng et al., 2009), later evolving into collaborative filtering with geographic weighting (Ye et al., 2011) and matrix factorization approaches that incorporate spatial constraints (Lian et al., 2014). More advanced hybrid models combined proximity, popularity, social, and semantic data (Zhang & Chow, 2015). The most recent developments employ large language models (LLM) to interpret natural language queries and generate personalized POI suggestions (Google Inc., 2024). These predictions are typically delivered through dynamic map pins, ranked lists, and AI-based suggestions from open-ended or contextual prompts (e.g., “quiet bars near the lake”).

These features form the backbone of contemporary navigation applications. Despite significantly improving responsiveness to environmental changes in real-time, they remain narrowly focused on optimizing external efficiency metrics. Consequently, these systems lack the capacity to interpret the situational intentions, cognitive states, or emotional needs of users during navigation. This limitation is rooted in the following two design

constraints, which are embedded in the current GPS-assisted navigation applications upon which these modern apps are developed:

One-Size-Fits-All Prediction Model: Current navigation applications rely on generalized predictive models that use the same logic for all users. These models primarily rely on population-level data or static profiles derived from external metrics, such as average travel time and traffic density. Consequently, all users receive the same suggestions, routes, and timing estimates regardless of their spatial ability, familiarity with the area, or fluctuating needs.

Lack of transparency: Predictive architectures use complex and opaque models, often based on proprietary algorithms that do not allow the user to understand why a particular destination, route, or POI is suggested. The companies that develop these commercial applications do not disclose the criteria used to generate recommendations, hindering a critical understanding of how the system works.

The development of a more human-centered predictive layer has been widely anticipated. Cartographic and cognitive theories have consistently argued that the same methodological approaches that underpin traffic and mobility forecasts, spatial-temporal modeling, sequence learning, and context fusion can, in principle, be applied to user-centered navigation. Instead of inventing entirely new algorithms, a conceptual shift is needed, from only predicting the **external conditions of a journey**—traffic flows, travel times, or optimal routes—to modeling the **behavior patterns of navigation** and understanding how these patterns are influenced by both **individual differences** and the **environmental context** in which navigation occurs. Examples from other domains demonstrate the feasibility of this approach. A relevant line of work is predicting the next application used, where the goal is to anticipate the app that a user will open next based on a sequence of recent app launches (Katsarou et al., 2022). Applied to navigation, this means treating map use on a smartphone as a temporally structured stream of interactions rather than as isolated taps. Models designed for sequential data, such as long-short-term memory (LSTM) networks, are well-suited for this task, since they maintain a dynamic record of past events and determine the most relevant parts of the recent history to predict the user’s subsequent step (Greff et al., 2017). In this framework, the rhythm of taps before, during, and after consulting the map could signal behavioral states and needs: a hesitant pause followed by a burst of touches may indicate uncertainty, while a steady tempo may reflect confidence. For example, research on tappigraphy demonstrates that the temporal features of smartphone touch interactions can capture subtle variations in users’ behavioral and cognitive states (Balerna & Ghosh, 2018; Ceolini & Ghosh, 2023). By analogy, similar temporal patterns could be derived from mobile map use. Predictive systems capable of detecting such patterns could adapt in real time, for example, by adjusting the style of instructions when hesitation is inferred.

Sequence-to-sequence architectures extend RNN models by employing an encoder-decoder design that anticipates a short horizon of upcoming actions (Sutskever et al., 2014). This makes them particularly suitable for navigation, where predicting several consecutive steps is more informative than analyzing a single tap. Attention mechanisms further refine this process by focusing on the most informative segments of recent behavior. This enables the model to distinguish between, for example, a pause followed by a burst of taps and a steady rhythm. This selective focus improves accuracy and makes predictions more interpretable.

Finally, the meaning of an interaction sequence depends on the broader context in which it is situated. To capture this context, mobile computing researchers have explored **fusion approaches** that integrate behavioral sequences with contextual signals such as the time of day, day of the week, network status, and nearby POIs (Wang et al., 2021). While not specific to navigation, these studies demonstrate that by combining behavioral and contextual data, patterns can be revealed that are invisible in interaction data alone. This line of work highlights the broader potential of context-aware prediction, which can also inform the design of future navigation systems.

Overall, the literature illustrates a gradual shift in prediction research, from modeling **extensive conditions**, such as traffic or travel times, to approaches that also consider **behavioral patterns** and their contextual embedding. While existing studies suggest that this is possible, the integration of all three levels of context—**intrinsic**, **extrinsic**, and **behavioral**—into a unified predictive framework for mobile map applications remains largely unrealized.

2.3 Summary

This chapter has traced the development of a robust theoretical foundation for mobile map design by reviewing research in GIScience, HCI, mobile cartography, and cognitive science. Studies on retrieval and filtering have shown that spatial information must be relevant to the user’s situation (De Sabbata, 2013; Raper et al., 2007b; Reichenbacher, 2007). Visualization research has emphasized that mobile maps must be designed in accordance with human perceptual and cognitive processes to ensure that representations support attention, memory, and spatial reasoning (Fabrikant & Lobben, 2009; Montello et al., 2018). Interaction studies have outlined the dynamic aspects of mobile map interfaces and the central role of interaction embedded in daily activity (Roth, 2013). Prediction studies have shown that navigation is an intentional, goal-directed process that unfolds over time, and that systems should anticipate user needs as these processes develop (Raubal et al., 2004). Taken together, these contributions demonstrate that context cannot be understood as location alone, but must encompass the broader conditions that shape navigation. Therefore, context must encompass the extrinsic structure of the environment (Coutrot et al., 2022; Lynch, 1960; Montello & Sas, 2006), the intrinsic traits and states of the user (He & Hegarty, 2020), and the behavioral traces that reveal how interaction unfolds in real time (Savino et al., 2021). Effective mobile maps must integrate these three dimensions rather than focusing on external efficiency metrics alone (Bartling et al., 2023).

Three persistent obstacles explain why these principles have not yet been realized in everyday applications. The first is that collecting mobile map use data in the wild is inherently difficult, as it requires unobtrusive methods that can record moment-to-moment interactions without disrupting natural behavior or compromising privacy (Martinez-Martin et al., 2018; Montag & Baumeister, 2023; Reichenbacher et al., 2022).

The second issue is the lack of data integration, as intrinsic, extrinsic, and behavioral contexts are typically studied in isolation, producing fragmented results rather than a coherent picture of how navigation unfolds in practice (Bartling et al., 2022, 2023). The third issue is that existing systems preclude holistic, accurate prediction by continuing to maintain a narrow focus on extrinsic metrics such as ETA or route length and rarely incorporating behavioral traces (Ballatore & Bertolotto, 2015). These models also seldom

integrate intrinsic traits. Consequently, predictive models remain blind to the combined influence of person, place, and interaction, which theory has long recognized as essential to mobile navigation.

The following chapter introduces the methodological framework developed in this thesis to address these three issues to facilitate data collection. Guided by the principle of parsimony, unobtrusive tappigraphy is employed to capture fine-grained interaction traces in the wild (Balerna & Ghosh, 2018; Ceolini et al., 2022). This method provides ecologically valid behavioral data while protecting privacy (Reichenbacher et al., 2022). These behavioral traces are then applied in combination with intrinsic measures of spatial ability, spatial anxiety, and gender (He & Hegarty, 2020; Hegarty et al., 2006; Lawton, 1994), as well as integrated with extrinsic descriptors of urban form, such as network centrality, circuitry, and orientation entropy (Boeing, 2017; Coutrot et al., 2022). This will provide a comprehensive understanding of how mobile map applications are used during navigation in a real-world environment. This thesis then presents a context-fusion LSTM architecture that predicts near-term interaction patterns from the temporal sequence of touches, enriched with intrinsic and extrinsic features. Thus, the methodological design not only extends existing work but also operationalizes the central claim of this review: mobile navigation must be understood and modeled as the joint interplay of intrinsic, extrinsic, and behavioral contexts.

3

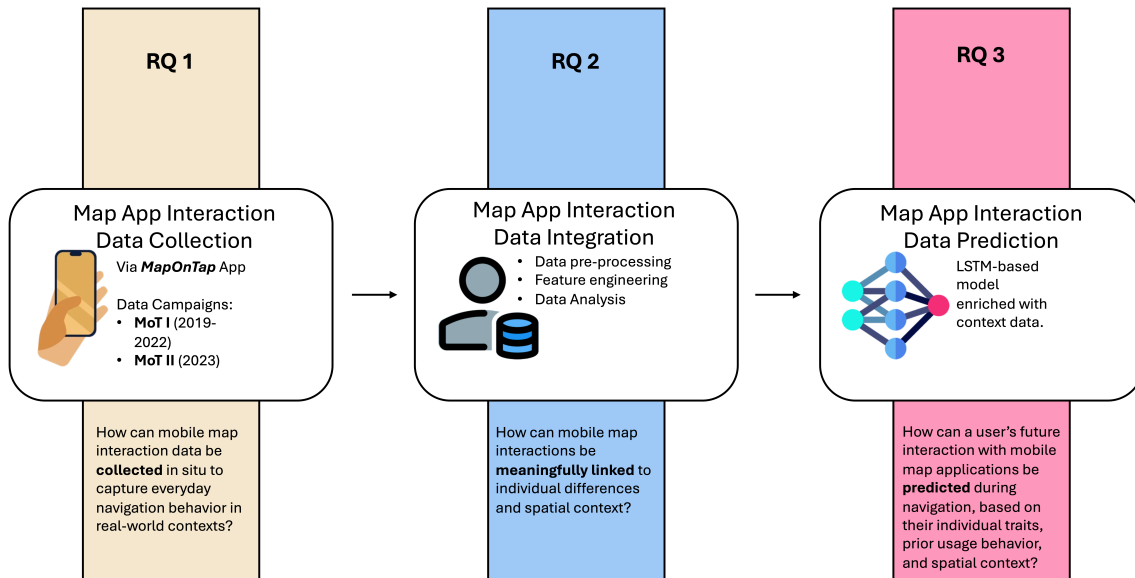
CHAP-
TER

Methodology

*“Now produce your explanation
and pray make it improbable.”*
Oscar Wilde, *The Importance
of Being Earnest*

This methodology chapter establishes a comprehensive and reproducible pipeline for investigating context-aware mobile map interactions across the following three stages: (1) the unobtrusive collection of map–app interaction data in everyday settings, (2) the systematic integration of intrinsic, behavioral, and extrinsic contextual elements to facilitate the understanding of real-world mobile map interaction, and (3) the predictive modeling of map–app interaction using advanced neural architectures that integrate sequential and contextual information.

Figure 3.1.: Overview of the project workflow, aligning each thesis research question (RQ1–RQ3) with its corresponding methodological stage: data collection, integration, and prediction.



Each methodological stage is carefully aligned with its corresponding research question (see Section 1.4, Research Questions and Approach), ensuring that the data, characteristics,

and analyzes directly address the identified knowledge gaps (see Figure 3.1).

Transparency and Reproducibility

To ensure transparency and facilitate replication, all analysis code and data processing scripts developed for this thesis have been made openly available in the project’s [GitLab](#) repository. Supplementary materials referenced in the text, including complete questionnaires, supplementary tables, and additional figures, are provided in Appendix A and are referenced throughout the thesis.

3.1 Map App Interaction Data Collection

To address the **first research question**, which asks how interactions with mobile maps can be collected in situ to capture real-time daily navigation behavior, **tappigraphy** was employed as the primary data collection method for this thesis. As noted in the section on related work 2.2.1, tappigraphy does not record any content that users have viewed or generated (e.g., text messages or photos), but rather documents the timing and sequence of touch interactions, as well as the names and categories of applications running in the foreground (Balerna & Ghosh, 2018). This intentional minimalism reflects the principle of parsimony, collecting only the information that is strictly necessary to reveal meaningful patterns in daily smartphone use without compromising user privacy or utilizing invasive data collection techniques.

Following this principle, tappigraphy served as the methodological foundation for the unobtrusive collection of data on user interactions (i.e., behavioral context) during real-world phone sessions, specifically focusing on mobile map applications. Previous research has demonstrated tappigraphy’s ability to record both small-scale and large-scale smartphone interactions, providing ecologically valid insights into everyday patterns of use, including where a broader category of map applications is used (Reichenbacher et al., 2022). However, the same methodological minimalism that ensures simplicity, privacy, and ecological authenticity also inherently limits the depth of interpretation. Tappigraphy itself offers little insight into the motivations of users, the situational contexts, or the physical environments in which interactions occur. Therefore, despite the elegance and value of its minimalism, using tappigraphy alone inevitably restricts explanatory power and reduces experimental control (Reichenbacher et al., 2022). This thesis advances the tappigraphy approach by extending it into a richer, contextually informed, yet still parsimonious framework by implementing two dedicated data collection campaigns. Importantly, this enrichment does not represent a rejection of parsimony but rather reinterprets the principle at a higher analytical level. Rather than indiscriminately collecting all available contextual data, the refinement strategically incorporates only the supplementary signals that are strictly necessary for meaningful interpretation, including voluntary GPS tracking to geographically anchor interactions to the extrinsic context and standardized questionnaires to capture relevant individual differences to provide the intrinsic context. This careful balance preserved the foundational strengths of tappigraphy (privacy, ecological validity, and methodological simplicity) while significantly enhancing the depth and interpretability of the collected data. This research used a custom Android application, **MapOnTap**, to operationalize this refined approach, seamlessly integrating tappigraphy with GPS logging and questionnaire data into a coherent, unified methodological framework. The following

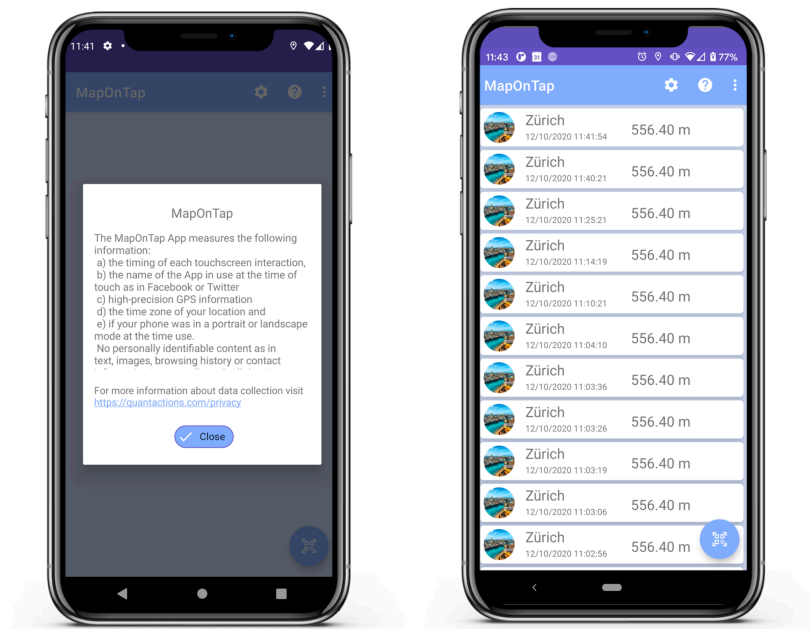
sections offer in-depth descriptions of the design and implementation of the MapOnTap application and the organization and procedure of the two data collection campaigns.

3.1.1 The MapOnTap Application

The name **MapOnTap** refers to both the research project (abbreviated as MoT) and the custom Android application developed specifically for data collection. Participants were instructed to install the MapOnTap app on their smartphones to passively record interaction data related to their mobile map use. The app was developed in collaboration with *QuantAction Sàrl* and is built using their proprietary privacy-preserving tappigraphy technology, which enables unobtrusive logging of touchscreen events and the context in which app usage occurs (see *TapCounter*¹)

MapOnTap (see Figure 3.2) is currently free to download and use and is available only for *Android* devices via the Google Play Store². MapOnTap adheres to the standard privacy and permission policies of the Play Store.

Figure 3.2.: *MapOnTap Android application interface. Left: the privacy dialog shown when the app is launched for the first time. It informs participants about the types of data collected (e.g., touchscreen events with a timestamp, app name, GPS location, and device state) and links to a detailed privacy policy. Right: An example of recorded location entries (city, timestamp, and altitude) logged during GPS-enabled sessions. These interface elements demonstrate the app’s commitment to transparency and unobtrusive design for in-situ behavioral data collection.*



MapOnTap implements tappigraphy by passively recording a series of touchscreen interactions (taps) performed by the user on a smartphone during an active phone session. A phone session is defined as the interval between when the user unlocks the screen and

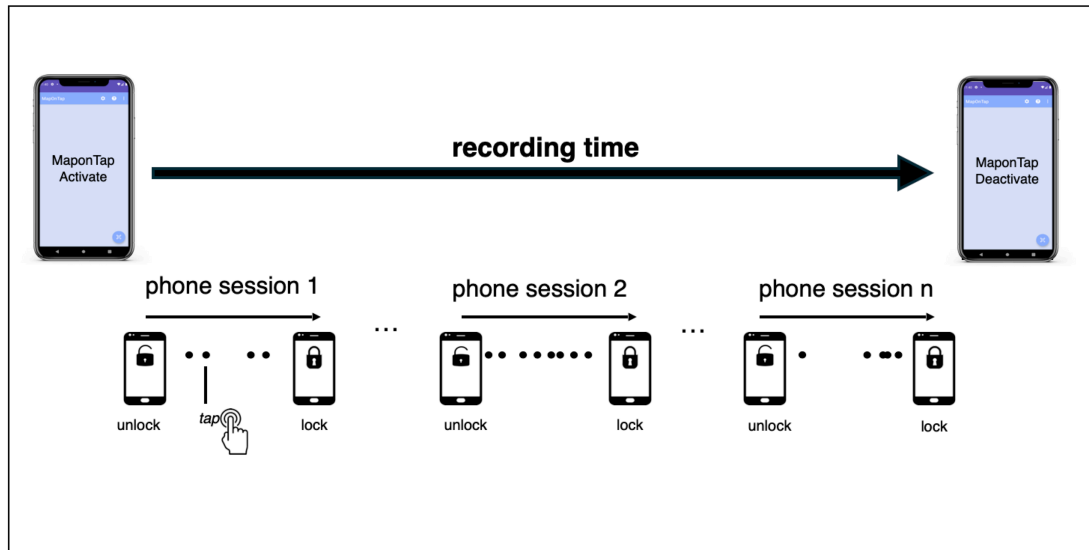
¹<https://play.google.com/store/apps/details?id=com.quantactions.tapcounter>

²<https://play.google.com/store/apps/details?id=ch.uzh.geo.mapontap>

starts tapping, and when the device is either locked again or enters sleep mode (see Figure 3.3).

The MapOnTap application logs all touch events with precise timestamps (in milliseconds), the start and end times of the phone session, and the name (e.g., Instagram or Google Maps) and category (e.g., Social or Travel and Local) of the foreground application according to the Google Play Store. To ensure participant privacy and data security, all tappigraphy data are first encrypted locally with SQLCipher 256-bit AES encryption on the user’s device before being transmitted via secure end-to-end encrypted channels to QuantActions’ cloud infrastructure. Upon upload, each data packet is tagged with a randomly generated pseudonymized participant ID; no direct personal identifiers are used. The encrypted data are subsequently stored in a PostgreSQL database instance running on the *s3it ScienceCloud* platform at the University of Zurich. Access to the database is strictly controlled using role-based access policies, and all data at rest remain encrypted, ensuring compliance with institutional and GDPR requirements.

Figure 3.3.: *Structure of Tap Data Collection Across Phone Sessions. Schematic representation of the MapOnTap data logging process. Touchscreen interactions (taps) are recorded during each phone session, defined as the period between unlocking and locking the device. The app remains active in the background, passively collecting timestamped tap events for every session over the entire recording period, spanning both MapOnTap campaigns (MoT I: 2019–2022; MoT II: 2023), during which participants contributed continuous data for a minimum of two weeks.*



Each mobile device is assigned a randomized device ID within the MapOnTap app. In parallel, every participant is assigned an anonymized participant ID, which enables the secure and pseudonymous labeling of interaction data. The temporary link between the participant ID and the device ID allows data from the same individual to be maintained consistently throughout the study period, including cases where participants switch devices, reinstall the app, or use more than one device (e.g., a smartphone and a tablet). The device and participant IDs are never stored together. At the completion of the study, the link between the device ID and the participant ID will be permanently deleted from the

QuantActions platform and the UZH infrastructure, ensuring that participants cannot be re-identified from the behavioral dataset.

In addition to tappigraphy-linked information, MapOnTap also allows for the optional collection of sensor data from those participants who have provided explicit consent. These data include:

- 📍 **GPS coordinates:** Location samples include latitude, longitude, and altitude, collected every five seconds when significant movement is detected. Adaptive sampling conserves battery life by pausing GPS when the device is stationary, as determined by motion sensors.
- 💡 **Ambient light sensor:** Illuminance (in lux) is recorded every 5 seconds to help infer the indoor vs. outdoor context based on the ambient light thresholds.
- ➔ **Motion sensors:** Periodic readings from the gyroscope (rotation rate) and accelerometer (linear acceleration) detect device movement and orientation (portrait / landscape), and user activity.
- 🧭 **Compass heading:** Orientation relative to magnetic north is derived from the magnetometer and corrected for the device tilt using accelerometer data.
- 🚶 **Mobility state:** The user's mobility is classified by the Google Activity Recognition API, which combines multiple sensor inputs to determine whether they are walking, running, cycling, driving, or stationary.

Sensor data are only collected with explicit permission from the operating system and the participant. Participants can deactivate individual sensors or opt out of sensor logging entirely at any time through the app settings or device privacy controls.

All sensor data are transmitted directly to the PostgreSQL server hosted on S3IT through encrypted channels. To maximize privacy and prevent participant re-identification or crosslinking, sensor and tappigraphy data are stored in logically separate tables within the same secure database. These tables share no direct identifiers or join keys. This data architecture enforces the technical and logical separation of behavioral and sensor information, further safeguarding participant privacy.

The protocol was designed to comply with institutional ethics requirements and best practices for responsible digital phenotyping research, including adherence to the principles of data minimization and participant autonomy.

Most importantly, the MapOnTap app does not collect or transmit any **personally identifiable information (PII)**. Personal messages, emails, phone numbers, SMS content, visited websites, contact lists, calendar entries, and text input from apps are never accessed or recorded. Additionally, the app does not log the screen coordinates of touch events; only the occurrence and timing of touches are recorded, never their position or associated content (e.g., passwords).

Having described the design, functionality, and privacy safeguards of the MapOnTap application, the following section details its deployment through the two data collection campaigns: MoT I and MoT II.

3.1.2 Data Collection Campaigns

Two separate data collection campaigns were conducted using the MapOnTap application. While both campaigns used the same MapOnTap application and core tappigraphy logging infrastructure, they differed in scope, participant sample, and the types of data collected (see Table 3.1 for an overview).

Table 3.1.: *Overview of data types collected during the two MapOnTap campaigns (MoT I and MoT II). Note: SBSOD, SA, and demographic data were collected using external questionnaires, not through the MapOnTap app.*

Data Type	MoT I	MoT II	Collection Method
Tappigraphy Data	✓	✓	MapOnTap app
GPS Data (if consented)	✓	✓	MapOnTap app
Sensor Data (if consented)	✓	✓	MapOnTap app
SBSOD (Sense of Direction)	✗	✓	Questionnaire (Hegarty et al., 2002)
SA (Spatial Anxiety)	✗	✓	Questionnaire (He & Hegarty, 2020)
Demographics (Age, Gender)	✗	✓	Questionnaire

Note. A check mark next to a data type indicates that it was collected during the campaign. Tappigraphy, GPS, and optional sensor logs were recorded via the MapOnTap application. Questionnaire measures (SBSOD; SA) and demographics were collected outside the app.

Participants for both campaigns were recruited via the MapOnTap project website³, which served as the central platform for enrollment and registration. The website provided detailed information on the project’s scientific objectives and data collection procedures and included a comprehensive statement on data privacy and ethics. Eligible participants were between 18 and 85 years old, had an Android smartphone with a built-in GPS sensor, understood the English language, and were capable of providing informed consent. Individuals were excluded if they used non-Android devices, lacked GPS sensors, shared their phones with others, used styluses, or had difficulty following instructions. Recruitment announcements were widely disseminated via the official UZH mailing list, LinkedIn, and the Geography Department Research website.

MoT I

This first data collection campaign was reviewed by the *Ethics Committee of the UZH PhF* and registered under its ethics protocol No. 19.12.6⁴. Data were collected between July 2019 and March 2022 from 53 participants. They were instructed to download the MapOnTap app from the Google Play Store and use their smartphones as they normally would for at least two consecutive weeks. However, since the study was active for two years, participants could continue contributing data for as long as they wanted. They retained full autonomy to cease data collection, disable specific sensors, or uninstall the application at any point without providing a reason. Importantly, no personal demographic information (e.g., gender, age, nationality) was collected during this data collection campaign, which

³<https://microsite.geo.uzh.ch/mapontap>

⁴<http://www.phil.uzh.ch/de/forschung/ethik.html>

purposefully relied only on information collected via MapOnTap. Participants in this campaign received no financial compensation, but were entered into a lottery to win a tablet.

MoT II

The second campaign was also reviewed by the *UZH PhF Ethics Committee* and was registered with protocol No. *23.03.24*. MoT II took place between March and June 2023 and recruited 61 participants. Like MoT I, this campaign collected data via the MapOnTap application, using the same protocols and procedures. However, MoT II included the collection of demographic information (gender and age range). A series of spatial cognition questionnaires was also administered (the complete questionnaires are provided in Section A.1 of the appendix) as follows:

- ✓ **Santa Barbara Sense of Direction Scale (SBSOD).** A 15-item, 7-point Likert scale self-assessment to evaluate individuals' perceptions of their ability to navigate and orient themselves in space, including both positively and negatively worded items (Hegarty et al., 2002).
- ✓ **Spatial Anxiety Questionnaire (SA).** This self-assessment measure includes 12 navigation-focused items, rated on a 5-point scale He and Hegarty, 2020. The instrument combines the eight rephrased items of the original spatial anxiety scale (Lawton, 1994) with the five additional items introduced by Lyons et al., 2018 to ensure a more precise and targeted assessment of spatial anxiety during navigation tasks.

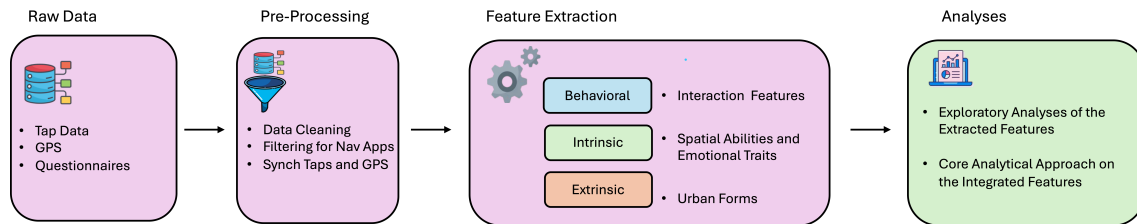
The questionnaires were presented in a randomized order. Data collection for each participant lasted for a minimum of two weeks. After this initial period, participants were notified that they could uninstall the app, as specified in the study's consent form. They were also free to continue using MapOnTap voluntarily. Participants who continued to use the app after the initial period had their additional data processed under the same ethical review and data privacy safeguards. Data recording automatically stopped for everyone at the end of the formal recording period in August 2023. This occurred on the QuantActions platform for tappigraphy and on our server for GPS data. In return for their participation, each participant received a 20 CHF digital voucher at the end of the second week of recording. Unlike MoT I, where participants were entered into a lottery, this direct compensation reflected the increased requirements of the MoT II data collection, including GPS tracking and questionnaire completion.

These two data collection campaigns gradually expanded on the minimal tappigraphy approach by selectively introducing additional measures. This refinement maintained a careful balance of methodological simplicity, explanatory depth, and participant privacy, ensuring that any additional data were collected strictly on a voluntary basis and explicitly aligned with specific research objectives. Once these datasets had been established, the next step involved analyzing how these contextual elements enriched the understanding of mobile map app usage. The integrated analyses detailed in the following section provide a comprehensive account of the relationships among interaction patterns, individual differences, and spatial contexts.

3.2 Map App Interaction Data Integration

To address the **second research question**, which asked how mobile map interactions can be meaningfully linked to individual differences and geographic environments, this section outlines the methodological steps taken to integrate map interaction data (behavioral context), intrinsic user characteristics (individual differences captured using standardized questionnaires), and extrinsic geographic features (environmental context derived from GPS data). By combining these distinct data streams, the aim of the analysis was to provide deeper insights into how mobile map applications are employed during real-world navigation. Using the complete datasets from the MoT I and MoT II campaigns, these methodological steps were all implemented in *Python* (v. 3.11.7) using *Visual Studio Code* (v. 1.86). The integration involved three primary stages: **pre-processing**, **feature extraction**, and **data analysis** (see Figure 3.4). The next section begins by describing the **pre-processing** pipeline, detailing the cleaning, filtering, and synchronization of the raw data streams to prepare them for feature extraction and subsequent analysis.

Figure 3.4.: *Map App Interaction Data Integration Pipeline. Raw data from taps, GPS, and questionnaires underwent pre-processing, synchronization, and feature extraction across three contextual domains: behavioral, intrinsic, and extrinsic. Analyses included clustering (grouping interaction patterns), archetypal analysis (identifying representative behavior profiles), correlation (quantifying relationships), and regression modeling (predicting interactions based on contextual variables).*



3.2.1 Data Pre-Processing

Data pre-processing followed a multi-step pipeline designed to ensure the quality, integrity, and analytical relevance of the data. This initial phase involved standard **data cleaning** pipeline procedures, including the removal of duplicates, handling of missing values, correction for data inconsistency, elimination of outliers, and standardization of variables.

After completing the general quality control steps, the next critical task was to isolate phone sessions specifically related to navigation. The MapOnTap application continuously collects data whenever the phone is active, meaning that it logs interactions from all foreground applications, regardless of their purpose. Consequently, the raw dataset contained numerous sessions during which navigation was not taking place; for example, when participants were checking social media, messaging, or using entertainment apps. Therefore, to ensure a meaningful analysis of the interactions with mobile maps, it was essential to carefully filter the data and select only those sessions during which navigation apps were actively used. This allowed us to focus our analyses strictly on navigation-related behavior.

To ensure that our analysis specifically addressed navigation behavior (as required for RQ2), we retained only phone sessions where at least one actively used foreground app was navigation-related. This was done by retaining sessions categorized as “**Travel and Local**” (TL) or “**Maps and Navigation**” (MN) according to the Google Play Store (see Section A.2, Table A.1 for a complete list). The identified apps were then manually inspected to confirm their relevance to navigation. The full list of selected applications is provided in Section A.3 of the appendix; see A.2. We also excluded phone sessions involving the MapOnTap data collection app itself (classified as MN) to prevent possible bias. Although navigation-related applications were the primary focus, information about other application activities occurring within the same phone session was preserved to allow the analysis to incorporate the broader context of digital behavior during navigation.

The final major pre-processing step was **synchronizing the sensor data** with the tappigraphy-based phone session data. For the present study, we focused exclusively on GPS sensor data to establish the spatial context of the mobile map interactions. This required the synchronization of two independent data streams: (1) GPS sensor samples and (2) tappigraphy logs. Aligning these streams ensured that each behavioral event, such as a tap within a navigation app, could be accurately linked to a specific location and time point. Since these two data streams were generated separately, timing discrepancies were possible, due to device behavior, sensor polling intervals, or user privacy settings.

To resolve these discrepancies, we performed a temporal join between touch events and GPS samples. For every touch event, we used a binary search algorithm on sorted GPS times to identify the GPS sample with the closest timestamp. This approach improved computational efficiency from linear to logarithmic time per lookup, which was important given the scale of the dataset.

We set a conservative maximum alignment tolerance of 30 seconds and considered as valid matches only GPS samples collected within the 30 seconds before or after a touch event. Touch events that lacked a GPS sample within this window were marked as missing to prevent incorrect spatial attributions. Phone sessions with fewer than five valid GPS–tap matches were also excluded from further analysis to maintain a minimum threshold for spatio-temporal coverage and data quality.

All synchronization, filtering, and data joining were implemented in *Dask* (v. 2023.11.0), which provides distributed and parallel processing and allows for scalability among many participants and sessions. *PyArrow* (v. 14.0.2) was used for efficient on-disk storage and high-throughput input/output (I/O), leveraging the Apache Arrow columnar memory format and Parquet file structure for the efficient handling of large tabular datasets.

This pipeline reduced the number of GPS-linked sessions available for downstream analysis. This was expected given that GPS tracking was voluntary and not continuous. We ensured that subsequent analyses, such as spatial feature extraction, would be based on robust and representative data by retaining only sessions that demonstrated reliable spatiotemporal alignment. A detailed summary of session counts before and after synchronization, as well as statistics on session exclusion and GPS–tap matches, is provided in the appendix, Section A.5.

After cleaning and synchronizing the dataset, we moved on to the next stage of the analysis pipeline: feature extraction. This involves transforming the raw data into interpretable variables that represent the three contextual dimensions—behavioral, intrinsic, and extrinsic—that are central to this thesis. Extracting these features was essential

for comprehensively understanding everyday mobile map interactions and enabled us to meaningfully relate touch dynamics to user differences and urban contexts.

3.2.2 Feature Extraction

Following the data cleaning and pre-processing steps, interpretable features were extracted to optimally capture the behavioral, intrinsic, and extrinsic contexts relevant to our research questions.

Behavioral Context

To fully capture the behavioral context of mobile map usage, tappigraphy characteristics were calculated on two complementary scales (see Table 3.2). This dual scale approach recognizes that human interactions with smartphones occur on multiple temporal scales: the immediate moment-to-moment dynamics captured within individual phone sessions (**session-based features**) and the broader habitual patterns of use that gradually emerge over days or weeks (**participant-based features**).

Session-based features were used to quantify the detailed dynamics of interactions within individual phone sessions. These metrics captured the frequency with which users tapped on map applications (map taps), the proportion of taps devoted specifically to map applications, and the intensity of the interaction, measured as taps per second (tap rate). Furthermore, the proportion of time spent on map apps relative to the entire phone session reflected the user’s focus on navigation tasks. Switching behavior was measured by considering the total number of transitions between apps (all app switches), as well as the proportion of transitions involving maps. Finally, variability in interaction rhythm was characterized according to the mean and coefficient of variation of inter-tap intervals (mean ITIs and CV-ITIs, respectively). Combined, these metrics offered a detailed picture of a user’s engagement with navigation applications, capturing the intensity and regularity of their mobile map usage.

Participant-based features were calculated by aggregating data across all sessions for each participant, thereby summarizing their typical navigation behavior throughout the recording period and allowing links to be identified with individual abilities and traits. Key participant features include the mean interval between map usage events (**inter-map interval; IMI**) and the **session initiation rate (SIR)**. IMI captures typical intervals between one map tap and another. For each participant, we calculated the time (seconds) between the start of consecutive navigation sessions (a navigation session was considered any phone session with one tap in the MN or TL categories), applied a $\log_{10}$ transformation to reduce the right bias, and the 25th percentile of the participant’s transformed values was retained as the IMI. Lower IMI values indicate shorter (more frequent) typical intervals; using the lower quartile reduces sensitivity to occasional very long gaps. The SIR represented how often a participant initiated new navigation sessions throughout the full distribution of phone sessions and was defined as the inverse of the mean of the $\log_{10}$ transformed IMI (which has no unit). Larger SIR values indicate more frequent initiations of navigation sessions. Sessions with zero MN/TL taps were excluded, and only data from participants with at least two navigation sessions (that is, less than one intersession interval) were retained. Together, these metrics reflect how frequently and regularly participants return to navigation apps.

Table 3.2.: *Overview of tappigraphy features extracted within phone sessions and across participants. Session-based features capture the detailed interaction dynamics within individual phone sessions. Participant-based features summarize users’ overall interaction patterns throughout the entire recording period, facilitating comparisons across individuals with different navigation behaviors.*

Session-Based Features	
All Taps	Total count of touchscreen taps during a phone session.
Map Taps	Count of taps within map apps during a phone session.
Proportion of Map Taps	Map taps divided by total taps in a session.
All App Switches	Number of foreground app changes during a session.
Switches from Maps	Number of switches from a map app to a non-map app.
Switches to Maps	Number of switches from a non-map app to a map app.
Proportion of Map Switches	Proportion of all switches involving a map app.
Map Duration	Total time (s) spent in map apps within a session.
Session Duration	Length (s) of the entire phone session.
Proportion of Time on Maps	Fraction of session duration spent on map apps.
Tap Rate (taps/s)	Number of map taps per second during map app use.
Mean Inter-Tap Interval (ITI)	Mean time (s) between consecutive map taps.
CV of ITI (CV-ITI)	Coefficient of variation (SD/mean) of ITI in map apps.
Transitions per Minute	Number of app-category switches per minute during the session.
Participant-Based Features	
Inter-Map Interval (IMI)	Mean elapsed time (s) between consecutive navigation sessions per participant. Lower values indicate more frequent map use.
Session Initiation Rate (SIR)	Inverse of the mean log-transformed interval between navigation sessions; higher values reflect more frequent initiation of map sessions.
Tap Density (taps/km ²)	Total number of taps divided by the area (km ²) of the participant’s activity space (convex hull of all GPS locations); quantifies spatial intensity of smartphone use.

Another participant-based feature we characterized was **tap density**, calculated as the total number of taps divided by the area (in km²) of each participant’s activity space, defined as the convex hull of all GPS locations. Tap density quantifies the spatial intensity of smartphone use and facilitates the direct comparison of usage patterns between individuals with different mobility ranges. Refer to Table 3.2 for a comprehensive summary of these features, including detailed descriptions. In addition to the behavioral context, we characterized the intrinsic context of mobile map use by extracting features from two validated psychometric questionnaires and demographic information. The following section details this process.

Intrinsic Context

Intrinsic context was quantified using standardized scores from two validated psychometric questionnaires and self-reported demographic information. Mean scores were calculated for each participant for both the SBSOD and the SA, and then standardized (using z-scores) for analysis. These continuous participant-based measures capture individual differences in spatial ability (SBSOD) and spatial anxiety (SA), and were used as predictor, moderator, and grouping variables in all subsequent statistical models. Self-reported gender (coded as M / F) was included as a categorical intrinsic characteristic (see Table 3.3).

Table 3.3.: *Participant-Based Intrinsic Features Derived From Questionnaires.*

Individual Difference Variables	
SBSOD score (standardized)	Standardized Santa Barbara Sense of Direction score (higher scores indicate better perceived spatial ability).
SA Score (standardized)	Standardized SA Questionnaire score (higher scores indicate greater perceived spatial anxiety).
Gender	Self-reported gender, categorical variable (e.g., male, female, other).

Note. SBSOD = Santa Barbara Sense of Direction scale (Hegarty et al., 2002); SA = Spatial Anxiety scale (He & Hegarty, 2020)).

While participants also reported their age group, this was not included in the analysis due to the highly unbalanced distribution: 28 participants were 25–34 years old, followed by nine aged 18–24 years, six aged 35–44 years, and only three in the 45–64 year old range. Given the limited statistical power for the older age categories and the categorical format of the age data, we excluded age from regression and interaction models. With both behavioral and intrinsic features clarified, the next section turns to the extraction of features related to the extrinsic context.

Extrinsic Context

As discussed in Section 2.2.3, the extrinsic context is often considered a passive backdrop against which navigation occurs in mobile map applications. However, as clarified in 2.2.2, spatial context actively shapes individuals perceptions, interpretations, and mental representations of their navigational experiences (Lynch, 1960; Montello & Sas, 2006). The urban environment, in particular, exhibits substantial variability in terms of spatial configuration, complexity, and street network structure, which imposes distinct cognitive demands that influence spatial cognition (Coutrot et al., 2022). Despite recognizing this, researchers have not yet systematically linked the extrinsic context to interactions with mobile maps.

To capture this largely unexplored dimension, we expanded the modeling of extrinsic context to include urban morphology metrics that capture the structural characteristics of the street network surrounding each navigation session. For each session, we extracted localized spatial features from OpenStreetMap (OSM) using the Python library *OSMNx* (v.2.0.2) (Boeing, 2017). A buffer of 500 meters was calculated around the convex hull of each session’s GPS trace to define the navigated spatial footprint (ranging from 0° to 180°). Within this area, the walkable road network was extracted, projected to metric

coordinates, and annotated with the road segment orientations. From this localized graph, we computed key characteristics that reflect the urban form (see Table 3.4): orientation entropy, centrality of proximity, centrality of betweenness, average degree of nodes, and circuitry.

Table 3.4.: *Urban Form Features Derived From Localized Street Networks*

Urban Form Feature	Description
Orientation entropy	Shannon entropy of road segment orientations grouped into 36 bins (10° steps), adjusted to avoid boundary effects. Higher values indicate irregular, radial layouts; lower values indicate more orderly, grid-like patterns. Captures angular complexity.
Circuitry	Ratio of actual road segment length to its straight-line (Euclidean) distance. Higher values reflect more winding, indirect streets—typically harder to navigate.
Mean closeness centrality	Average closeness of all nodes in the local road network. Higher values indicate that locations are generally fewer steps away from each other, signaling higher spatial integration.
Mean betweenness centrality	Average number of times nodes (intersections) fall on the shortest paths between other nodes. Higher values imply that key junctions serve as critical connectors.
Mean degree centrality	Average number of direct connections per node (intersection). Higher values reflect greater local connectivity and more navigational choices.
Distance from home (km)	Euclidean distance from each session’s centroid to the participant’s home location (modal GPS coordinate); captures spatial context/familiarity for each navigation episode.

Note. These features quantify the geometric and topological complexity of urban form and provide measures of spatial context that are relevant for navigation behavior.

These features were selected to represent both the angular and topological properties of the environment. Centrality metrics (closeness and betweenness) were log-transformed to address right-skewness. All variables in urban form were then standardized (z-scores) before their inclusion in statistical models. Sessions were excluded if they had incomplete GPS traces or if network extraction had failed. Additional pre-processing and quality control steps are detailed in Appendix, Section A.6.

In addition to urban morphology, we incorporated “distance from home” as an extrinsic contextual feature for exploratory analyses. For each session, the home location was defined as the mode of GPS coordinates for each participant across all sessions. We preserved privacy by slightly perturbing (adding random spatial noise of approximately 100 meters) the GPS coordinates, ensuring that individual locations could not be precisely identified. This method protected participant privacy by summarizing location data as a single generalized point (modal location) rather than revealing specific routes or detailed location traces.

We calculated the straight-line (Euclidean) distance from this home location to the centroid of each session’s GPS trace. This “distance from home” metric facilitated the assessment of variations in spatial app usage and navigation behaviors as users move away from their typical home base, allowing us to control for spatial familiarity in subsequent analyses.

In summary, after extracting meaningful features that capture the behavioral, intrinsic, and extrinsic contexts of mobile map interactions, the next step involved systematically analyzing the rich dataset, which is detailed in the following section.

3.2.3 Data Analysis

The final stage of the data integration pipeline involved performing exploratory analyses on both MoT I and MoT II datasets. During this stage, we focused on the basic metrics derived from tappigraphy, GPS data, and questionnaires. These analyses assessed overall data quality and identified preliminary patterns of mobile map use. The insights provided by these analyses guided our choice of analytical strategies and informed our subsequent integrated modeling efforts. Following these initial exploratory studies, the core analyses were designed to uncover clearer and more interpretable mobile interaction patterns across all phone sessions and participants. At both levels, we linked behavioral interaction data with individual traits such as the sense of direction and spatial anxiety, as well as with features of urban form. Our analyses aimed to reveal how behavioral, intrinsic, and extrinsic contexts collectively influence mobile map interactions. Let us start with a brief summary of the exploratory analyses.

Exploratory Analyses of the Extracted Features

We initially performed three sets of exploratory analysis to systematically investigate the three-level contextual features (intrinsic, behavioral, extrinsic) extracted from the MoT I and MoT II datasets. These exploratory steps served two central purposes: to assess the quality of the collected data and to identify preliminary patterns for guiding subsequent analytical decisions. The first set of analyses (Zingaro & Reichenbacher, 2022; Zingaro et al., 2023) focused on the relationship between map interactions and GPS data, using the MoT I campaign and a broader definition of map applications (i.e., those in the “MN” and “TL” categories as defined in Section 3.2.1). The second set (see Zingaro et al., 2024) examined the relationship between user traits and map interactions, leveraging data collected from the MoT II campaign. For this analysis, we employed a more restricted definition of mobile map applications (still “MN” and “TL”, but strictly related to navigation; see Section 3.2.1). Across these exploratory studies, we implemented data analysis pipelines in Python, relying on libraries such as *scipy*, *geopandas*, and *numpy*, and using *matplotlib* and *seaborn* for visualization.

The first exploratory study (Zingaro & Reichenbacher, 2022) employed the distance from home (see Table 3.4) distribution to divide interaction behaviors for mobile maps and non-maps (all the other categories of apps that participants used during the recording period) into different spatial categories: near-home (0–3.5 km), mid-range (100–200 km), and travel-dominant (>200 km). To visualize and further investigate the spatial patterns of smartphone interactions, we applied kernel density estimation (KDE), a nonparametric method commonly used to estimate the probability density distributions of spatial point data (Silverman, 2018). KDE allowed us to identify spatially dense regions of smartphone

interaction relative to participants’ estimated home locations for mobile map and non-map applications. We used tap density to facilitate meaningful comparisons between participants with varying activity levels and travel ranges (see Table 3.2). Using this normalized measure, we compared participants’ tap densities between map apps and non-map apps, enabling an examination of whether the intensity of the interaction differed systematically by app category in relation to spatial behaviors.

A second exploratory study of the MoT I dataset (Zingaro et al., 2023) used two complementary analytical methods: hierarchical clustering analysis (HCA) and archetypal analysis (AA) to reveal potential map application use patterns at varying distances from participants’ homes. The use of a mobile map application was defined as standardized tap-count distributions across predefined spatial distance bins. In preparation for these analyses, taps were binned into distance categories using Fisher–Jenks natural breaks, optimizing within-class homogeneity and between-class variability. The tap distribution of each participant was then z-standardized to ensure comparability across profiles.

HCA is an unsupervised algorithm that forms ordered subgroups that can help individualize data groups that are either more closely or distantly related (Murtagh and Contreras, 2017). We used Ward’s criterion to optimize homogeneity within clusters by minimizing the sum of squares within the cluster (Zingaro et al., 2023). HCA is typically represented by a dendrogram, with the height of the branches representing the distance or dissimilarity between the clusters. To determine the optimal number of clusters, we partitioned the dendrogram to maximize the distance between nodes between trees (Murtagh and Contreras, 2017).

We also applied AA, an unsupervised machine learning technique, for comparison purposes. AA finds the unique combinations of features (i.e., pure types) within a dataset (i.e., archetypes) that best represent its properties (Bartling et al., 2022; Eugster and Leisch, 2009). Data points in the dataset are then positioned on a spectrum between archetypes and are not assigned to only one particular archetype (unlike the results of the cluster analysis). With archetypal analysis, we can assess each data point’s membership of these different archetypal signatures (similar to cluster analysis) while preserving individual differences (Bartling et al., 2022). To assess the quality of the model fit, we used the residual sum of squares (RSS), the sum of the squared differences between the original data points and their corresponding reconstructed points, derived from the archetypes.

A lower RSS indicates a better fit, meaning that the archetypes selected by the model represent the underlying structure of the data more accurately. Conversely, a higher RSS implies that the archetypes do not sufficiently capture the inherent patterns within the dataset. The optimal number of archetypes depends on how RSS changes as additional archetypes are added. Typically, the optimal number is selected at the point where the RSS no longer demonstrates any significant improvement with additional archetypes. This criterion ensures a balance between explanatory power (goodness of fit) and model parsimony (simplicity). Unlike the other analyses, for AA we used RStudio (v.1.4) and the R package, “archetypes” (Eugster & Leisch, 2009).

During the third exploratory study (Zingaro et al., 2024), we extracted the following session-based metrics from the tappigraphy interaction data: app switch frequency (e.g., total number of app switches and total number of map switches); proportion of map switches; duration of map app use per session; total interaction time; and proportion of time spent on the map app (see Table 3.2 for details). We then performed correlation analyses using the complete dataset, further stratifying these analyses by gender. The

rationale for the gender-specific analysis originates from prior evidence indicating that men generally express greater confidence in their sense of direction (Kozlowski & Bryant, 1977), while women tend to experience higher levels of spatial anxiety during navigation (Lawton, 1994). Correlation analyses were conducted using either Pearson or Spearman correlation coefficients, depending on the results of Shapiro-Wilk tests to ascertain the normality of the distributions behavioral scores and transformed variables. All analyses used an alpha significance level of 0.05 to determine statistical significance.

These exploratory analyses provided critical insights to inform the refinement of data collection strategies used for the MoT I and MoT II campaigns. They highlighted distinctive patterns of mobile map usage based on simple extracted features. The next step was to integrate multiple feature dimensions into a cohesive analytical framework employing more advanced statistical and computational methods, which are presented in the next section.

Core Analytical Approach to the Integrated Features

Our core analytical framework was developed by systematically extending the insights gained from the initial exploratory analyses. While the exploratory stage provided an overview of the patterns observed in mobile map interactions, the core analyses were designed to uncover clearer and more interpretable interactions at two complementary levels of granularity: across all phone sessions and across participants. At both levels, we linked behavioral interaction data with individual traits such as sense of direction and spatial anxiety, as well as with features of the urban form. The framework therefore integrated both intrinsic and extrinsic contexts with behavioral traces, enabling the examination of how these dimensions jointly shape mobile map usage. To achieve this, we structured our analytical approach around two interrelated studies.

The **first study** involved combining uniform manifold approximation and projection (UMAP) with density-based spatial clustering (DBSCAN) (see Viswanath et al., 2024) to analyze three log-transformed session-based interaction metrics: interaction frequency (taps per second), interaction regularity (coefficient of variation of inter-tap intervals, CV-ITI), and switching frequency (see Table 3.2 for details). The choice of these three metrics was not arbitrary but was rather informed by our preliminary explorations and understanding of which metrics could robustly capture critical aspects of the dynamics of moment-to-moment navigation applications.

To ensure robust and consistent input quality for dimensionality reduction, we carefully addressed missing data points using feature-wise median imputation, a method known for its resilience to data skewness and outliers. We then standardized the data using z-score normalization, placing all metrics on comparable scales and ensuring balanced contributions to subsequent analytical steps. We chose UMAP precisely because it excels in capturing both local relationships and global structures hidden within complex, high-dimensional behavioral datasets. This preserves the rich nuances inherent in everyday interactions with mobile maps.

Following dimensionality reduction, we applied DBSCAN to identify naturally emerging clusters within the reduced UMAP space. The strength of DBSCAN is its ability to recognize clusters without imposing predetermined numbers or shapes. This approach facilitates the emergence of authentic behavioral patterns directly from the data, undistorted by preconceived constraints or assumptions. We calibrated the DBSCAN parameters ($\epsilon = 1$ and minimum samples = 100) to optimize the sensitivity and specificity of the

clustering process, balancing the detection of interaction modes with the risk of noise or meaningless partitions. After validating the stability and interpretability of these session-level clusters with rigorous statistical testing (using Kruskal–Wallis and Dunn’s post hoc analyses), we integrated these fine-grained session patterns with stable participant-level traits, including SBSOD and SA. We used regression modeling to systematically assess whether the distribution of an individual’s session types across the identified behavioral clusters could reliably predict their persistent spatial traits while controlling for demographic influences such as gender.

We further expanded our analysis across participants by using *IMI*, inspired by recent developments in tappigraphy research Ceolini and Ghosh, 2023; Ceolini et al., 2022), and *SIR* (see Table 3.2 for details). These metrics were specifically developed to encapsulate each participant’s general pattern of interaction over time regarding both frequency (how often participants initiate map interactions) and regularity (the consistency with which those interactions are spaced across time). Again, we used regression models to investigate the predictive relationship between these participant-level metrics and stable individual traits (sense of direction (SBSOD) and spatial anxiety (SA)), including gender as a covariate to disentangle trait-based differences from demographic effects.

In our **second study**, we extended our analytical framework by explicitly integrating the three contextual dimensions of daily mobile map use (behavioral, intrinsic, and extrinsic) into a unified analytical approach. Building upon the session-based interaction patterns identified in our prior analyses, specifically tap rate (interaction frequency) and CV-ITI (interaction regularity) as referenced in Table 3.2, we systematically combined these behavioral metrics with intrinsic individual traits (sense of direction [SBSOD], spatial anxiety [SA], and gender; see Table 3.3) and extrinsic urban form descriptors (see Table 3.4). The purpose of these analyses was to rigorously examine how individual traits and urban structure jointly shape real-world interactions with mobile map applications used during navigation.

To this end, we employed two primary **robust regression models (RLMs)** using Huber’s M estimator, a technique selected for its resilience to outliers and deviations from normality. The final regression formulas were as follows (all predictors were z-standardized):

Model 1 – Interaction Frequency

$$\begin{aligned} \text{tap_rate}_z = & \beta_0 + \beta_1 \cdot \text{orientation_entropy}_z + \beta_2 \cdot \text{betweenness_log}_z \\ & + \beta_3 \cdot \text{degree}_z + \beta_4 \cdot \text{circuitry_log}_z + \beta_5 \cdot \text{SA}_z \\ & + \beta_6 \cdot \text{SBSOD}_z + \beta_7 \cdot \text{Gender_M} + \varepsilon \end{aligned} \quad (3.1)$$

Model 2 – Interaction Regularity

$$\begin{aligned} \text{cv_iti}_z = & \beta_0 + \beta_1 \cdot \text{orientation_entropy}_z + \beta_2 \cdot \text{closeness_log}_z \\ & + \beta_3 \cdot \text{degree}_z + \beta_4 \cdot \text{circuitry_log}_z + \beta_5 \cdot \text{SA}_z \\ & + \beta_6 \cdot \text{SBSOD}_z + \beta_7 \cdot \text{Gender_M} + \varepsilon \end{aligned} \quad (3.2)$$

The rationale for selecting robust regression models was supported by the extensive preliminary analyses. Initially, we evaluated whether mixed-effects modeling was required due to repeated measures at the participant level. Intraclass correlation coefficients (ICCs)

for the tap rate and CV-ITI were modest ($ICC = 0.130$ and $ICC = 0.106$, respectively), suggesting limited within-subject clustering. Given the relatively low ICCs and our theoretical emphasis on capturing variability within individual navigation sessions across diverse urban contexts, we considered fixed-effects modeling appropriate. We contemplated using ordinary least squares (OLS) regression but ultimately excluded it due to critical violations of assumptions, notably heteroscedasticity and non-normal residual distributions. Furthermore, a multicollinearity assessment among urban form variables revealed high inter-correlations, particularly between proximity centrality and betweenness (Pearson $r > .87, p < .001$), and variance inflation factors ($VIF > 18$).

To mitigate these issues, we developed robust parallel regression models for each outcome, including either betweenness or closeness centrality separately. This strategy enabled a clear assessment of each centrality measure’s unique contributions without compromising statistical validity. Both models employed robust linear regression with Huber weighting, and model performance was evaluated using the median absolute deviation (MAD). Alternative model comparisons (betweenness versus closeness centrality) informed the final selection, which is presented in the Results section.

Given the geographic diversity of our sample, we conducted sensitivity analyses, excluding two participants residing outside Europe (one in the United States and one in Pakistan) whose GPS traces indicated distinct urban contexts that could affect generalizability. These analyses yielded stable results with slightly attenuated coefficients, confirming that the findings were not unduly influenced by geographic outliers. Detailed descriptions of model selection, sensitivity analyses, and considerations of participant locations are provided in the Appendix Section A.6.

To further explore how individual traits might moderate the relationship between urban morphology and navigation behavior, we conducted moderation analyses that incorporated interaction terms. Specifically, we tested whether sense of direction (SBSOD), spatial anxiety (SA), and gender moderated the effects of urban metrics (orientation entropy, closeness centrality, circuitry) on tappigraphy-based engagement outcomes. Each moderation model included standardized main effects, control variables, and interaction terms, estimated using robust linear regression with Huber’s M estimate. Significant interactions were further investigated using simple slope analysis, visualized at ± 1 standard deviation of the moderator. The complete model equations and additional details are presented in the Supplementary Materials A.8.

Having established this integrated analytical framework, which reveals how behavioral, intrinsic, and extrinsic contexts intertwine to shape real-world mobile map interactions, we moved toward a deeper and more predictive understanding. The following section details a computational model that harnesses these nuanced contextual relationships to predict future mobile map interactions, translating our theoretical insights into actionable predictions about user behavior.

3.3 Map App Interaction Data Prediction

To address our **third research question**, which asks how future user interactions with a mobile map application during navigation can be predicted based on the user’s characteristics, previous map usage, and spatial context, we adopted a modeling approach capable of

handling sequential, high-dimensional, and heterogeneous data, reflecting the real-world navigation complexity of mobile map applications: deep learning (DL) (see Figure 3.1).

DL is a family of machine learning algorithms based on multilayer neural networks that are well-suited for modeling complex, nonlinear relationships in large-scale data (Bishop & Bishop, 2023). The adoption of DL techniques has increased with the growth of big data (Najafabadi et al., 2015). The availability of large and complex datasets and advances in computational power have enabled researchers to work with high-dimensional, sequential, and heterogeneous information that often exceeds the capabilities of traditional statistical methods (Najafabadi et al., 2015).

DL methods have been widely used for modeling dynamic behavioral signals, spatio-temporal data, and various contextual characteristics (see Bishop and Bishop, 2023; Prince, 2023), making them particularly relevant for our research on mobile app usage and context-aware interaction.

Our predictive modeling task is formally framed as a binary classification problem as follows:

Given a user’s recent behavioral sequence of taps on a mobile device’s screen and their additional contextual information (intrinsic and extrinsic), our goal is to estimate the probability that their next tap within a short temporal window will be on a map application (as opposed to a non-map application):

$$\hat{p} = P(y = 1 \mid X_{\text{seq}}, X_{\text{ctx}})$$

where $X_{\text{seq}} \in \mathbb{R}^{T \times F}$ is the behavioral sequence (a matrix of T recent tap events, each with F features), $X_{\text{ctx}} \in \mathbb{R}^C$ is a vector of static or slowly varying contextual features (including user traits and urban form metrics), and $y \in \{0, 1\}$ is a binary label indicating whether the next tap is on a map app.

To capture both the temporal and contextual dimensions of the prediction task (that is, the next mobile map interaction), the model must represent both short- and long-range dependencies within user interaction sequences while integrating static or slowly changing contextual factors. As highlighted in the previous section 2.2.3, next-app prediction treats users’ interactions with mobile devices as structured sequences of actions (Katsarou et al., 2022), an appropriate approach for predicting the next-map-app use.

To this end, we used **long short-term memory (LSTM)** networks, a type of recurrent neural network (RNN) specifically designed to handle sequential data by maintaining an internal memory of past inputs via gated mechanisms (input, output, and forget gates) (Prince, 2023). These gates control how information is updated and preserved, enabling LSTMs to keep behaviorally relevant inputs over extended time spans while filtering out less useful signals, thereby mitigating the vanishing gradient problem that affects conventional RNNs. As a result, LSTMs are particularly well suited to modeling tap sequences, in which key behavioral regularities can unfold across multiple interactions. Note that we opted for a compact LSTM architecture instead of a Transformer. Transformers capture sequential dependencies using self-attention and positional encodings (Vaswani et al., 2017). In our case, the data consist of short sequences (roughly 10–50 taps), irregular event timings (explicitly represented by Δt), and only a limited number of labeled sequences for each participant. Under these circumstances, gated recurrent models are more sample-efficient,

can natively incorporate Δt as a standard input feature without dedicated temporal encoding schemes, and avoid the additional parameters and memory computational cost that self-attention would incur for sequences of this length.

LSTMs have also demonstrated strong performance in prior work on behavioral and mobile interaction prediction (e.g., (Greff et al., 2017; Mei et al., 2020; Ordóñez & Roggen, 2016)). However, while sequential modeling effectively captures user behavior over time, relying on temporal sequences alone ignores crucial contextual and individual influences. Factors such as spatial anxiety, sense of direction, gender, and urban form descriptors critically shape when, where, and how users engage with map applications.

To address these limitations, our model explicitly integrates temporal interaction patterns with contextual features using early feature fusion to produce a richer, more unified representation of the user’s state. Inspired by recent context-sensitive prediction frameworks, see (Wang et al., 2021), our approach synthesizes both the detailed temporal dynamics of user interaction sequences and the broader contextual landscape in which these interactions occur.

This combined methodological approach significantly improves the model’s predictive accuracy and interpretability, moving beyond simple behavioral sequencing to a nuanced understanding of engagement with mobile maps that is shaped by both personal and environmental contexts. Next, we describe the model architecture in detail.

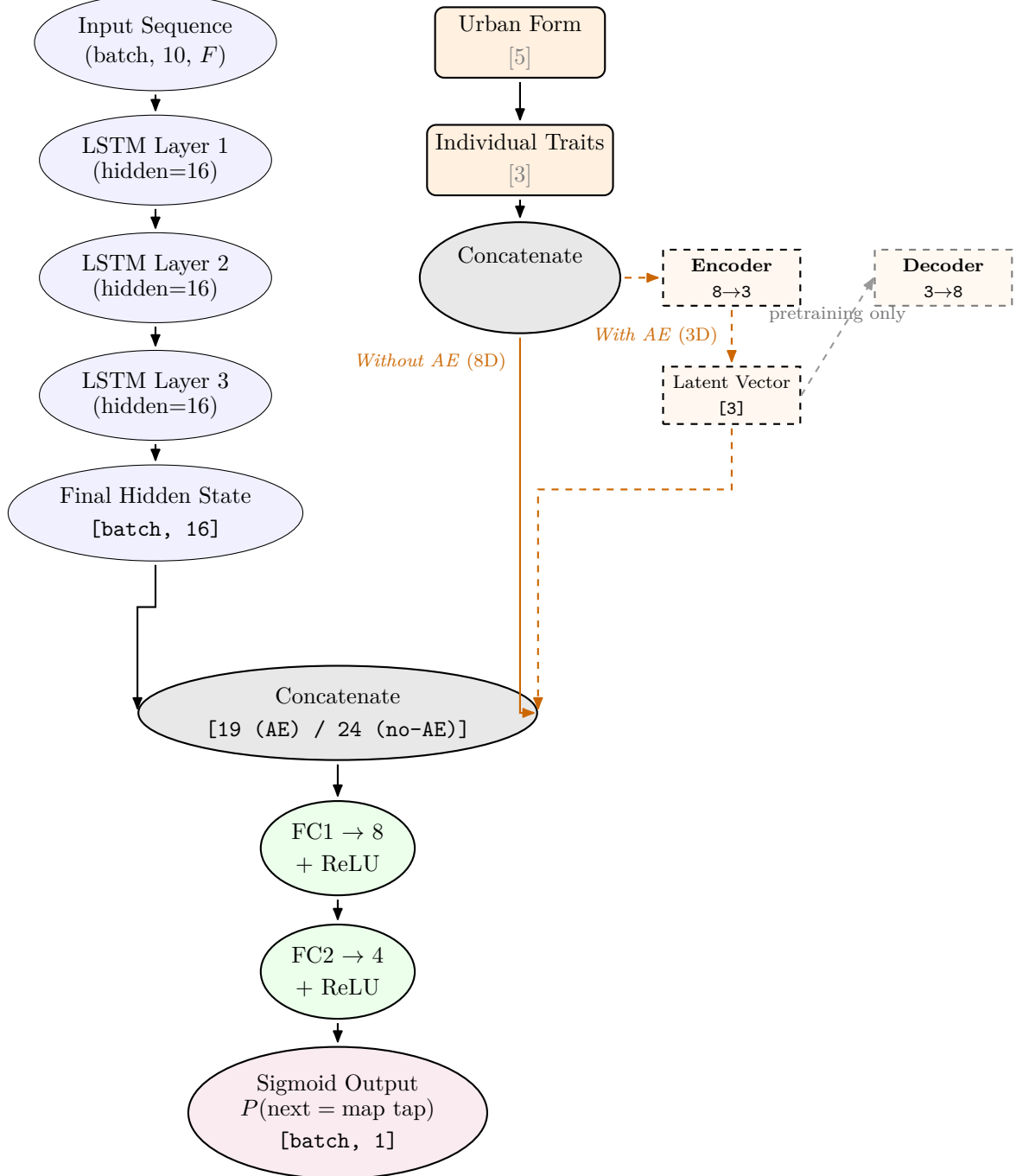
3.3.1 Model Architecture

Predicting interactions with the map app requires the integration of both recent user behavior patterns and the structure of the urban environment. We therefore developed a **two-branch, context-aware neural network** in PyTorch (see Figure 3.5), which consisted of the following:

- (a) **A Behavioral Interaction Branch:** a stack of three LSTM layers that encode the temporal structure of user–app interactions (blue boxes, top left).
- (b) **An Urban and Individual Trait Context Branch:** this encodes extrinsic and intrinsic factors using either the full eight-dimensional context vector, referred to as “**Without AE**”, or with a compressed three-dimensional latent embedding from an autoencoder referred to as “**With AE**” (orange boxes top right) .
- (c) **A Fusion Head:** this concatenates the LSTM hidden state with the context representation and maps the fused vector to a binary probability of MAP vs. NON-MAP (in the green/red boxes, at the bottom of the diagram).

In the following, we detail the data preparation, model design, and training procedures for each branch; reproducibility codes are available from the dedicated project [GitLab](#) repository.

Figure 3.5.: Architecture of the context-aware LSTM fusion model. The sequence branch (blue on the left) encodes recent app interaction patterns, while the context branch (orange on the right) encodes urban form and individual traits either directly (8D, Without AE) or via a latent 3D autoencoder (With AE). The fusion head (green and red) then integrates these two sources to make the final prediction.



Behavioral Interaction branch

Branch (a) uses a stack of **three LSTM layers** to model the temporal dynamics of users' mobile interaction patterns (hidden size = 16, dropout = 0.3). Ablation experiments were

conducted to determine the best hidden sizes, including 32 and 64 units, which are reported in the Appendix, Section A.9.

Data were drawn from the MoT II dataset. After pre-processing and category filtering, the dataset comprised 56,438 phone sessions. To ensure data quality, sessions with no interaction data or very sparse app category representation were excluded (see Appendix, Section A.9 for exclusion thresholds). Input sequences were created by aggregating tap events from each phone session into one-minute intervals, where each timestep corresponded to the number of taps in one of 27 app categories (input tensor shape $[batch, 10, 27]$).

Each session was labeled as map-related if it contained at least one tap in a navigation or travel app, and as non-map otherwise. Before inputting sequences into the LSTM, they were normalized feature-wise using global means and standard deviations. We also applied temporal differencing to highlight behavioral changes over time. Sequences were maintained at their natural length before windowing.

For training purposes, each behavioral sequence was divided into overlapping windows of a fixed duration of 10 minutes, resulting in 43,248 training windows (4392 positive; 10.16% positive), 5461 validation windows (9.98% positive), and 5290 test windows (9.34% positive). Each window inherited the **session-level label** (MAP vs. NON-MAP). The dataset was thus highly imbalanced, with only 10% of windows labeled as map-related.

To improve generalization and robustness to real-world noise, we systematically explored stochastic augmentation strategies applied to the training set alone; the validation and test windows remained unaltered. The search space comprised three types of augmentation:

- *Time-step masking*, in which tap counts for randomly selected time steps were replaced with zeros to simulate missing activity ($mask_prob \in \{0.05, 0.10, 0.15, 0.20\}$);
- *Gaussian noise injection*, where each feature value was perturbed by zero-mean Gaussian noise with standard deviation $\sigma \in \{0.05, 0.10, 0.15, 0.20, 0.30\}$;
- *Feature dropout*, which randomly removed entire app-category channels within a window ($feat_drop_prob \in \{0.00, 0.05, 0.10, 0.15\}$).

The complete factorial design ($4 \times 5 \times 4$) yielded 80 unique configurations, each trained and evaluated under the same optimization regime. The configuration that achieved the highest F1 validation score was designated as the best model with $mask_prob = 0.20$, $\sigma = 0.15$ and $feat_drop_prob = 0.00$. All test set metrics reported for the augmented model correspond to this configuration. For more information, refer to the project’s [GitLab](#) repository.

We applied two complementary strategies to address strong class imbalance. First, we used weighted random sampling in training batches using *WeightedRandomSampler*. Each window is assigned a sampling weight that is inversely proportional to its class frequency, ensuring that underrepresented samples (those related to maps) are selected more frequently (balanced mix of map and non-map windows with an average positive ratio = 50.06% in each mini-batch) without generating synthetic data during each epoch, mitigating bias toward the majority class and supporting reliable prediction of map app usage. Second, we used a class-weighted loss by setting the *pos_weight* parameter in the binary cross-entropy with logits loss set to $\frac{n_{neg}}{n_{pos}}$, which was computed once on the training set. This technique more heavily penalized misclassified map sessions while preserving the correct calibration in the validation and test sets. Early stopping was applied with a patience of 20 epochs, monitoring the validation loss. Please refer to Appendix A.9 for a complete description of

the behavioral data pipeline, including session filtering, tensor construction, normalization, differencing, window extraction, augmentation methods, and weighted sampling, along with all code, parameter settings, and reproducibility procedures.

Urban and Individual Trait Context Branch

Branch (b) encodes the interaction-related extrinsic and intrinsic context. We constructed a context feature vector for every session by **combining** five metrics of **urban form** (orientation entropy, betweenness, closeness, circuitry, and degree) with three **participant traits** (spatial anxiety, sense of direction, and gender, encoded as a single binary indicator). This resulted in a compact **eight-dimensional vector** that captured both the structural complexity of the user’s environment and key individual differences known to shape navigation behavior.

All features were standardized (mean = 0, SD = 1) to ensure comparability and prevent the dominance of high-magnitude variables. To ensure that the contextual inputs corresponded exactly to the interaction sequences in the LSTM branch, both datasets were aligned using the same session identifiers and training, development, and test partitioning. Only sessions that were present in both datasets were retained ($N = 35,089$ phone sessions).

The resulting aligned datasets were stored as tuples and paired with their corresponding LSTM session tensors to produce fusion-ready inputs. Then, two alternative approaches were considered to integrate this context vector into the predictive model:

With AE (autoencoder): the eight-dimensional context vector was compressed into a **three-dimensional embedding** using a dedicated **autoencoder**. The autoencoder was trained to minimize information loss, thereby summarizing environmental and individual factors while reducing the dimensionality and potential redundancy of the characteristics. Only encoder weights were retained, allowing each session context vector to be assigned to its compact latent representation during training and inference.

Without AE (direct input): The eight-dimensional raw context vector was fed directly into the fusion head without compression. This provided a baseline for comparison to assess whether reducing dimensionality improved predictive performance.

In both cases, the resulting context representation (either latent or direct) was concatenated with the behavioral sequence embedded in the fusion stage of the model, allowing predictions to reflect both recent interaction dynamics and stable environmental and trait-based influences. Details on feature selection, context vector construction, autoencoder architecture, and training for this part of the model are provided in the Appendix A.9, along with the implementation.

Fusion Head

In the final stage (c), the last hidden state from the behavioral LSTM (16 units) was concatenated with the context representation. Two variants of the context representation were tested: the 3D latent vector from the autoencoder (With AE) and the raw 8-dimensional context vector (Without AE). This resulted in a fused vector of 19 or 24 dimensions, respectively, which was then passed to the fusion head. The fused vector was

passed through a **multi-layer perceptron** (MLP) comprising two fully connected layers (FC1 = 8 units, FC2 = 4 units, both with rectified linear unit (ReLU) activations), followed by a final sigmoid output neuron that predicted the probability of the next tap being on a map app. During training, both the behavioral LSTM and the context autoencoder were kept frozen to ensure stable feature extraction and prevent overfitting. Only the fusion head parameters were updated, optimized with binary cross-entropy loss (with logits). Training used the Adam optimizer (learning rate = 10^{-3} , weight decay = 10^{-4}) with a batch size of 32. Early stopping was applied with a patience of 20 epochs, monitored for validation loss. Class imbalance was addressed as described for the behavioral branch, using weighted random sampling and class-weighted loss. Appendix A.9 documents all architectural details about this final fusion stage.

After training the model, the next step was to examine why it made particular predictions. The predictive performance alone did not reveal how the contextual information, such as urban form metrics or individual traits, influenced the model’s output. To achieve this insight, we conducted an explainability analysis that focused on the fusion head, integrating behavioral and contextual representations in order to classify whether the next tap would occur in a map app.

3.3.2 Explainability Analysis of Fusion Head Predictions in Everyday Mobile Map Use

We used SHAP (SHapley Additive Explanations) to understand the influence of contextual features on the fusion head’s predictions (Lundberg & Lee, 2017). SHAP is a method that assigns a value to each input feature, illustrating how much it pushes the model toward or away from predicting a “map tap.” Positive values mean the feature made a “map tap” prediction more likely, while negative values mean the opposite. We chose SHAP because it is well-suited for complex models (Molnar, 2022) such as the one used in this thesis.

The analysis was applied to both configurations of the fusion head, the **With AE** variant and the **Without AE** variant. In both cases, the LSTM branch that learned from tap sequences was frozen and its 16-dimensional sequence representation was combined with the contextual input. SHAP values were only calculated for the contextual features so that their role could be studied independently from that of the sequence features. Explanations were produced using the development dataset.

Clear feature names made the explanations interpretable. For the raw model, features included betweenness, closeness, degree, orientation entropy, circuitry, SBSOD score, SA score, and gender. For the latent model, the features were labeled AE latent 1, AE latent 2, and AE latent 3. We averaged the absolute SHAP values to rank the most influential features. We also repeated the analysis separately for map windows and non-map windows, which produced three sets of outputs for each configuration: an overall summary, a map-only summary, and a non-map summary. We visualized the results with graphs that depicted both the strength and the direction of each feature’s influence and with bar graphs that compared positive and negative contributions.

An ablation study tested the significance of groups of features for prediction. We re-trained the fusion head three times: once with all features included, once including only urban features, and once with only individual traits. This allowed us to see how the performance changed with the removal of certain groups of features. We evaluated each version in terms

of accuracy, F1 score, and ROC-AUC, as well as precision, recall, and F1 score for the map and non-map classes separately, all at a threshold of 0.50.

All analyses were run in Python using the SHAP library (version 0.48.0). We wrote custom routines to prepare the LSTM outputs, join them with context features, compute the SHAP values, and generate plots and tables. Combined, these analyses not only clarify the accuracy of the model's predictions, but also why it arrives at its decisions.

In summary, this methodology establishes a solid pipeline for understanding and modeling how individuals interact with mobile maps in everyday life by combining unobtrusive data collection (RQ1), systematic feature integration (RQ2), and predictive modeling with explainability across the three contextual levels, intrinsic, behavior, and extrinsic (RQ3). The following chapter presents the results derived from these methods and demonstrates how they address the three research questions.

4

CHAP-
TER

Results

*“Man, I cried, how ignorant art
thou in thy pride of wisdom.”
Frankenstein, Mary Shelley*

This chapter presents the empirical findings derived from the methodological framework outlined in Section 3. The results are organized sequentially according to the three research questions (RQ1–RQ3, see Section 1.4).

The first set of results, which address RQ1, focuses on unobtrusive and parsimonious data collection. We report descriptive statistics from the two MapOnTap campaigns (MoT I and MoT II; see Section 3.1.2), covering tappigraphy and GPS data in MoT I, and the same data in MoT II, complemented by the questionnaire measures). The second set of results, addressing RQ2, focuses on the data integration pipeline. We first report exploratory analyses on both MoT I and MoT II datasets, which delve into the basic metrics derived from tappigraphy, GPS data, and the questionnaires. These analyses evaluated the overall quality of the data and identified preliminary patterns of mobile map use, guiding our choice of analytical strategies and informing our subsequent integrated modeling efforts. We then report the results of core analyses designed to clarify and render more interpretable the mobile interaction patterns observed throughout all phone sessions and across participants. Our analyses aimed to reveal how behavioral, intrinsic, and extrinsic contexts collectively influence mobile map interactions. The third set of results addresses RQ3, presenting the findings of the context-aware LSTM fusion model. We report the model’s accuracy and performance per class, compare variants using either the full eight-dimensional context vector or its latent 3D autoencoder, and analyze error patterns using confusion matrices. To elucidate the “black box” of the model, we complement these performance metrics with SHAP-based explainability analyses that highlight which contextual features most strongly influence predictions.

Key findings are summarized in tables and figures throughout this chapter. Extended descriptive statistics and robustness checks are provided in the Appendix A.4, while

complete app inventories, preprocessing scripts, and supplementary plots are available in the project’s [GitLab](#) repository. The following subsections address each research question in turn, beginning with descriptive results from MoT I and MoT II (RQ1).

4.1 Descriptive Results of Mobile Map App Interaction Data Collection

As reported in Section 2.2.1, tappigraphy is a highly parsimonious approach to collecting touch events from the foreground application while an individual is actively using a phone. It has been widely adopted in neuroscience research to aid in understanding human behavioral rhythms (Balerna & Ghosh, 2018; Ceolini et al., 2022). However, when applied to research on the use of mobile map applications, this informational parsimony does not produce further insights into the users themselves and provides only limited experimental control (Reichenbacher et al., 2022). To address this, we designed two dedicated data collection methods to test the limits of the parsimony principle by examining how little information is sufficient to retain meaningful insights into everyday mobile map interaction. This was achieved by gradually collecting additional contextual information. MoT I was added to the voluntary GPS logging of tappigraphy data, providing an initial baseline. MoT II collected the same data, as well as information from demographic and self-report measures, enriching the dataset with participant traits. The following descriptive statistics summarize the scope and composition of the data obtained from both campaigns.

4.1.1 MoT I Descriptive Overview

In our first baseline data collection, MoT, I included *53 participants*. In keeping with its intentionally minimal design, no demographic information was collected. Therefore, the analyses focused exclusively on phone session data retrieved from tappigraphy and voluntary GPS logging.

Participants were asked to contribute to data collection for at least two consecutive weeks, but could continue for longer if they wished to. The tappigraphy coverage per participant spanned from a minimum of *1 hr 57 min*, (0.08 days) to a maximum of *868 days* ($M = 94$ days, $SD = 139$). See Table 4.1.

We collected GPS data for 42 participants. The recordings were voluntary and participants could enable or disable GPS logging at any time. The coverage ranged from a minimum of 0.12 days (2 h 56 min) to a maximum of 384.67 days, with a mean of 102.98 days ($SD = 119.28$ days). See Table 4.1. The complete summaries of the raw data and the complete list of applications are available in the project repository [GitLab](#).

Table 4.1.: *Participant Coverage in the MoT I Campaign*

Metric	Part.	Min	Max	Mdn	M	SD
Tappigraphy coverage (days)	53	0.08	868.00	39.00	94.00	139.00
GPS coverage (days)	42	0.12	384.67	47.18	102.98	119.28

Note. Values are given in days. The tappigraphy minimum of 0.08 days corresponds to 1 hour 57 minutes; the GPS minimum of 0.12 days corresponds to 2 hours 56 minutes.

Only participants with data available for both tappigraphy and GPS were included in the analyses, resulting in a sample of 38 participants (71.7% of those with data from either source). The descriptive statistics were calculated according to phone sessions and participants.

Phone Session Characteristics

Overall, we recorded 119,713 phone sessions, comprising a total of 12,492,705 taps. The number of taps per session ranged from 1 to 10,369 ($M = 104.36$, $SD = 254.14$; $Mdn = 27.00$). The duration of the sessions ranged from 0.00 to 28,942.21 seconds ($M = 252.15$ s, $SD = 638.78$ s; $Mdn = 49.10$ s). Map-only sessions accounted for 2.97% of all phone sessions and were substantially shorter, ranging from 0.00 to 2,897.66 seconds ($M = 33.05$ s, $SD = 110.81$ s; $Mdn = 6.91$ s). See Table 4.2 for a summary.

Table 4.2.: *MoT I Tappigraphy Statistics by Phone Session*

Metric	Min	Max	Mdn	M	SD
Taps per session	1.00	10,369.00	27.00	104.36	254.14
Session duration (s)	0.00	28,942.21	49.10	252.15	638.78
Map-only duration (s)	0.00	2,897.66	6.91	33.05	110.81

Note. Map-only sessions constituted 2.97% of all sessions.

When aggregating by participants, the number of phone sessions per participant ranged from 4 to 20,965 ($M = 3069.56$, $SD = 4510.05$; $Mdn = 1,204$). Map-only sessions ranged from 1 to 1,269 per participant ($M = 118.47$, $SD = 251.10$; $Mdn = 16$), representing between 0.11% and 24.14% of all sessions ($M = 2.68\%$, $SD = 4.60\%$; $Mdn = 1.41\%$). The proportion of taps occurring in map applications ranged from 0.05% to 68.53% ($M = 3.66\%$, $SD = 11.27\%$), while the proportion of total session time spent in map applications ranged from 0.27% to 48.16% ($M = 6.03\%$, $SD = 8.60\%$). See Table 4.3 for more details.

Table 4.3.: *MoT I Tappigraphy Statistics Aggregated by Participants*

Metric	Part.	Min	Max	Mdn	M	SD
Sessions per participant	38	4.00	20,965.00	1,204.00	3,069.56	4,510.05
Map-only sessions per participant	38	1.00	1,269.00	16.00	118.47	251.10
Map-only share of sessions (%)	38	0.11	24.14	1.41	2.68	4.60
Map-app taps (share, %)	38	0.05	68.53	1.41	3.66	11.27
Time in map apps (% of session)	38	0.27	48.16	3.40	6.03	8.60

Note. The data refer to the 38 participants who were retained. Participants exhibited significant variability in their use of mobile maps. The number of sessions ranged from four to over 20,000, with a median of approximately 1200. Map-only sessions represented a small proportion of the total number of sessions ($M = 2.68\%$) across participants, and direct interaction with map apps was modest compared to general phone use ($M = 3.66\%$ of taps and $M = 6.03\%$ of session time).

Applications Used and Categories Distribution

Tappigraphy captures every interaction with the smartphone's touchscreen, along with the foreground application being used and its associated Google Play category (see Section 3.2.1).

In our cohort, participants used 1087 unique apps spanning 32 Google Play categories, as well as one uncategorized label ("-1"). A complete list of all recorded applications and their associated categories is provided in the project's [GitLab](#) repository.

For this thesis, we focused specifically on applications categorized as Maps and Navigation (MN) and Travel and Local (TL), which provide the most consistent and comprehensive approximation of map-related usage available within the Play Store classification scheme. This broader grouping enabled us to capture a wide spectrum of interactions with map services, ranging from real-time GPS navigation to travel-related searches and local exploration. Importantly, although the Play Store classified the custom-built MapOnTap application as an MN app (a standard assignment during application registration), all tap interactions generated within MapOnTap itself were excluded from the analysis to avoid methodological artifacts.

The MN category included 25 different apps, used by 33 of 38 participants (86.84%), while the TL category comprised 63 different apps, used by 36 of the participants (94.74%).

The 10 most commonly used navigation-relevant apps are shown in Figure 4.1. These were *Google Maps* (TL; $n = 35$ participants), *SBB Mobile*, which is the Swiss Federal Railways app (MN; $n = 28$ participants), *Swisstopo*, the Swiss national mapping app (MN; $n = 8$ participants), *Booking.com: Hotels Travel* (TL; $n = 7$), *Flightradar24 Flight Tracker* (TL; $n = 7$), *PubliBike*, a Swiss bike-sharing platform (TL; $n = 7$ participants), *ZVV Fahrplan*, Zurich's public transport app (TL; $n = 5$), *Airbnb* (TL; $n = 5$), *local.ch: booking platform*, a Swiss local business and service finder used for travel-related reservations (TL; $n = 4$), and *Uber* (MN; $n = 4$).

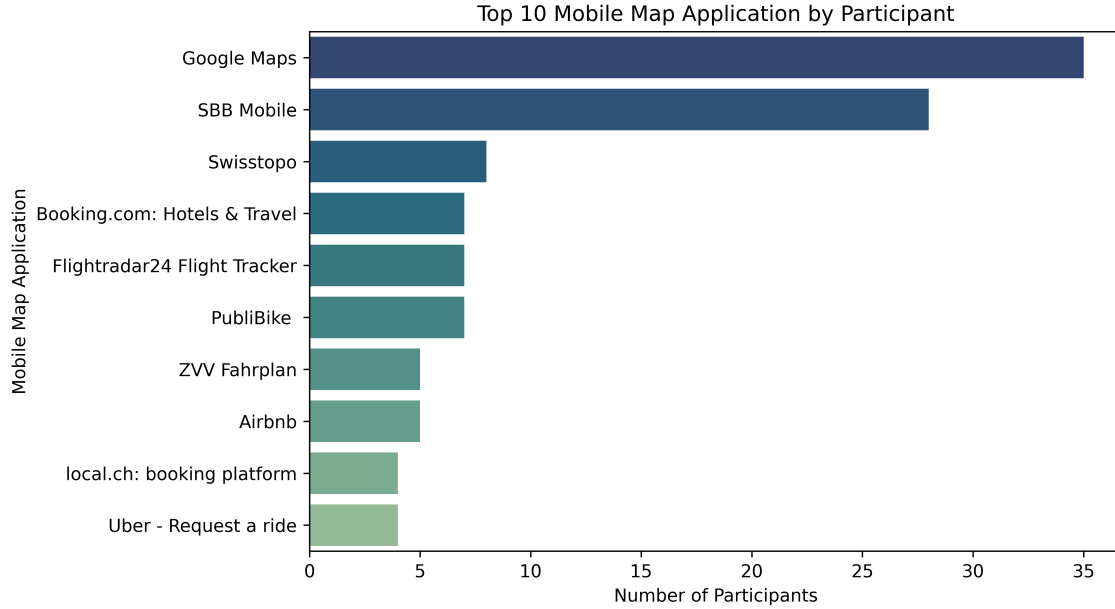
For participants, the MN taps ranged from 17.00 to 49,384.79 ($M = 4803.09$, $SD = 8939.27$), and the TL taps ranged from 251.00 to 38,130.87 ($M = 9472.71$, $SD = 9175.03$); see Table 4.4.

Table 4.4.: *Within-Participant Mean Taps per Session Across Map-Related App Categories. Participants engaged in more intensive interaction with Travel & Local (TL) apps than with Maps & Navigation apps (MN), as reflected in the higher median and mean values.*

Category	Min	Max	Mdn	M	SD
Maps & Navigation (MN)	17.00	49,384.79	1,946.54	4,803.09	8,939.27
Travel & Local (TL)	251.00	38,130.87	6,879.67	9,472.71	9,175.03

Note. Values are each participant's average number of taps per session, calculated only for those with at least one session in the respective category. Higher values indicate greater interaction per session.

Figure 4.1.: *Top 10 Most Frequently Used Map-Related Apps Across Participants. Google Maps was by far the most widely used app, whereas others were adopted by comparatively fewer participants. The bars indicate the number of participants who used each app at least once during the study period.*



GPS Sessions Characteristics

Having provided an overview of application categories and the specific use of mobile map applications, we next describe the GPS dataset that formed the basis for subsequent spatial analyses.

Participants contributed between 1 and 10,015 GPS sessions ($M = 1310$, $SD = 2194$; $Mdn = 329$). The number of GPS points per participant ranged from 20 to 416,998 ($M = 76,780.07$, $SD = 112,655.28$; $Mdn = 28,009.5$). The average GPS session duration was 104.3 s ($SD = 57.7$ s; $Mdn = 98.4$ s; $IQR = 86.2$ – 111.2 s), with a minimum of 15 s and a maximum of 10,015 s (see Table 4.5).¹ In total, 2,978,096 GPS points were recorded. Complete summary statistics are available in the project repository [GitLab](#).

GPS statistics show that the dataset contains a large quantity of diverse spatio-temporal traces, while also highlighting significant heterogeneity, such as non-uniform distributions of sessions and GPS points among participants. This variability is typical of real-world recordings, especially voluntary ones. It is important from a methodological perspective as it highlights the need to consider these differences cautiously when analyzing map use over spatial distributions.

¹Two negative session durations (min = -21,471 s) were identified and removed during quality control.

Table 4.5.: *GPS Usage Statistics Across Participants. Participants exhibited highly variable GPS engagement, with some recording only a handful of sessions and others exceeding 10,000. On average, sessions were short in duration, but the total number of GPS points per participant was substantial.*

Metric	Min	Max	Mdn	M	SD
GPS sessions per participant	1.00	10,015.00	329.00	1,310.00	2,194.00
GPS points per participant	20.00	416,998.00	28,009.50	76,780.07	112,655.28
GPS session length (s)	15.00	10,015.00	98.40	104.30	57.70

Note. Statistics are based on the filtered *MoT I* cohort.

4.1.2 MoT II: Descriptive Overview

Our second data collection campaign, MoT II, expanded on the principle of parsimony guiding our data collection design by incorporating additional contextual information beyond tappigraphy and GPS. In addition to the unobtrusive recording of touch events in phone sessions and voluntary location traces, MoT II introduced standardized measures of individual traits, such as SBSOD (Hegarty et al., 2006) and SA (He and Hegarty, 2020), as well as demographic information including age group and gender. These intrinsic measures complemented the tappigraphy and GPS data, facilitating the examination of how stable individual differences interact with everyday mobile map use, an aspect that could not be addressed in MoT I.

The enrolled cohort included 61 participants (32 male and 29 female) between 18 and 64 years of age. The largest age group was 25–34 years old ($n = 39$ participants). The coverage of tappigraphy per participant ranged from 0.01 to 33.62 days ($M = 17.35$, $SD = 5.48$; $Mdn = 16.56$). GPS coverage ranged from 0.01 to 96.28 days ($M = 18.57$, $SD = 12.76$; $Mdn = 15.89$).

Table 4.6.: *Participant Coverage in the MoT II Campaign. Although participants typically contributed about 16–18 days of tappigraphy and GPS data, coverage varied substantially across the sample, with some providing only minimal data and others extending their data contribution to several months.*

Metric	Part.	Min	Max	Mdn	M	SD
Tappigraphy coverage (days)	61	0.01	33.62	16.56	17.35	5.48
GPS coverage (days)	61	0.01	96.28	15.89	18.57	12.76

Note. Coverage reflects the number of days each participant contributed valid tappigraphy or GPS data.

Importantly, unlike in MoT I, where all phone sessions were retained for analysis, MoT II applied a stricter filtering criterion. For this dataset, we focused exclusively on navigation-related phone sessions, as detailed in Section 3.2.1. Therefore, the descriptive statistics we report refer only to the subset of sessions and their corresponding GPS traces retained following this filtering step. Descriptive analyses of the raw data are provided in the Appendix, Section A.4.

Applications Used and Categories Distribution

The overall MoT II dataset contained 1029 unique applications in 33 categories (including the uncategorized “1”). However, for MoT II, the analysis was restricted to mobile map applications that were directly relevant to navigation tasks. This ensured that the analysis focused on apps primarily used for spatial decision making, wayfinding, or mobility support and excluded sessions with peripheral app use unrelated to navigation behavior. We therefore filtered phone sessions by MN and TL categories that involved a set of only 26 applications and excluded all MapOnTap self-interactions (Methods, Section 3.2.1). Importantly, sessions may include concurrent activity from other app categories (e.g., communication, social media) and, in some cases, additional MN/TL apps beyond the focal 26. Thus, although the analyses are designed to focus on navigation-specific use, they also capture the broader multitasking context in which navigation typically occurs. The implications of this interplay between applications are addressed in the Discussion section. After this filtering step, the session had a total of 364 co-occurring applications from the other categories. Among these co-occurring apps, there were 30 additional to the 26 selected MN and TL apps.

Participants used between a minimum of one and a maximum of five MN apps ($M = 1.89$, $Mdn = 1$) and between a min of 1 and a max of 9 TL apps ($M = 1.82$, $Mdn = 1$). Across the total sessions 4,499, the MN taps represented 1.76% of all the taps and the TL taps for 6.36%.

The 10 most commonly used navigation-relevant apps are presented in Figure 4.2. *Google Maps* (TL) was by far the most common, used by 49 participants, followed by *SBB Mobile* (MN; $n = 31$). Other high-frequency applications included the German Rail Journey Planner *DB Navigator* (TL; $n = 6$) and the Swiss Leisure Routes Platform *Schweizmobil* (TL; $n = 6$), which offers planning for hiking, cycling, and public transport routes. Public transportation and reservation apps also appeared, such as *Swisstopo* (MN; $n = 4$), the official Swiss national map service; *ZVV Fahrplan* (MN; $n = 4$), Zurich’s public transport schedule and route planner; and *Booking.com* (TL; $n = 4$), an international hotel and travel booking platform. Long-distance bus service *FlixBus* (TL; $n = 3$), Swiss public transport ticketing app *FAIRTIQ* (MN; $n = 3$), and the Italian train reservation app *ProntoTreno* (MN; $n = 3$) were less frequent but still within the top 10.

Having identified navigation apps in the MoT II navigation-filtered dataset, we examined how they were used by describing the characteristics of the phone sessions.

Phone Session Characteristics

The final MoT-II cohort included 56 participants (25 women, 31 men) and 4499 phone sessions strictly related to navigation apps. The duration of the sessions was highly variable, ranging from 0.04 to 14,710.02s, ≈ 4 hours ($M = 243.48$ s, $SD = 723.87$ s; $Mdn = 51.24$ s), indicating that most of the sessions were relatively short, while only a few extended to several hours. Only a small subset of sessions consisted exclusively of map use (“map-only” sessions), which were markedly shorter (0.04–1,544.67 s; ≈ 25.7 minutes; $M = 30.20$ s, $SD = 97.17$ s; $Mdn = 11.52$ s) and represented 19.49% of all navigation sessions.

When aggregating sessions per participant, the number of sessions per participant ranged from a minimum of two to a maximum of 596 ($M = 80.34$, $SD = 112.61$; $Mdn = 44.00$). The map-only sessions ranged from one to 243 ($M = 18.33$, $SD = 37.36$; $Mdn = 8.00$),

Figure 4.2.: *Top 10 Most Frequently Used Navigation-Relevant Apps in the MoT II Dataset. Counts represent the number of participants using each app at least once during the study period. Google Maps dominated usage, while apps such as DB Navigator (official Deutsche Bahn rail planner) and FAIRTIQ (Swiss public transport ticketing) illustrate country-specific adoption.*

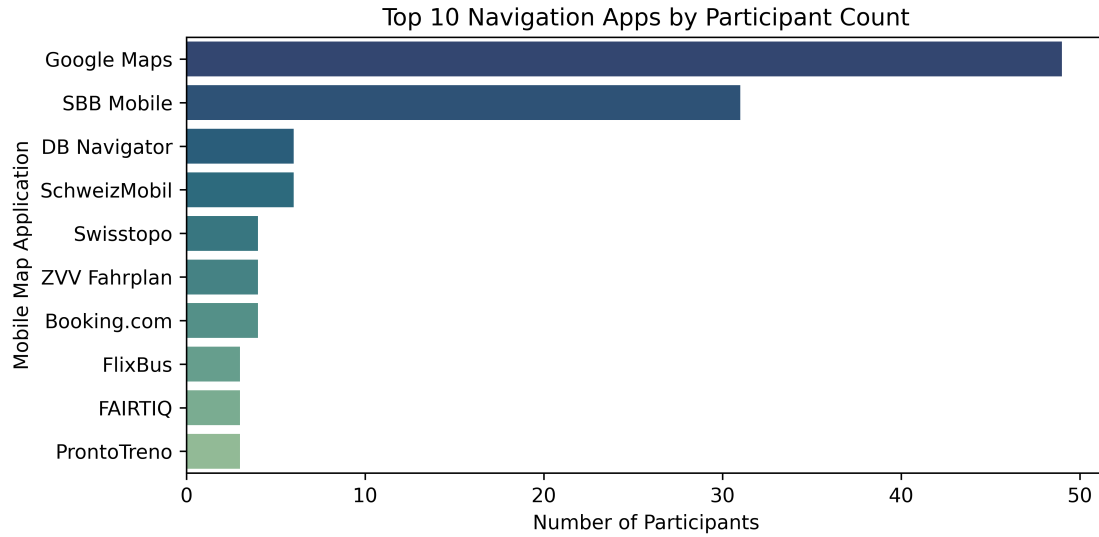


Table 4.7.: *Tappigraphy Statistics at the Session Level in the MoT II Navigation-Filtered Cohort. Navigation sessions were generally short (see Table 4.7, with a median duration of approximately 51 seconds (less than 1 minute), while the longest sessions extended to over 4 hours. Map-only sessions were much shorter (median ≈ 12 seconds, maximum ≈ 26 minutes) and accounted for 19.49% of all navigation sessions.*

Metric	Min	Max	Mdn	M	SD
Taps per session	2.00	5,482.00	73.00	186.74	314.75
Session duration (s)	0.04	14,710.02	51.24	243.48	723.87
Map-only duration (s)	0.04	1,544.67	11.52	30.20	97.17

Note. Values are reported per session. “Map-only” indicates sessions where all recorded interaction occurred within a navigation app.

representing 1.47 to 48. 50% of each participant’s sessions ($M = 18.13\%$, $SD = 10.60\%$; $Mdn = 15.92\%$). The proportion of taps in the map applications ranged from 1.46% to 76.28% ($M = 14.76\%$, $SD = 14.84\%$; $Mdn = 9.79\%$). The proportion of total session time spent on map apps ranged from 0.00% to 9.33% ($M = 2.61\%$, $SD = 2.49\%$; $Mdn = 1.77\%$). See Table 4.8.

GPS Session Characteristics

Across the MoT II navigation-filtered cohort, participants contributed a highly variable number of GPS sessions (see Table A.5). The number of sessions per participant ranged from two to 2321 ($M = 727.42$, $SD = 541.15$; $Mdn = 627.00$). The total number of GPS

Table 4.8.: *Tappigraphy Statistics Aggregated by Participant in the MoT II Navigation-Filtered Cohort. Participants varied widely in their use of map apps: while some engaged with only a few sessions, others accumulated hundreds of them. Map-only sessions constituted a notable but highly variable share of participants' overall activity.*

Metric	Min	Max	Mdn	M	SD
Sessions per participant	2.00	596.00	44.00	80.34	112.61
Map-only sessions per participant	1.00	243.00	8.00	18.33	37.36
Map-only share of sessions (%)	1.47	48.50	15.92	18.13	10.60
Map-app taps (share, %)	1.46	76.28	9.79	14.76	14.84
Time in map apps (% of session time)	0.00	9.33	1.77	2.61	2.49

Note. Values are aggregated at the participant level. Percentages represent the proportion of sessions, taps, or session time attributable to navigation apps within each participant's overall activity.

points recorded also varied considerably, from 54 to 258,594 per participant ($M = 73,583.42$, $SD = 61,001.77$; $Mdn = 68,789.50$).

GPS recordings were generally brief at the session level. The median session lasted only 45 s ($M = 224.41$, $SD = 619.95$; both in seconds, that is, $M \approx 3.7$ minutes, $SD \approx 10.3$ minutes), although some extended to more than 6 h (23,422 s). Observation windows, the total duration of data collection per participant, ranged from less than one day (0.01 days) to almost three months (80.03 days). On average, the participants contributed approximately 17 days of GPS data ($SD = 11.51$; $Mdn = 15.59$).

Table 4.9.: *GPS Statistics in the MoT II Navigation-Filtered Cohort. Participants exhibited highly variable GPS engagement, with some recording only a few sessions and others exceeding 2000. On average, sessions were short in duration, but the total number of GPS points per participant was substantial.*

Metric	Min	Max	Mdn	M	SD
GPS sessions per participant	2.00	2,321.00	627.00	727.42	541.15
GPS points per participant	54.00	258,594.00	68,789.50	73,583.42	61,001.77
GPS session length (s)	0.00	23,422.17	45.03	224.41	619.95

Statistics are based on the filtered MoT II cohort. Observation windows spanned from less than one day to nearly three months (0.01–80.03 days).

Sense of Direction and Spatial Anxiety Statistics

In MoT II, we extended the data collection beyond tappigraphy and GPS logging by including two standardized self-report measures: the Santa Barbara Sense of Direction scale (SBSOD, Hegarty et al., 2006) and the Spatial Anxiety scale (SA, He and Hegarty, 2020). These instruments were added to capture stable individual differences in spatial ability and emotional responses to navigation tasks, providing a basis for linking trait variability to real-world engagement with mobile map applications. The descriptive statistics are shown in 4.10

Among the 56 participants, the SBSOD scores ranged from 2.13 to 6.73 (on a 7-point scale), reflecting substantial variability in self-reported sense of direction ($M = 4.93$; $SD = 0.97$; $Mdn = 5.04$). This distribution indicates that the sample included participants with both a very low and very high self-reported sense of direction, although most clustered around moderately high scores. SA scores ranged from 1.23 to 3.92, with an average of $M = 2.35$ ($SD = 0.63$; $Mdn = 2.46$). Thus, most participants reported moderate levels of spatial anxiety, although some scored near the extreme ends of the scale.

Table 4.10.: *Descriptive Statistics for Individual Difference Measures in the MoT II Cohort. Sense of direction scores were moderately high on average, while spatial anxiety scores indicated generally moderate levels across participants.*

Measure	Min	Max	Mdn	M	SD
Santa Barbara Sense of Direction (SBSOD)	2.13	6.73	5.04	4.93	0.97
Spatial Anxiety (SA)	1.23	3.92	2.46	2.35	0.63

Note. $n = 56$. SBSOD = Santa Barbara Sense of Direction scale; SA = Spatial Anxiety scale. Min = minimum, Mdn = median, M = mean, SD = standard deviation.

Finally, as noted in Section 4.1.2, MoT II also recorded demographic information alongside SBSOD and SA self-reports. Of the 56 participants retained after preprocessing, 25 were women and 31 men. Participants were unevenly distributed between age groups: 10 were aged between 18 and 24 years, 38 between 25 and 34 years, five between 35 and 44 years, two between 45 and 54 years, and one between 55 and 64 years. Due to this imbalance, especially the strong concentration in the 25 and 34 group, age was not included in the analysis (see Section 3.2.1). The descriptive results reported here therefore focus on the SBSOD and SA measures in combination with reported gender.

For SBSOD, women reported lower self-perceived sense of direction scores ($M = 4.32$, $SD = 0.94$), ranging from a minimum of 3.07 to a maximum of 6.20, compared to men ($M = 5.32$, $SD = 0.83$), whose scores ranged from 2.13 to 6.73. In contrast, women reported higher spatial anxiety scores ($M = 2.50$, $SD = 0.40$, range = 1.54–3.62) than men ($M = 1.99$, $SD = 0.57$, range = 1.23–3.92). Together, these results suggest there are gender differences in both self-reported spatial abilities and spatial anxiety (see Table 4.11).

Table 4.11.: *Descriptive Statistics for SBSOD and Spatial Anxiety (SA) Scores by Gender. Women reported a poorer sense of direction and higher spatial anxiety than men.*

Measure	Gender	Min	Max	M	SD
SBSOD (7-point scale)	Women	3.07	6.20	4.32	0.94
	Men	2.13	6.73	5.32	0.83
SA (5-point scale)	Women	1.54	3.62	2.50	0.40
	Men	1.23	3.92	1.99	0.57

Note. SBSOD = Santa Barbara Sense of Direction scale (Hegarty et al., 2006). SA = Spatial Anxiety scale (He & Hegarty, 2020). Higher SBSOD scores indicate stronger self-reported sense of direction; higher SA scores indicate greater spatial anxiety.

The two MoT campaigns explored how parsimonious data collection could be maintained while gradually adding the contextual elements necessary to understand daily mobile map

use. The first campaign, MoT I, established a minimal baseline by relying exclusively on tappigraphy and voluntary GPS data logging. This setup captured the overall scale and heterogeneity of smartphone use, but illustrated that sessions in which mobile map apps were used were rare and unevenly distributed across participants. Additionally, the Play Store categories MN and TL, used to identify map-related applications, are much broader and include applications not relevant to navigation. The second MoT II campaign enriched the dataset with demographic information and standardized self-report measures of sense of direction (SBSOD) and spatial anxiety (SA). A stricter filter was also applied to isolate only the navigation sessions, avoiding diluting the results by including MN and TL apps not directly related to navigation. Even with a reduced number of sessions, the dataset retained substantial richness, and the inclusion of SBSOD and SA scores allowed for the description of stable individual differences relevant to real-world navigation behavior.

While these descriptive statistics highlight both the potential and the limitations of the two data collection campaigns, their variability motivated further exploratory studies during which raw data were transformed into features that could be used to examine relationships between taps and GPS data, between individual traits and taps, and ultimately across all three levels of information combined. This is the focus of the second research question of this PhD dissertation (RQ2) and will be addressed in the following section.

4.2 From Exploration to Integration: Contextualizing Mobile Map App Interactions

This section reports the results that address how interactions with mobile maps can be meaningfully linked to individual differences and spatial context (RQ2). To answer this question, we draw on the features extracted across the three contextual levels (behavioral, intrinsic, and extrinsic; see 3.2.2) and report two complementary sets of results that mirror the analysis plan described in Section 3.2.3.

First, we present exploratory analyses (see Section 4.2.1) that establish preliminary empirical associations between extrinsic and behavioral contexts, including: (i) density and distributional analyses of tap activity, examined as a function of GPS-derived context, such as distance-from-home indicators (Zingaro & Reichenbacher, 2022); (ii) unsupervised profiling, performed using HCA and AA to identify recurring interaction patterns of mobile map use across distances from home (Zingaro et al., 2023). The association between intrinsic traits and behavior is then explored; and (iii) linking SBSOD, SA, and gender across a set of tap-related measures for map and non-map activity. (Zingaro et al., 2024)

In the second part, we move beyond preliminary exploratory checks and report the results of the analyses of the relationships between behavior, intrinsic, and extrinsic context to gain an integrated understanding of everyday navigation. In this section 4.2.2, we more comprehensively capture the dynamics of interaction, using more complex measures of engagement and structured modeling, as described in detail in Section 3.2.3. The results of two complementary studies are reported below.

The first study describes associations among the SBSOD and SA scores and within-session interaction groups, derived to capture the regularity and frequency of interactions of the mobile map as well as app-switching behavior, using UMAP to reduce dimensionality and DBSCAN for grouping (see 3.2.3). Then, the results of robust regression analyses are

reported, predicting SBSOD and SA scores using longer-term engagement metrics. These metrics were derived by aggregating the phone sessions according to participant (see 3.2.3) and quantifying the frequency and temporal regularity of navigation app usage.

The second study shifted the focus to explanatory modeling. We present an integrated analytical framework combining behavioral metrics within the session (see 3.2.3) with intrinsic traits and extrinsic urban form descriptors. We describe how behavioral outcomes were regressed on these predictors, using robust regression and moderation analyses to examine how traits and urban structure jointly influence interaction dynamics, and whether their effects vary under different environmental conditions. Finally, we report model fit, diagnostic evaluations, and sensitivity checks to assess the robustness of the findings. We begin the following section by reporting the results of the three exploratory analyses.

4.2.1 Exploratory Analyses of Mobile Map Interaction Across Contexts

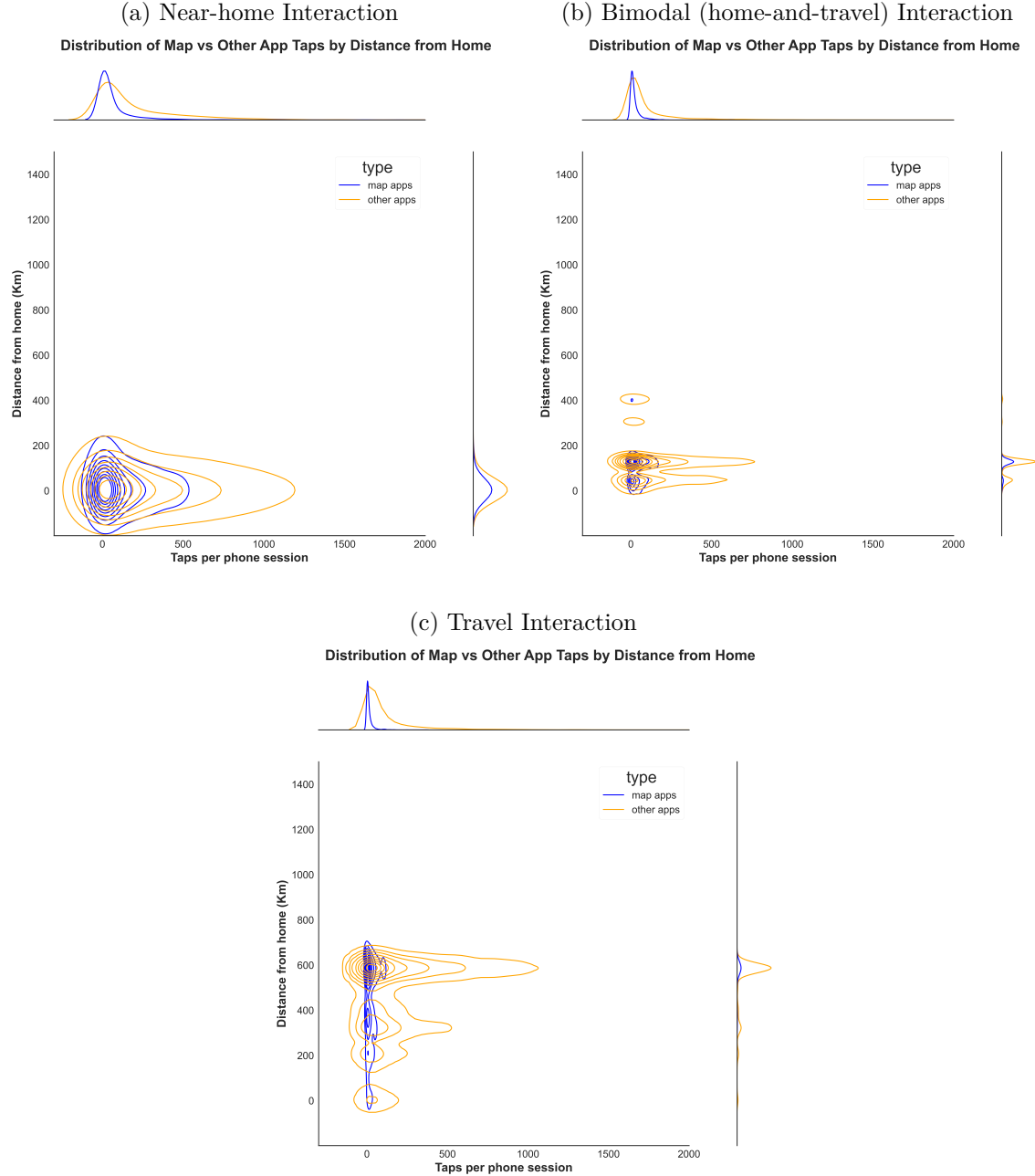
This subsection summarizes the three exploratory studies that provided the basis for the integrated analyses (see Section 4.2.2). The first two studies, based on MoT I data, report extrinsic-behavior results by examining how the environmental context (e.g., distance from home) influences mobile map interactions. The third study, based on MoT II data, reports intrinsic behavior results, linking individual traits such as sense of direction (SBSOD), spatial anxiety (SA), and gender to interaction patterns. These studies are detailed in the following publications:

1. Zingaro, D., & Reichenbacher, T. (2022). Exploratory analysis of mobile app usage in relation to distance from home. *Proceedings of the 17th International Conference on Location Based Services (LBS 2022)*, 59–67. <https://doi.org/10.5167/uzh-224336>
2. Zingaro, D., Bartling, M., & Reichenbacher, T. (2023). Exploring Map App Usage Behaviour Through Touchscreen Interactions (Short Paper). In *Proceedings of the International Conference on Geographic Information Science (GIScience 2023)*. Leibniz International Proceedings in Informatics (LIPIcs), Article 95. <https://doi.org/10.4230/LIPIcs.GIScience.2023.95>
3. Zingaro, D., Reichenbacher, T., Bartling, M., & Fabrikant, S. I. (2024). Exploring the Relation Between Sense of Direction and Spatial Anxiety in Everyday Mobile Map App Use (Short Paper). In *16th International Conference on Spatial Information Theory (COSIT 2024)* (Proceedings, Volume 315, pp. 15:1–15:8). Leibniz International Proceedings in Informatics (LIPIcs). <https://doi.org/10.4230/LIPIcs.COSIT.2024.15>

MoT I - Exploratory Extrinsic-Behavior Results

In the first exploratory study (Zingaro & Reichenbacher, 2022), we tested whether individuals' phone interactions were influenced by their distance from home, and specifically whether this effect was more pronounced for map-related apps. The underlying hypothesis was that being farther away from familiar locations increases the need for guidance or information. If this is true, then interactions with map-related apps should increase more

Figure 4.3.: *Kernel Density Estimates (KDEs) of Taps per Session as a Function of Distance From Home in MoT I. Each panel shows the interaction patterns for one participant, comparing map-apps (Maps and Navigation + Travel and Local; blue) and other apps (orange). The three recurring interaction profiles illustrated are: (a) near-home-dominant, with peak activity at 0–3.5 km; (b) bimodal, with two peaks at 0–3.5 km and 100–200 km; and (c) travel-dominant, with greater activity at longer distances.*



Note. Adapted from Zingaro and Reichenbacher (2022). *Exploratory analysis of mobile app usage in relation to distance from home.* In *Proceedings of the 17th International Conference on Location Based Services (LBS 2022)* (pp. 59–67). University of Zurich. <https://doi.org/10.5167/uzh-224336>

strongly compared to other types of apps as the distance from home increases (Ishikawa et al., 2008). To test this proposal, we analyzed the tappigraphy and GPS data of 38

participants (see Section 4.1 for descriptive analyses) and divided all the taps into two categories: taps in map-related apps (summing the taps of both MN and TL categories) and taps for all other apps (all other categories of the dataset). Across all of our phone sessions, there were 489,231 taps in map applications and 12,003,474 taps in other applications, which already suggests the wide discrepancy in our dataset on mobile map use.

To give spatial meaning to these interactions, we calculated the distance from the home location of each participant (detailed information on home location estimation is provided in Section 3.2.2) by calculating the Euclidean distance from home to the geographic centroid (average location) of the recorded GPS coordinates of each session using the Web Mercator projection, commonly used for online maps). The session distances varied widely, from 0 km (home-based) to 1393 km, although most of the sessions occurred relatively close to home ($M = 243$ km, $SD = 353$; $Mdn = 8.8$ km).

Because participants differed greatly in how often they used their phones and how much they traveled, we normalized their activity by calculating a measure called the “tap density”, which was the median number of taps divided by the median distance traveled from home (taps per kilometer). Using medians for the analysis helped minimize the influence of unusually high activity or days involving long travel. Tap densities varied significantly (from 0.016 to 31,875 taps/km; $M = 1,287$, $SD = 5281$, $Mdn = 2.63$ taps/km). This large variation suggested that participant’s phone interactions might change depending on how far they traveled from home. Rather than analyzing all interactions together in a single average, we grouped sessions by their distance from home and analyzed interactions separately within these groups. We used KDEs to visualize these grouped interactions, revealing three distinct interaction patterns among participants. Similar patterns were observed for interactions with map apps and other types of app (see Figure 4.3 a–c):

1. **Near-Home Interaction:** Most taps occurred within 0–3.5 km of home ($n = 9$ participants; 24%).
2. **Bimodal (Home-and-Travel) Interaction:** Two peaks of taps, one around 0–3.5 km and a second at a mid-range distance (100–200 km; $n = 18$ participants, 47% of the total).
3. **Travel Interaction:** Tap activity increased significantly at greater distances from home ($n = 11$ participants, 29% of the total).

To test whether distance from home had a stronger effect on map-specific behavior compared to behavior on other types of application, we compared tap densities separately for map apps and other apps. Wilcoxon signed rank tests demonstrated significant increases in tap density away from home for both map apps ($Z = 17.0$, $p < .001$) and other apps ($Z = 15.0$, $p < .001$). Thus, being farther from home increased overall phone interaction, but users did not exhibit increased use of map applications compared to their use of other apps.

These findings offer two key insights: (a) distance from home significantly influences how intensely people interact with their phones, and (b) the frequency of taps alone does not uniquely indicate map-specific behavior since other apps show similar patterns. In other words, being away from home generally increases phone interactions, but does not necessarily reflect map-specific usage.

In the second exploratory study (Zingaro et al., 2023), we tested whether participants’ use of mobile map apps at varying distances from home organizes into a small set of recurring

shared patterns and whether these patterns differ between the two navigation-related Play Store categories (MN and TL), or if they are entirely unique to each individual. To ensure tapping behavior was comparable among participants, we transformed raw tapping and GPS data into standardized **distance interaction profiles** by calculating the distance from home, following the procedure described in Section 3.2.2. After removing 2121 extreme session distance records using an interquartile range filter (2.85% of the original 74,304 records), the cleaned dataset comprised 72,183 session–distance records from 30 participants (range = 0–1393 km; $M = 243$ km, $SD = 353$ km, $Mdn = 8.8$ km). The strong right skew (small Mdn , large M , and SD) reflects a concentration of sessions close to home with a long tail of infrequent long distance trips; the IQR filter removed extreme outliers so that subsequent binning and clustering would capture common distance-dependent patterns rather than being driven by a few rare journeys. We partitioned these distances into 100 classes with the Fisher–Jenks natural-breaks algorithm because it produces bins that reflect the intrinsic structure of the data rather than arbitrary cutoffs. For each participant, and separately for the MN and TL taps, we counted the taps per distance bin and then z -standardized those counts within each participant so that the subsequent grouping would be driven by the *shape* of each participant’s distance profile (distributional pattern) rather than by absolute tap volume. The descriptive statistics for the participants and MN and TL categories are available in the project repository, [GitLab](#).

We applied two complementary unsupervised techniques to the standardized distance–interaction profiles to uncover recurring spatial-usage patterns for the two distinct mobile map application categories.

First, we applied agglomerative HCA with Ward linkage to group participants according to the *shape* of their distance–interaction profiles. Ward linkage was chosen because it builds clusters by joining pairs that produce the smallest increase in within-cluster variation; that is, it creates groups that are as internally similar as possible. HCA produces a tree-like diagram (a *dendrogram*) that illustrates how participants merge into progressively larger groups (Figure 4.4 (a)). We inspected this dendrogram and cut it at the largest jump between nodes (the largest vertical gap), yielding a stable two-cluster solution. These two clusters are described below:

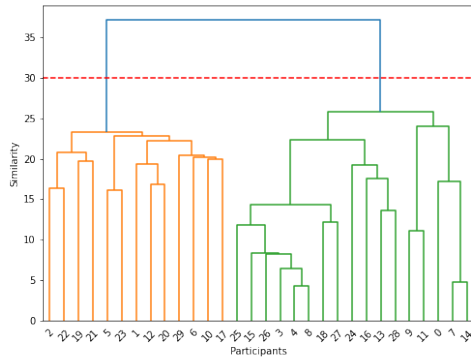
- **Cluster_{Near Home}** ($n = 17$ participants): interaction concentrated very close to the estimated home location (strong positive z -scores in the smallest distance bins, with patterns seen in both MN and TL).
- **Cluster_{Travel}** ($n = 13$ participants): Interaction spread across larger distances, with multiple peaks at intermediate and long ranges (notably 5–10 km, 100–150 km, 300 km; see Figure 4.4 (b)).

The HCA solution therefore divides participants into two broad spatial-use types: a local, near-home mode and an extended, travel-related mode. It showed that both MN and TL taps contribute to each signature, implying that these spatial patterns are not confined to a single Play Store category.

Because HCA assigns each participant to a single cluster, more subtle or mixed behaviors can be overlooked. To address this issue, we also applied AA, which captures a continuous spectrum of “pure types” (archetypes) and then represents each participant as a convex combination of these archetypes (see 3.2.3). We evaluated archetypal models for different numbers of archetypes by inspecting the residual sum of squares (RSS). A four-archetype solution (RSS 0.68) provided the best balance between minimizing reconstruction error and

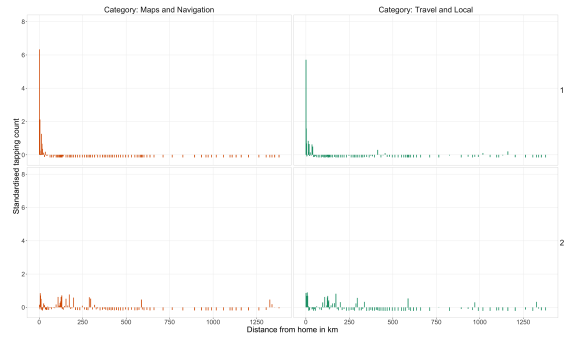
Figure 4.4.: *MoT I: Unsupervised Profiles of Distance-Related Tapping Behavior.* Panels (a–d) present the results from the hierarchical clustering analysis (HCA) and archetypal analysis (AA). The HCA dendrogram (a) and mean cluster profiles (b) indicate two main groups: participants whose tapping was concentrated near home and those whose activity extended into travel distances. AA (c, d) further decomposed behavior into four archetypes with two dominant signatures: a compact near-home profile (A_2) and a more dispersed travel-oriented profile (A_3). Together, both approaches highlight a consistent distinction between near-home and travel-related interaction patterns.

(a) HCA dendrogram (Ward linkage)



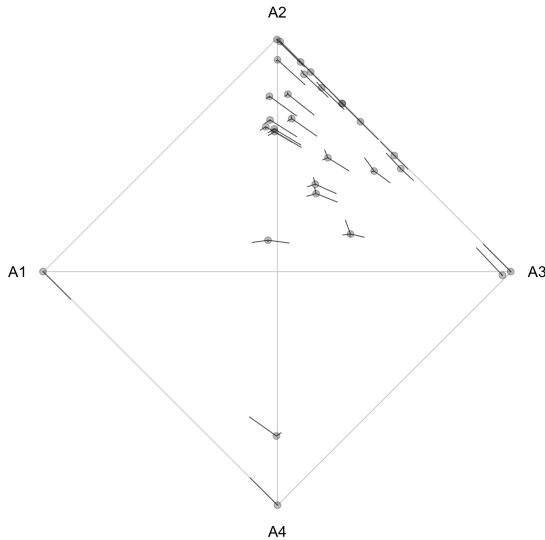
Cut at largest inter-node jump $\Rightarrow$ two clusters.

(b) HCA cluster profiles (two clusters)



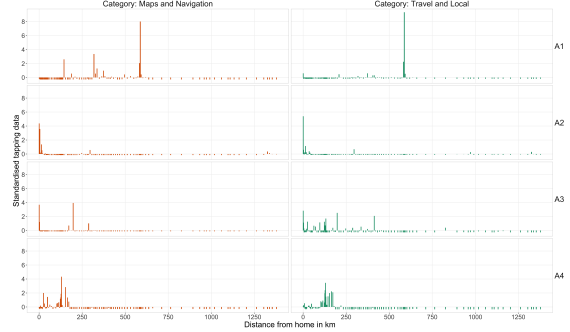
Cluster sizes: 17 (near-home interactions), 13 (travel interactions).

(c) Archetypal simplex (four-archetype solution)



Points = participants; vectors indicate principal affiliation directions.

(d) Archetype profiles (MN/TL)



A_2 : near-home; A_3 : up to ~ 200 km; A_1/A_4 : mid-range focus.

Note. MN = Maps & Navigation; TL = Travel & Local. Adapted from Zingaro et al. (2023). *Exploring map app usage behavior through touchscreen interactions (Short Paper).* In *Proceedings of the International Conference on Geographic Information Science (GIScience 2023)*. Leibniz International Proceedings in Informatics (LIPIcs). <https://doi.org/10.4230/LIPIcs.GIScience.2023.95>

maintaining interpretable, well-supported archetypes that represented consistent behavioral patterns shared by several participants rather than mathematical artifacts. Figure 4.4 (c) shows participants in the archetypal simplex as well as directional membership vectors indicating the strength and principal direction of each participant's affiliation. The four archetypes (A1–A4) can be summarized as follows (Figure 4.4 (d)):

- **A₂ (home behavior).** Dense, high-intensity interactions concentrated close to home (mainly within 13 km). Both MN and TL taps contribute substantially to this profile. This archetype attracted the largest number of participants and therefore represents the dominant near-home signature of the sample.
- **A₃ (travel behavior).** More dispersed interaction extending up to 200 km, with a relatively higher share of TL taps which were consistent with travel or planning-related activity. Many participants exhibit substantial affiliation to this archetype.
- **A₁ (long-range) and A₄ (mid-range).** These archetypes capture less common signatures: A1 emphasizes very long-range interactions (peaks roughly in the 300–600 km band) and A4 emphasizes mid-range activity (100–200 km). Both A1 and A4 have low support (one or two participants each) and thus function as outliers or low-frequency extremes in the sample.

The simplex plot (Figure 4.4 (c)) shows that most of the participants are positioned between archetypes rather than aligning with a single extreme. This indicates that there are gradual transitions between interaction modes rather than sharp partitions.

Comparing AA and HCA results shows that both methods converge on two broad behaviors: near home and traveling. However, AA reveals a richer structure by exposing intermediate and rare extreme signatures, showing that many participants are mixtures of archetypes rather than members of a single discrete cluster. This graded representation is visible in the simplex plot, where most of the participants lie between archetypal vertices instead of at a single corner.

Combined, the results of this second exploratory study demonstrate that clusters based on minimal signals, such as tap count and distance from home, can be useful for identifying recurring tap patterns on mobile map applications as a first step towards understanding the extrinsic–behavior contextual interplay. However, this minimal enrichment is insufficient to explain the meaning of these interactions for the user. This motivated the next campaign, which added layers of context (e.g., users' traits and demographics).

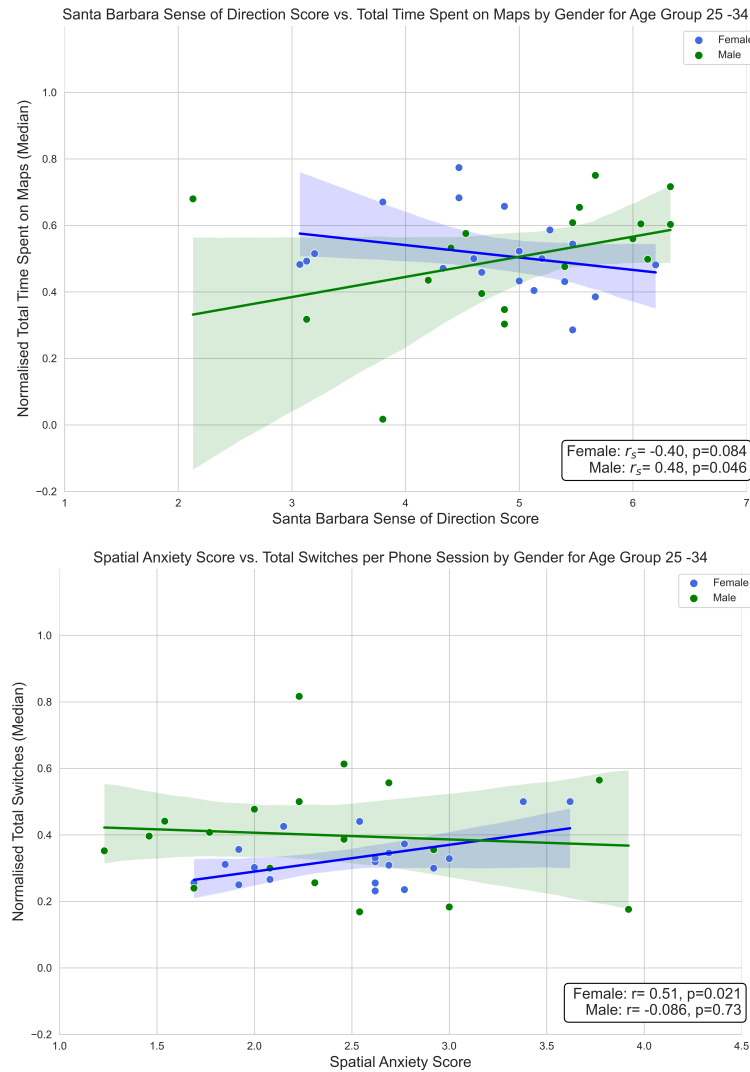
After exploring these primary spatial interaction patterns in MoT I, we will turn to the exploratory analyses of the data gathered in MoT II. MoT II was designed to extend the data collection to include individual traits, ability measures, and demographics, in addition to the already established tappigraphy and sensor data.

MoT II - Exploratory Trait–Behavior Results

In MoT II, we examined the relationship between individual differences in the SBSOD and SA scales and the patterns of interaction related to everyday navigation captured by the tappigraphy data. Drawing on prior research, we hypothesized that participants with higher SBSOD scores would spend less time using map apps and would switch between

map apps and other apps more often. Conversely, we hypothesized that participants with higher SA scores would spend longer using map apps and engage in less switching between map apps and other apps.

Figure 4.5.: *Gender-Specific Correlations in the 25–34 Age Band (MoT II).* Among men (top panel), more time spent using map apps per session was positively associated with sense of direction ($r_s = .48$, $p = .046$, 95% CI [.02, .77]). Among women (bottom panel), more frequent app switching per session was positively associated with spatial anxiety ($r = .51$, $p = .021$, 95% CI [.09, .78]). These findings suggest that navigation app use patterns may relate differently to spatial traits across genders.



Note. Zingaro et al. (2024). *Exploring the Relation Between Sense of Direction and Spatial Anxiety in Everyday Mobile Map App Use (Short Paper)*. In *16th International Conference on Spatial Information Theory (COSIT 2024)*. Leibniz International Proceedings in Informatics (LIPIcs) (pp.15:1–15:8) . <https://doi.org/10.4230/LIPIcs.GIScience.2023.95>

Due to the uneven age distribution of the filtered sample, gender-specific analyses were restricted to the most represented age group (25–34 years; $n = 38$). Unless otherwise specified, all tests were two-sided with $\alpha = .05$ and 95% confidence intervals.

We computed correlations between each trait measure and the phone session features defined in Section 3.2.2, including the number of app switches to or from map apps, total app switches, proportion of map switches, and time spent in map apps. These variables were extracted from sessions filtered by navigation, logarithmically transformed when necessary, normalized per participant to the range $[0, 1]$, and summarized using medians per participant. Only the significant correlations are reported here; full correlation analyses, including all additional features, are provided in the Appendix A.7.

Trait Check (Full MoT II Sample, $N = 38$).

SBSOD and SA were inversely correlated ($r = -.33$, $p = .013$, 95% CI $[-.55, -.07]$), replicating the canonical negative association between sense of direction and spatial anxiety.

Gender-Specific Associations (25–34 Band, $n = 38$).

Among men, higher SBSOD was associated with *more* time spent in map apps per phone session (Spearman $r_s = .48$, $p = .046$, 95% CI $[.02, .77]$; Figure 4.5 top panel). Among women, higher SA was associated with *more* app switches per session (Pearson $r = .51$, $p = .021$, 95% CI $[.09, .78]$; Figure 4.5 bottom panel).

Complementary analyses with the other phone session features that are summarized in Section 3.2.2 were not significant; these are reported in Appendix A.7.

Our previous exploratory analyses of MoT I and MoT II showed that our data from multiple sources (taps, GPS, questionnaires) could provide insights into where people tap on map applications and whether time spent on mobile apps or the frequency of switching between them could be influenced by sense of direction, spatial anxiety levels, or potential gender biases. However, these metrics are rather simple and do not reflect information about how frequently these taps occur over time for a specific application, nor do they reflect whether this interaction is regular within and across phone sessions for our participants. The same simplicity also applies to GPS data measures such as distance from home, which neglects other aspects of the complexity of the urban environment that individuals navigate.

In the next section, we translate raw tappigraphy and GPS traces into more comprehensive temporal interaction metrics. This allows us to examine how the intrinsic traits first, and then the intrinsic and extrinsic dimensions, jointly influence the intensity, regularity, and structure of daily map application use.

4.2.2 Core Analyses: Integrating Behavioral, Intrinsic, and Extrinsic Features

We conducted two core studies based on the MoT II dataset. These analyses enabled us to jointly consider the intrinsic (e.g., sense of direction, spatial anxiety, and gender), behavioral (e.g., interaction dynamics captured via tappigraphy), and extrinsic (e.g., urban form features; see Section 2.3) dimensions of navigation behavior.

The first study quantified the frequency and regularity of mobile map interactions that are strictly related to navigation at two different temporal scales: (1) Across phone sessions (i.e., aggregated according to participant) to capture the stable tendencies of each individual over time; and (2) within phone sessions to capture moment-to-moment interaction dynamics occurring inside active phone sessions when users deploy mobile map applications strictly related to navigation. Then, we used these features to understand their association with SBSOD and SA.

The second study extended the analysis to include within-session features and individual traits by incorporating urban form features into predictive models. We tested whether spatial context, alongside stable individual traits, jointly explains variations in interaction patterns.

Collectively, these studies provide a clearer, more interpretable account of how mobile map interactions unfold and how they are shaped by intrinsic, behavioral, and extrinsic factors. This section summarizes the findings of two complementary manuscripts:

1. Zingaro, D., Reichenbacher, T., & Fabrikant, S.I. (under review). *How Spatial Abilities, Emotional Traits, and Urban Form Shape Interaction with GPS-Embedded Map Applications*. Submitted to *ACM Transactions on Spatial Algorithms and Systems* (Special Issue on Urban Mobility).
2. Zingaro, D., Reichenbacher, T., & Fabrikant, S.I. (in preparation). *A Parsimony-Based Approach to Detecting Individual Differences in Spatial Abilities and Emotional Traits through Touchscreen Interactions in GPS-Assisted Navigation*. Target: *Behavior Research Methods*.

We begin by displaying the results of the first study of the core analyses.

Interaction Dynamics and Trait Associations Across and Within Phone Sessions

Instead of simply counting how many times participants used navigation apps (a simple tap measure), we wanted to focus on when and how often users returned to them by extracting temporal features across participants that could reveal habits of reliance on navigation that may be related to individual traits. As mentioned in the Methods section 3.2.2, we extracted two measures to do this. The IMI captures the typical time gap between two navigation sessions. Shorter IMI values mean that a person tends to return to navigation quickly and frequently, whereas longer IMI values mean that they leave more time between uses. Because very long breaks can distort the average, we applied a $\log_{10}$ transformation to the gaps and used the 25th percentile, following similar studies on ITI intervals based on tappigraphy data (see Ceolini et al. (2022) for details). The SIR indicates how frequently someone initiates new navigation sessions throughout the study period. It is calculated from the average logarithmic transformation intervals and then inverted, so higher values mean more frequent initiation. Unlike IMI, which focuses on a typical gap, SIR provides a broad indicator of a participant's general tendency to re-engage with navigation apps. Together, IMI and SIR provide complementary perspectives: IMI highlights the short-term rhythm of returning to navigation, while SIR captures the longer-term rate of initiating sessions. The analysis covered 3585 navigation sessions from 55 participants (after cleaning,

accounting for phone session with more than one tap over our MN/TL categories of interest to compute the time intervals). We then followed these three steps: (1) we described IMI and SIR to aid understanding of the variability of the retrieved engagement features; (2) we tested simple correlations of IMI and SIR with SBSOD and SA to see whether individual traits are related to these measures; and (3) we ran regression models to evaluate these associations while controlling for gender as a covariate.

Table 4.12.: *Descriptive Statistics for Participant-Level Variables (N = 55). The table shows that IMI values were highly skewed, ranging from only a few seconds to several days, while SIR values were more stable across participants. Trait scores also showed variability, with SBSOD spanning most of the 1–7 scale and SA ranging from low to moderately high values.*

Variable	M	SD	Min	25%	50%	75%	Max
IMI (25th percentile)	12,227.60	45,625.14	8.99	57.82	121.47	236.18	244,640.50
SIR	0.1781	0.0341	0.1597	0.1752	0.1752	0.1752	1.7750
SBSOD Score	4.4955	1.1164	2.13	3.20	4.87	5.27	6.73
SA Score	2.2987	0.5278	1.23	1.77	2.62	2.62	3.92

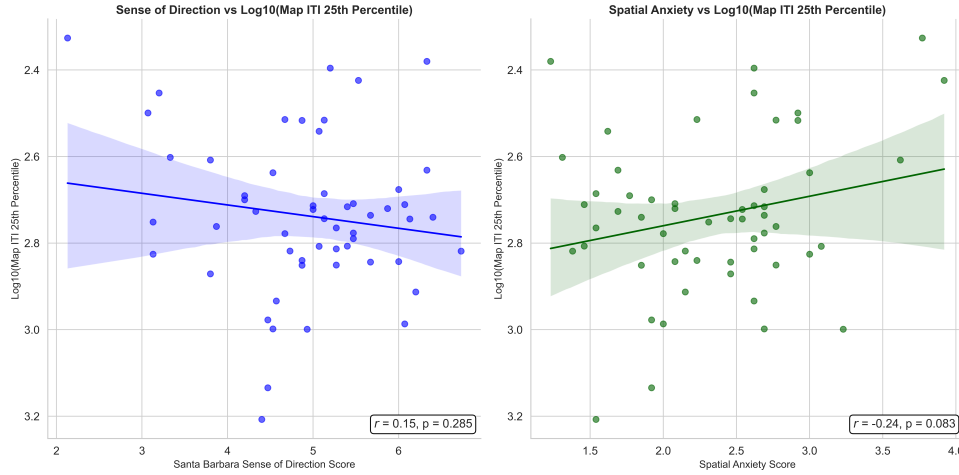
Note. IMI = inter-map interval; values represent each participant’s 25th percentile of inter-session gaps (s) between navigation sessions (sessions with ≥ 1 tap in MN or TL). SIR = session initiation rate, calculated as the inverse of the mean $\log_{10}$ -transformed inter-map intervals. SBSOD = Santa Barbara Sense of Direction scale (Hegarty et al., 2002); SA = Spatial Anxiety scale (He & Hegarty, 2020). Only participants with at least two valid navigation sessions were included ($n=3585$); sessions with zero MN and TL taps were excluded.

As shown in Table 4.12, the descriptive statistics underline the clear differences in navigation behavior and traits between participants. For IMI, the average gap between navigation sessions was approximately 12,228 seconds (approximately 3.4 hours); however, the values ranged from 9 seconds to more than 244,000 seconds (approximately three days), indicating uneven intervals between navigation sessions. SIR values were more stable, with a mean of 0.18 ($SD = 0.03$), but still ranged from 0.16 to 1.78, reflecting the variation in the frequency with which the participants started new sessions. Individual traits scores varied less. SBSOD had a mean of 4.50 ($SD = 1.12$) on a scale of 1 to 7 points and ranged from 2.13 to 6.73. SA had a mean of 2.30 ($SD = 0.53$) on a scale of 1 to 5 points and ranged from 1.23 to 3.92. The descriptive statistics provided the foundation for testing the systematic relationship between IMI and SIR with sense of direction and spatial anxiety using correlations.

IMI showed no significant associations with trait measures. The correlation with SBSOD was small and positive ($r = 0.147$, $p = 0.285$), while the correlation with SA was small and negative ($r = -0.236$, $p = 0.083$). While the IMI–SA result suggested that more frequent navigation might be linked to slightly lower spatial anxiety, this did not reach statistical significance. Figure 4.6 depicts the scatter plots for both relationships. Gender-stratified analyses also showed no significant effects. Among women, IMI was correlated with SBSOD at $r = 0.232$ ($p = 0.275$) and with SA at $r = -0.298$ ($p = 0.158$). Among men, the corresponding values were $r = 0.134$ ($p = 0.471$) for SBSOD and $r = -0.255$ ($p = 0.166$) for SA.

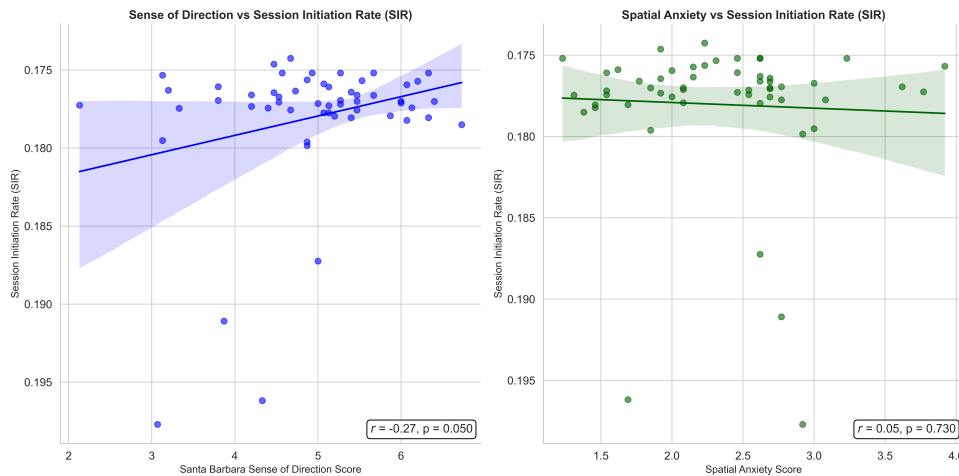
SIR was not clearly associated with trait measures (see Figure 4.7). The correlation with SBSOD was negative and at the statistical significance threshold ($r = -0.266$, $p = 0.050$)

Figure 4.6.: *Relationships Between Inter-Map Intervals (IMI) and Trait Measures. Scatter plots show Santa Barbara Sense of Direction (SBSOD) scores (left) and Spatial Anxiety (SA) scores (right) plotted against $\log_{10}$ -transformed 25th percentile IMI values. Correlations: SBSOD, $r = .15$, $p = .285$; SA, $r = -.24$, $p = .083$. Lines represent least-squares fits; shaded areas indicate 95% confidence intervals.*



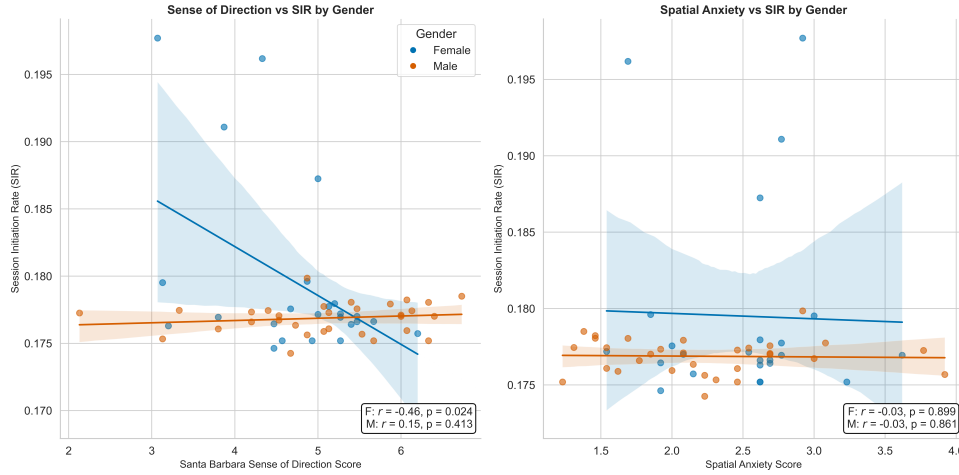
without a clear association. The correlation between SIR and SA was close to zero and was not significant ($r = 0.048$, $p = 0.730$). The gender-stratified results (see Figure 4.8) suggested that the negative correlation between SIR and SBSOD was stronger among women ($r = -0.459$, $p = 0.024$) and absent among men ($r = 0.153$, $p = 0.413$). The correlations between SIR and SA were negligible in both women ($r = -0.027$, $p = 0.899$) and men ($r = -0.033$, $p = 0.861$).

Figure 4.7.: *Relationships Between Session Initiation Rate (SIR) and Trait Measures. Scatter plots show Santa Barbara Sense of Direction (SBSOD) scores (left) and Spatial Anxiety (SA) scores (right) plotted against SIR, defined as the inverse of the mean $\log_{10}$ -transformed inter-map interval. Correlations: SBSOD, $r = -.27$, $p = .050$; SA, $r = .05$, $p = .730$. Lines represent least-squares fits; shaded areas indicate 95% confidence intervals.*



For the final step, we ran OLS regression models. Since IMI did not exhibit clear associations

Figure 4.8.: Relationships Between Session Initiation Rate (SIR) and Santa Barbara Sense of Direction (SBSOD) Scores Stratified by Gender. Scatter plots show SIR plotted against SBSOD for women (left) and men (right). Correlations: women, $r = -.46$, $p = .024$; men, $r = .15$, $p = .413$. Lines represent least-squares fits; shaded areas indicate 95% confidence intervals.



with the trait measures, we only report the results for SIR and SBSOD since the correlation was borderline significant ($p = .050$) and appeared to be stronger among women, raising the question of whether gender was driving this association. OLS regression allowed us to test this directly by including gender as a covariate. This analysis was therefore important in determining whether the observed SIR–SBSOD link reflects a genuine gender-related effect or whether it disappears once gender is taken into account. For completeness, regressions for IMI and SIR with SA are reported in the Appendix A.8.

For SBSOD, the regression model explained 7.1% of the variance ($R^2 = 0.071$, $p = 0.050$). The SIR coefficient was negative ($\beta = -56.91$, 95% CI $[-113.81, -0.02]$), indicating that participants who initiated navigation sessions more frequently reported a weaker sense of direction. Because the p -value was exactly at the threshold, this finding should be interpreted with caution. An a priori power analysis for a single predictor ($m = 1$) indicated that when $N = 55$, the statistical power to detect a moderate effect was approximately 0.79. This suggests that it is likely that moderate effects would have been detected, but replication is needed to confirm the authenticity of the SIR–SBSOD link.

Including gender as a covariate did not change the interpretation. The model explained 8.2% of the variance ($R^2 = 0.082$, $p = 0.107$), and the SIR effect was no longer significant ($\beta = -49.91$, 95% CI $[-109.56, 9.74]$, $p = 0.099$). Gender was also unrelated to SBSOD ($\beta = 0.22$, 95% CI $[-0.24, 0.68]$, $p = 0.419$).

The a priori power analysis for two predictors ($m = 2$) indicated that $N = 55$ provided only 0.68 power to detect a moderate effect. This means that the models were underpowered for moderate-to-small effects. In practice, the non-significant findings cannot be viewed as evidence of no effect, since small effects may have gone undetected, while the borderline significant findings cannot be considered conclusive and require replication with larger samples.

Taken together, these analyses using aggregated measures over participants showed that only the probability of starting a new session of navigation app (SIR) demonstrated a borderline negative association with participant sense of direction; this effect was particularly strong among women. The disappearance of the SIR–SBSOD effect once gender was included in the regression analysis suggests that the general SIR–SBSOD association reflects gender differences rather than a uniform effect across all participants. These weak associations are not unexpected given our limited dataset, which covers only a few weeks rather than the longer periods (months or years) typically needed to stabilize the intersession indices (Ceolini et al., 2022). Additionally, collapsing behavior into single-person-level scores obscures moment-to-moment variability, motivating the move to session-level analyses, where interaction patterns can be examined over shorter timescales and in relation to a situational context.

After finding only weak associations between person-based navigation measures (IMI and SIR) and individual traits, we turned to within-session analyses. Our goal was to capture the temporal dynamics within individual navigation sessions to examine whether short-term tapping patterns on mobile maps are associated with individual traits, such as sense of direction, spatial anxiety, and gender. We extracted three metrics that together describe different aspects of session-level behavior as described in our methodology section 3.2.2. The CV-ITI (coefficient of variation of inter-tap intervals) measures the regularity or irregularity of tapping within a navigation-related phone session, with higher values indicating more irregular rhythms. The tap rate (taps per second) measures the interaction intensity during the session, reflecting how actively the map app is being used.

Finally, the transitions per minute capture how frequently participants switch between different applications during the phone session. The raw distributions of the three session-level metrics were highly right-skewed; to address this skewness and stabilize the variance between sessions, we applied a transformation $\log(1 + x)$ before analysis.

This approach also handles zero values (e.g., sessions in which there were no transitions or very low tap rates), since adding 1 ensures that all inputs are positive before taking the logarithm. The final dataset accounted for $N = 3667$ phone sessions that were strictly related to navigation. Table 4.13 summarizes the distributions of the three session-level metrics. CV-ITI values (log-transformed) averaged 0.89 ($SD = 0.42$), with a range from 0 to 2.60, indicating substantial variation in tapping regularity across sessions. In contrast, the distributions for taps/s and transitions/min were highly compressed on the log scale ($M = 0.0008$, $SD = 0.0011$, range = 0–0.049 for taps/s; $M = 0.0027$, $SD = 0.0072$, range = 0–0.199 for transitions/min), reflecting that most sessions involved minimal interaction and little switching, although some sessions reached much higher values.

We aimed to test whether different combinations of tapping regularity (CV-ITI), interaction frequency (taps per second), and switching (transitions per minute) formed recurring patterns of navigation use. This is because we wanted to capture more complex dynamics of interaction within sessions where these three features can drive different aspects of interaction behavior specific to navigation, rather than treating each metric in isolation.

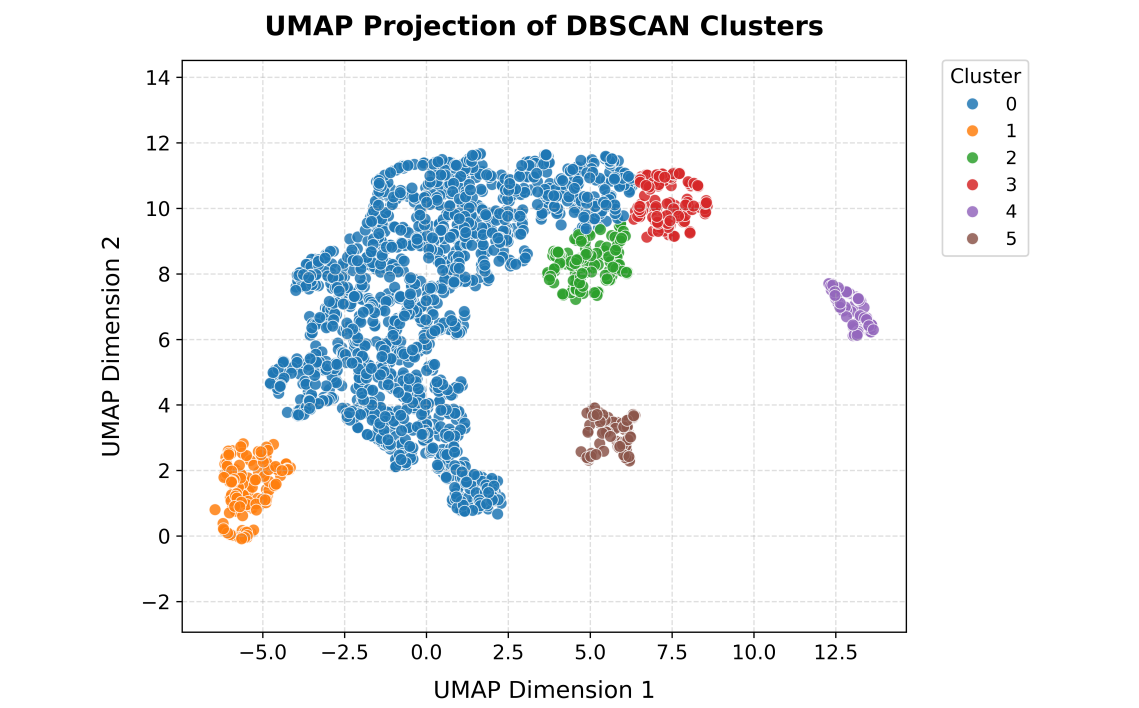
We therefore standardized the logarithmic-transformed metrics (z-scoring) and embedded them in a 2D space using UMAP. Then, DBSCAN was applied to detect dense session regions without fixing the number or shape of the clusters (see Section 3.2.3 for details on the methods and parameters). Low-density points were labeled as noise. This unsupervised

Table 4.13.: *Descriptive Statistics for Session-Level Interaction Metrics (N = 3667 Sessions). Tapping regularity (CV-ITI) exhibited substantial variability across sessions. In contrast, interaction frequency (taps per second) and switching behavior frequency (transitions per minute) were relatively low. Only a few sessions exhibited much higher values.*

Statistic	CV-ITI (log)	Taps/s (log)	Transitions/min (log)
M	0.893	0.000834	0.002707
SD	0.417	0.001098	0.007155
Min	0.000	0.000001	0.000000
Mdn	0.859	0.000622	0.000938
Max	2.601	0.048790	0.198891

Note. CV-ITI = coefficient of variation of inter-tap intervals; taps/s = taps per second; transitions/min = app transitions per minute. All values are based on $\log(1 + x)$ -transformed metrics to reduce skewness and stabilize variance. Higher CV-ITI values indicate greater tapping irregularity. The low averages for taps/s and transitions/min reflect that most navigation sessions involved minimal interaction and few switches, while the maximum values highlight occasional bursts of more intensive use.

Figure 4.9.: *UMAP Embedding of Session-Level Metrics (CV-ITI, Taps/s, Transitions/min) with DBSCAN Clusters. The two-dimensional projection shows six distinct clusters (colored regions) covering 69% of valid sessions, while the remaining sessions were classified as noise (label -1, not shown). One large cluster (Cluster 0) emerged, while the others were smaller and more compact, clearly separated in the embedding space. Axes represent UMAP components in arbitrary units.*



approach is well-suited for heterogeneous data, which is what real-world phone sessions produce, and resulted in six groups (n= 0–5), covering 2537 of 3667 valid sessions (approx-

imately 69%). The remaining sessions were labeled as noise. As shown in Figure 4.9, one large cluster (Cluster 0) emerged; this pattern suggests that most of the clustered sessions followed a broad interaction style. The remaining clusters were smaller and more compact, clearly separated in the embedding space, and formed more specific modes of behavior.

We then focused on describing and testing individual traits (SBSOD, SA, gender) across clusters to examine whether behaviorally defined interaction modes correspond to differences in spatial abilities or anxiety. Descriptive summaries of traits for each DBSCAN group are provided in Table 4.14. The mean spatial anxiety score ranged from 1.88 in Cluster 2 to 2.43 in Cluster 6, with the other groups falling between these values. The mean sense of direction score ranged from 4.64 in Cluster 3 to 5.13 in Clusters 4–5. The other clusters again fell within this range. These averages indicate that there was little variation and that the distributions substantially overlapped, suggesting that the behavior-defined interaction modes were not strongly differentiated by trait levels.

Table 4.14.: *Descriptive Statistics of Spatial Anxiety and Sense of Direction Scores for Participants Grouped into DBSCAN Clusters (Noise Cluster Excluded). Values show the number of sessions (n), means (M), and standard deviations (SDs) for each cluster. The table illustrates the slight variations in average trait levels across clusters, with overlapping distributions.*

Cluster	Spatial Anxiety			Sense of Direction		
	n	M	SD	M	SD	n
1	1876	2.21	0.54	4.93	0.98	1876
2	176	1.88	0.49	5.04	0.78	176
3	148	2.19	0.59	4.64	1.03	148
4	131	2.23	0.59	5.11	1.09	131
5	102	2.34	0.53	5.13	0.92	102
6	104	2.43	0.46	4.76	0.78	104

Note. SA = spatial anxiety (He & Hegarty, 2020); SBSOD = Santa Barbara Sense of Direction scale (Hegarty et al., 2006). Noise cluster (label -1) excluded.

We wanted to determine whether stable individual differences in sense of direction and spatial anxiety were reflected in these distinct clustered behavioral patterns. If these traits influence interaction, participants with different SBSOD or SA levels should exhibit different session distributions among the identified groups. Since participants could have sessions in multiple groups, we tested this by calculating their proportion of sessions in each group. We excluded one cluster from each model to avoid perfect multicollinearity. Using this approach, we tested whether the distribution of participants' sessions across behavioral modes was related to their stable individual traits.

In other words, the proportion of sessions in each cluster enabled us to summarize the typical interaction profile of an individual in a directly comparable manner between individuals. We used these proportions as predictors in regression models (OLS), along with gender, to examine whether participants with varying levels of sense of direction or spatial anxiety relied more on certain interaction patterns than others.

For SBSOD, the model explained 16.1% of the variance ($R^2 = .161$, adjusted $R^2 = .039$)

but was not statistically significant overall ($F_{7,48} = 1.318$, $p = .263$). None of the individual predictors reached significance, indicating that differences in cluster composition and gender did not explain the variation in sense of direction scores. For SA, the model explained 22.6% of the variance ($R^2 = .226$, adjusted $R^2 = .114$) and was not statistically significant overall ($F_{7,48} = 2.007$, $p = .074$). None of the predictors reached conventional significance, although gender ($p = .126$) and the proportion of sessions in Cluster 1 ($p = .081$) exhibited non-significant trends. These results suggest that the distribution of sessions across behavioral modes was not strongly related to spatial anxiety.

The six clusters represented common patterns of session-level interaction, but were not strongly associated with individual traits. Regression analyses showed that the share of sessions in each group did not explain differences in sense of direction and was not significantly related to spatial anxiety. The SA model showed only weak, non-significant trends (Cluster 1, $p = .081$; gender, $p = .126$), but no predictor reached conventional significance. Overall, the six groups describe general navigation use styles but did not clearly reflect stable traits such as sense of direction or spatial anxiety.

An Integrated Understanding of Mobile Map Interaction in Everyday Navigation

Finally, we examined the relationship between extrinsic and intrinsic variables and the dynamics of interaction within the session. For this analysis, we focused on two key metrics, CV-ITI and taps per second, which capture interaction frequency (see 3.2.3 for details). These measures provide a direct way to track the frequency and regularity with which participants interacted with navigation apps during a phone session. We excluded transitions per minute because switching behavior is more likely to reflect multitasking or app management rather than being directly driven by the navigation environment.

The spatial context was operationalized through urban form descriptors computed for each session's location (see 3.2.3 for details). These measures capture the structural characteristics of the street network, such as connectivity, directness, and spatial complexity, which can influence navigation behavior (Coutrot et al., 2022). Our aim here was to test whether moment-to-moment interaction patterns are shaped jointly by who the navigator is (their traits) and where they are navigating (the urban form).

The final dataset included 2399 GPS-embedded navigation sessions from 46 participants (28 men, 18 women). We analyzed these data in two stages, using:

1. **Robust regression models** to estimate the main effects of traits and urban form on the frequency and regularity of the interaction;
2. **Moderation models** to test whether the effects of traits vary according to environmental conditions, as captured by urban form.

Additional descriptive statistics, distributions, and details on urban metrics pre-processing (sensitivity analyses) are provided in the Appendix A.6.

Robust Regression Models. We estimated the main effects of traits (SBSOD, SA, and gender) and urban form descriptors on interaction frequency (taps/s) and regularity (CV-ITI) using robust OLS regression. We planned to use mixed-effects models with random

intercepts to account for the nested structure of the data (sessions within participants). However, the intraclass correlation coefficients (ICCs) for the two main outcome variables, tap rate (ICC = 0.130) and CV-ITI (ICC = 0.106), indicated only modest within-subject clustering. The preliminary mixed-effect models did not improve the fit of the model or materially change the results, but added complexity to the interpretation. Since our theoretical focus was on session-level behavior in different urban contexts rather than between-subject variance, we opted for fixed-effects models. We estimated these models using robust OLS to address potential heteroscedasticity.

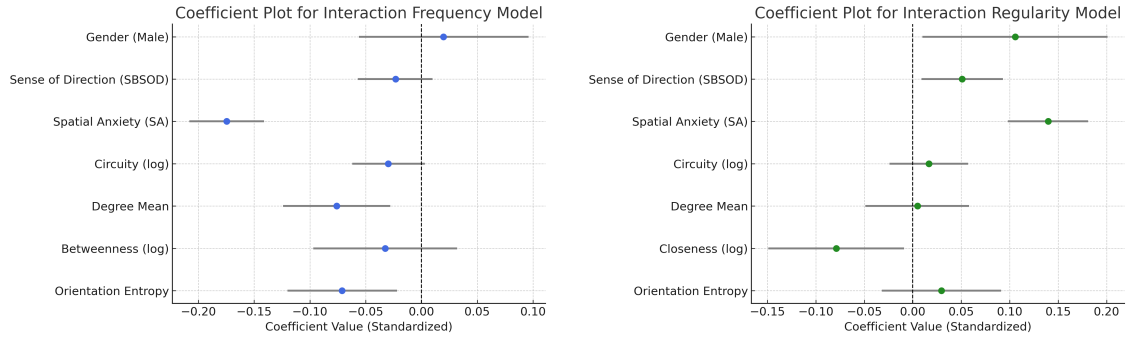
The multicollinearity between urban form variables presented an additional modeling challenge. Specifically, mean betweenness and mean closeness centrality were highly correlated (Pearson's $r > .87$, $p < .001$) and had inflated variance inflation factors (VIF > 18), which exceeded standard thresholds. To address this issue, we created two parallel robust regression models for each outcome variable (tap rate and CV-ITI): one that included closeness and excluded betweenness and another that included betweenness and excluded closeness. This approach facilitated an independent evaluation of each centrality measure while maintaining statistical validity.

Model 1 — Interaction frequency (tap rate). Higher spatial anxiety was associated with a lower interaction frequency ($\beta = -0.175$, $p < .001$). Two metrics of urban form also showed significant negative effects: orientation entropy ($\beta = -0.071$, $p = .005$) and degree centrality ($\beta = -0.076$, $p = .002$). Circuity displayed a marginal negative effect ($\beta = -0.030$, $p = .071$). The centrality of the relationship, sense of direction (SBSOD), and gender were not significant predictors.

Model 2 — Interaction regularity (CV-ITI). Spatial anxiety was positively associated with inter-touch interval variability ($\beta = 0.140$, $p < .001$), indicating more irregular interactions. Higher SBSOD ($\beta = 0.051$, $p = .017$) and male gender ($\beta = 0.106$, $p = .030$) were also associated with greater variability. Of the urban form metrics, only the centrality of closeness was significant ($\beta = -0.079$, $p = .028$), with greater closeness linked to more regular interaction rhythms. Orientation entropy, circuity, and degree centrality were not significant.

Figure 4.10 presents the standardized coefficients of the robust regression models that predict interaction frequency (tap rate) and regularity (CV-ITI). The figure highlights which traits and urban form descriptors emerged as significant predictors in each model.

Figure 4.10.: *Standardized Regression Coefficients and 95% Confidence Intervals From the Robust Models Predicting Interaction Dynamics. Panel (a) shows predictors of interaction frequency (tap rate), where spatial anxiety, orientation entropy, and degree centrality were significant negative predictors. This indicates that fewer taps occur in more anxious individuals and in more complex or connected environments. Panel (b) shows predictors of interaction regularity (CV-ITI), where spatial anxiety, sense of direction, and male gender were linked to greater irregularity, while closeness centrality predicted more regular interactions. Gray bars represent 95% confidence intervals.*



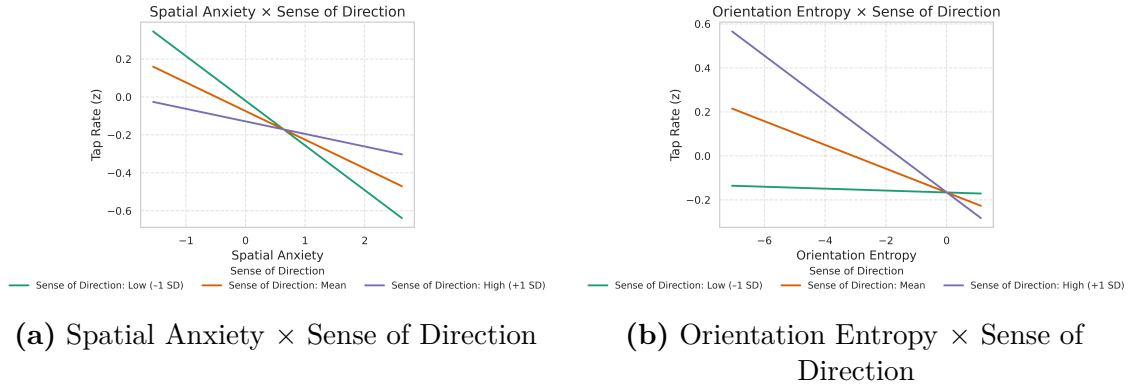
Based on the main effects identified in the robust regression models, we examined whether traits exert distinct influences on interaction frequency and regularity under various environmental conditions. Moderation models enabled us to test for context dependence by incorporating interaction terms between traits (SBSOD, SA, and gender) and urban form descriptors. This approach addresses the possibility that certain individual differences may become more or less influential depending on the spatial characteristics of the navigation environment.

Moderation Models. Five robust moderation models exhibited significant interactions (all predictors z-scored; two-sided tests).

- **Model 1 — SA × SBSOD (tap rate):** $\beta = 0.085, p < .001$
- **Model 2 — Orientation entropy × SBSOD (tap rate):** $\beta = -0.050, p = .008$
- **Model 3 — Circuity × SA (CV-ITI):** $\beta = 0.052, p = .009$
- **Model 4 — Closeness × SBSOD (CV-ITI):** $\beta = 0.064, p = .011$
- **Model 5 — Closeness × Gender (CV-ITI):** $\beta = 0.109, p = .037$

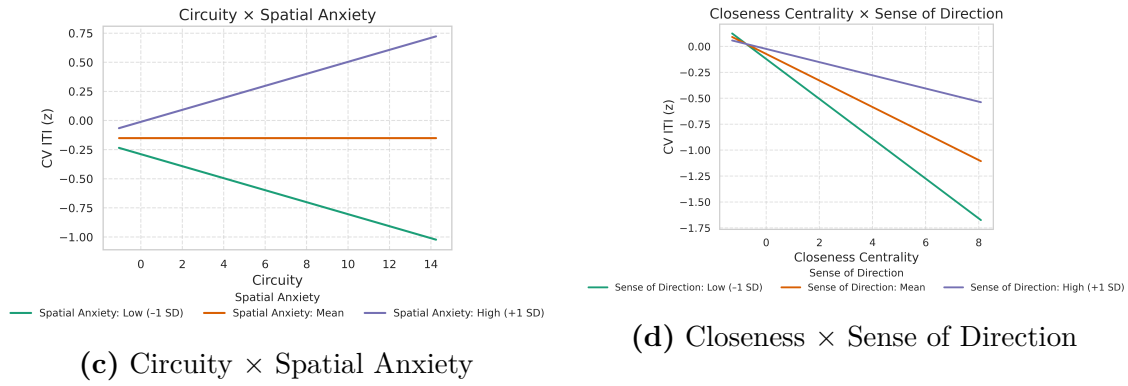
These models indicate that the effects of spatial traits and gender on interaction outcomes varied as a function of specific characteristics of urban form. Moderation models tested whether the strength or direction of the effects of traits on interaction metrics varies with environmental conditions (see Figures 4.11, 4.12). However, they did not address the possibility that environmental effects operate indirectly through intrinsic traits.

Figure 4.11.: *Moderation Effects on Tap Rate. Panel (a) shows that the influence of sense of direction on tap rate decreased with higher spatial anxiety. Panel (b) shows that sense of direction predicted tap rate more strongly in environments with higher orientation entropy (greater street network complexity).*



Note. Higher values on the vertical axis indicate greater tap rate (taps per second).

Figure 4.12.: *Moderation Effects on Interaction Regularity (CV-ITI). Panel (c) shows that spatial anxiety amplified the effect of network circuitry on interaction irregularity. Panel (d) shows that sense of direction was more strongly linked to CV-ITI in environments with higher closeness centrality.*



Note. CV-ITI = coefficient of variation of inter-touch intervals; higher values indicate more irregular interaction timing.

This study examined how personal traits and the urban environment combined influence interactions with mobile maps during navigation. Robust regressions revealed that spatial anxiety was the strongest and most consistent predictor of the of interaction dynamics. People with greater anxiety tended to tap less frequently and at less regular intervals than those with lower anxiety. Sense of direction and gender had less influence, and men exhibited slightly more irregular tapping than women. From an environmental perspective, the structure of the street network also mattered. More irregular layouts (higher orientation entropy) and more connected networks (higher degree centrality) were associated with less frequent tapping. In contrast, a higher centrality of proximity, indicating easier access to the surrounding streets, was associated with more regular tapping. When we tested interactions in the regression, we found that traits and the environment did not act in isolation, but rather reinforced one another. Sense of direction had a stronger influence when navigating in complex street networks, and spatial anxiety amplified the effect of network circuitry on irregular tapping. Gender differences in regularity also varied depending on network connectivity.

Together, these results demonstrate that the use of navigation apps is influenced by both the navigator and their location while using a mobile map application. Individual differences in anxiety and gender specifically interact with the structural properties of urban environments. This joint influence on how individuals interact with mobile maps led to the development of the prediction model, the detailed results of which we report in the following section.

4.3 Modeling Next Mobile Map Interaction with Context-Aware LSTM Fusion Model

The final part of this Results section addresses whether it is possible to **predict the next tap on the mobile map** by combining behavioral traces with intrinsic and extrinsic contextual information (RQ3). The task was to estimate the probability that the next tap during an ongoing phone session would occur in a Maps & Navigation (MN) or Travel & Local (TL) application. We treated this as a binary classification problem, coding taps in MN or TL apps as 1 and all other taps as 0.

To capture both **short-term dynamics** and individual-environment dimensions, we implemented a context-sensitive two-branch neural architecture in PyTorch. The architecture included three components: (i) a behavioral branch that modeled recent interaction sequences, (ii) a context branch that encoded the metrics and individual traits of urban form, and (iii) a fusion head that combined both sources of information into the final prediction (see Section 3.3.1).

In what follows, we present the results in three stages. First, we report the **overall predictive performance** of the selected model (Section 4.3.1). Second, we examine the contribution of the context branch, comparing the use of a compact 3D autoencoder embedding against the full eight-dimensional context vector (Section 4.3.2). Third, we analyze interpretability, identifying the contextual dimensions that most strongly influenced the predictions and how the latent dimensions of the autoencoder correspond to the original features (Section 4.3.3).

4.3.1 Overall Predictive Performance

To evaluate how well the augmented model we used (see 3.3.1 for details) could anticipate whether the next tap in a phone session would occur in a map application, we used *test set*, a portion of the data that had been completely set aside and never used during model training. This set contained $N = 5290$ windows of phone activity, of which 494 included a tap on a map app and 4796 did not.

Because a map tap is a relatively rare event, evaluating the model requires exploration beyond simply assessing overall accuracy. We therefore examined *precision* (the share of predicted map taps that were correct), *recall* (the share of actual map taps that the model successfully found), *F1 score* (the harmonic mean of precision and recall, balancing both concerns), and *ROC-AUC*. The ROC-AUC is the area under the receiver operating characteristic curve, which summarizes how well the model distinguishes between map and non-map taps across all possible thresholds. A value of 0.5 indicates the model’s performance is at the level of chance, while higher values mean the model is better at distinguishing map taps from non-map taps.

Table 4.15.: *Classification Threshold Analysis of the Test Set for the Augmented LSTM Model. The table shows how changing the decision threshold (p) shifts the balance between recall and precision. Lower thresholds capture almost all map taps (high recall) but produce many false alarms (low precision), while higher thresholds increase precision but miss more true events. The operating point chosen for this study was $p = 0.8$, which offered the best trade-off between sensitivity and specificity. Model configuration: time-step masking = 0.20, Gaussian noise $\sigma = 0.15$, feature dropout = 0.00, hidden size = 16.*

Threshold	Precision	Recall	F1 (pos)	ROC-AUC	Accuracy
0.5	0.106	0.951	0.191	0.743	0.249
0.6	0.112	0.925	0.200	0.743	0.308
0.7	0.129	0.838	0.223	0.743	0.455
0.8	0.216	0.567	0.313	0.743	0.767
0.9	0.441	0.389	0.413	0.743	0.897

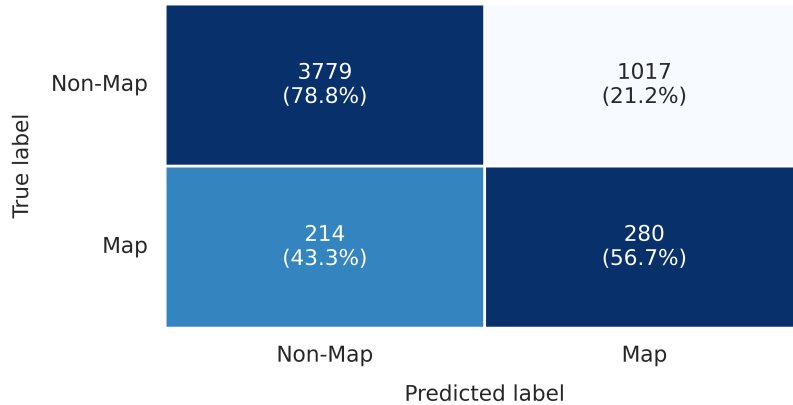
Note. ROC-AUC = area under the receiver operating characteristic curve; F1 (pos) = F1 score for the positive (map) class.

To understand how the choice of threshold affected performance, we systematically varied the *decision threshold* p between 0.5 and 0.9. The threshold is the cut-off probability above which the model labels the next tap as a map tap. Lower thresholds make the model more “eager,” catching more real events (higher recall) but also generating many false alarms (lower precision). Higher thresholds make the model more “cautious,” improving precision but missing many true map taps. Table 4.15 shows how the balance shifted between thresholds. For example, the recall was very high (0.95) at $p = 0.5$, but the precision was extremely low (0.11), which means that almost nine out of 10 predicted map taps were false alarms. At $p = 0.9$, precision improved (0.44) but recall decreased sharply (0.39), so most real map taps were missed.

The most balanced setting was found at $p = 0.8$, which maximized *Youden’s J statistic*

$= 0.355$. This statistic combines sensitivity (recall for the map class) and specificity (correctly rejecting non-map taps) into a single score. A value of 0 represents chance performance, while 1 reflects perfect balance. The threshold of 0.8 produced the highest J score, confirming it as the best trade-off between misses and false alarms.

Figure 4.13.: *Confusion Matrix for the Selected Augmented LSTM Model at $p = 0.8$. The figure shows that the model correctly identified most non-map taps (true negatives) and more than half of the real map taps (true positives), although many map taps were still missed (false negatives). Model configuration: time-step masking = 0.20, Gaussian noise $\sigma = 0.15$, feature dropout = 0.00, hidden size = 16.*



Note. TN = true negatives; FP = false positives; FN = false negatives; TP = true positives.

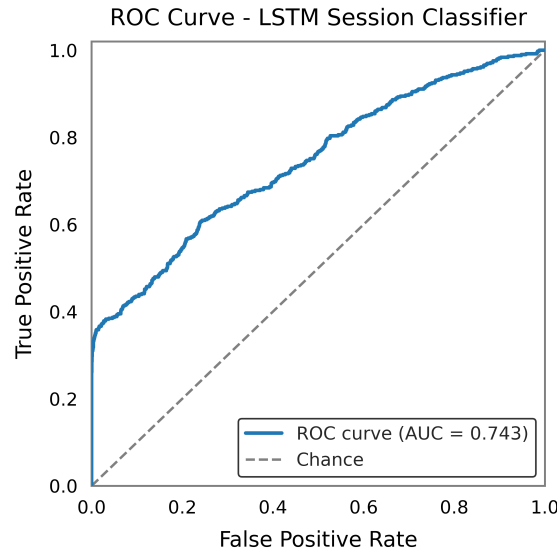
At this operating point, the model achieved an accuracy of 0.77 and a ROC-AUC of 0.74. More importantly, recall was 0.57, meaning the model successfully identified more than half of the true map taps. Precision was 0.22, meaning that approximately one in five predictions was correct. The corresponding F1 score was 0.31. The confusion matrix in Figure 4.13 illustrates these outcomes, showing 280 true positives (TP) and 3779 true negatives (TN) versus 214 false negatives (FN) and 1017 false positives (FP).

Finally, to evaluate the model’s discrimination capacity independent of any single threshold, we plotted the full ROC curve (Figure 4.14). This curve shows the trade-off between the true positive rate (recall) and the false positive rate at all thresholds. The area under this curve was 0.743, confirming that the model has a moderate ability to distinguish map from non-map taps, despite the imbalance in the dataset. This means that if we randomly take one map tap and one non-map tap, the model will assign the higher probability to the map tap about 74% of the time. In other words, the model is better than chance (50%) but still far from perfect separation, which is expected given that map taps are relatively rare in the dataset.

4.3.2 Contribution of the Context Branch: Autoencoder vs. Full Vector

After establishing the baseline predictive power of interaction sequences alone, we next sought to determine whether adding contextual information could improve the model. “Context” here refers to two stable factors: (i) *extrinsic features*, describing the street network in which the session took place (e.g., orientation entropy, circuitry, and centrality measures such as betweenness, closeness, and degree), and (ii) *intrinsic features*, including sense of direction (SBSOD score), spatial anxiety (SA score), and gender. We aimed

Figure 4.14.: Receiver Operating Characteristic (ROC) Curve for the Selected Augmented LSTM Model Evaluated on the Held-Out Test Set. The curve plots the true positive rate (TPR) against the false positive rate (FPR) across all classification thresholds. The area under the curve (AUC) is 0.743, showing that the model distinguishes map from non-map taps better than chance, although the performance is moderate. Model configuration: time-step masking = 0.20, Gaussian noise $\sigma = 0.15$, feature dropout = 0.00, hidden size = 16.



to ascertain whether combining this person–environment context with short-term tap sequences would help the model anticipate map taps more accurately.

We evaluated two ways of representing this eight-dimensional context vector. The first was a full standardized vector (*Without AE*), where each of the eight features was included directly after rescaling to the same range (mean = 0, SD = 1). The second was a compressed version learned using an *autoencoder (With AE)*. Since an autoencoder is a type of neural network trained to reconstruct its input data after having compressed it into a smaller latent space, we hypothesized that if it can recreate the inputs well, then the smaller embedding must capture the most important information while removing redundancy. In our case, the autoencoder reduced the eight input features to three latent dimensions (AE Latent 1, AE Latent 2, AE Latent 3).

In the test data, the autoencoder achieved a reconstruction loss of 0.084 on the training set and 0.087 in the development set. The low reconstruction error suggests that the 3D embedding preserved most of the variance in the original standardized features, enabling the effective reduction of dimensionality without substantial loss of information.

To assess whether the latent space preserved an interpretable structure, we examined Pearson’s correlations between each latent dimension and the original standardized features (Figure 4.15).

The three latent dimensions exhibited distinct and interpretable patterns of association with both urban form metrics and individual traits.

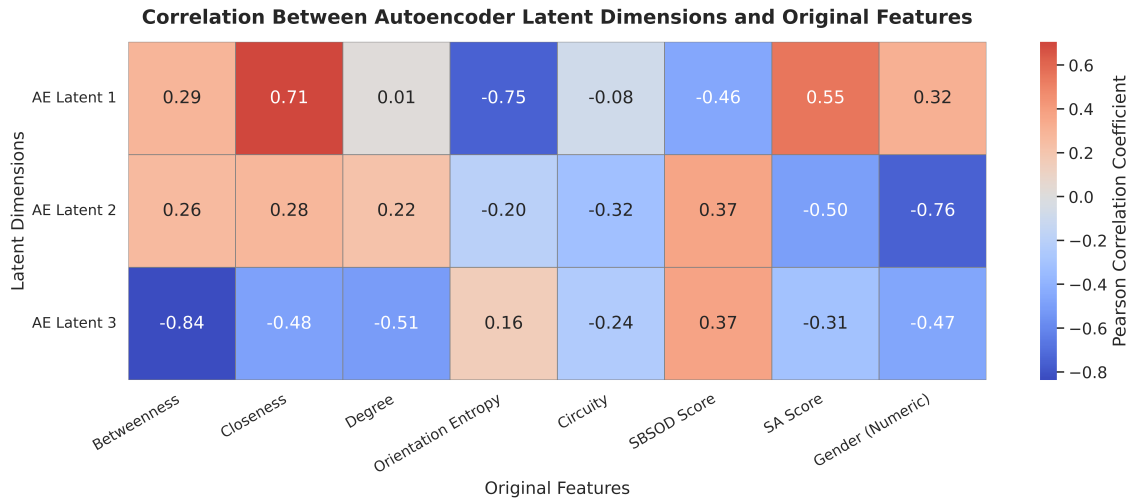
- **AE_latent_1** was most strongly and negatively correlated with *orientation entropy* ($r = -0.748$), and positively associated with *closeness* ($r = 0.705$) and *SA score*

($r = 0.553$);

- **AE_latent_2** was most strongly and negatively correlated with *gender* (numeric coding; $r = -0.758$) and *SA score* ($r = -0.499$), and positively with *SBSOD score* ($r = 0.370$);
- **AE_latent_3** was dominated by strong negative correlations with *betweenness* ($r = -0.837$) and *degree* ($r = -0.509$), alongside a moderate negative association with *closeness* ($r = -0.482$).

These results confirm that the autoencoder not only compresses the data for efficiency but also learns latent features with meaningful connections to both urban structure and individual traits.

Figure 4.15.: *Pearson Correlation Coefficients Between Autoencoder (AE) Latent Dimensions and the Original Contextual Features. The three latent dimensions showed distinct and interpretable patterns: AE latent 1 was negatively correlated with orientation entropy and positively correlated with closeness and spatial anxiety; AE latent 2 was negatively correlated with gender and spatial anxiety but positively associated with sense of direction; AE latent 3 was strongly negatively correlated with betweenness and degree, and moderately associated with closeness.*



Note. AE = autoencoder; SA = spatial anxiety; SBSOD = Santa Barbara Sense of Direction scale.

We then tested whether the reduced-dimensional AE contextual representation could match the predictive performance of the full vector when fused with the LSTM output. We compared the two fusion head configurations in the development set, assessing whether the AE’s dimensionality reduction, which is intended to minimize redundancy and improve efficiency, could match the predictive performance of the uncompressed contextual representation.

Table 4.16 summarizes the overall results of this comparison. In the development set, the full vector model (Without AE) achieved higher scores on all metrics: precision = 0.82, precision = 0.29, recall = 0.61, F1 = 0.39, and ROC-AUC = 0.80. The autoencoder model was weaker, with accuracy = 0.80, precision = 0.25, recall = 0.54, F1 = 0.34, and ROC-AUC = 0.75. These results suggest that compressing context into three dimensions came at a cost to predictive power.

Because overall metrics can mask class-specific differences, we also examined the results separately for map taps (rare positive events) and non-map taps (common negative events). Table 4.17 presents these per-class metrics. For map events, recall was almost identical in both configurations (0.39 with the full vector, 0.38 with the autoencoder). However, the full vector achieved higher precision, meaning that when it predicted a map tap, it was correct more often than not. Both models performed strongly for non-map events, with F1 scores exceeding 0.94 in each case.

Table 4.16.: *Overall Performance of the Fusion Head on the Development Set. The model using the full eight context features (Without AE) achieved higher accuracy, precision, recall, F1, and ROC-AUC than the autoencoder-compressed version (With AE), indicating that the uncompressed context provided stronger predictive power.*

Model	Accuracy	Precision	Recall	F1	ROC-AUC
With AE	0.800	0.246	0.542	0.338	0.753
Without AE	0.821	0.288	0.607	0.390	0.804

Note. AE = autoencoder; ROC-AUC = area under the receiver operating characteristic curve.

To illustrate error patterns more directly, Figure 4.16 depicts the confusion matrices for both configurations. On the development set, the full-vector model correctly identified 216 map taps and missed 140, while producing 535 false alarms and 2888 correct rejections of non-map taps. The autoencoder model detected fewer true map taps (193) and missed more (163), while also producing more false alarms (593) and slightly fewer correct rejections (2830). Expressed in proportions, the full-vector model caught approximately 61% of real map taps while flagging 16% of non-map taps incorrectly, whereas the autoencoder caught 54% of map taps but flagged 17% of non-map taps incorrectly.

Table 4.17.: *Per-Class Performance (Map vs. Non-Map) on the Development Set. The autoencoder version (With AE) showed higher precision for the map class, but at the cost of much lower recall compared to the full feature model (Without AE). For the non-map class, both models performed similarly well, with very high precision and recall.*

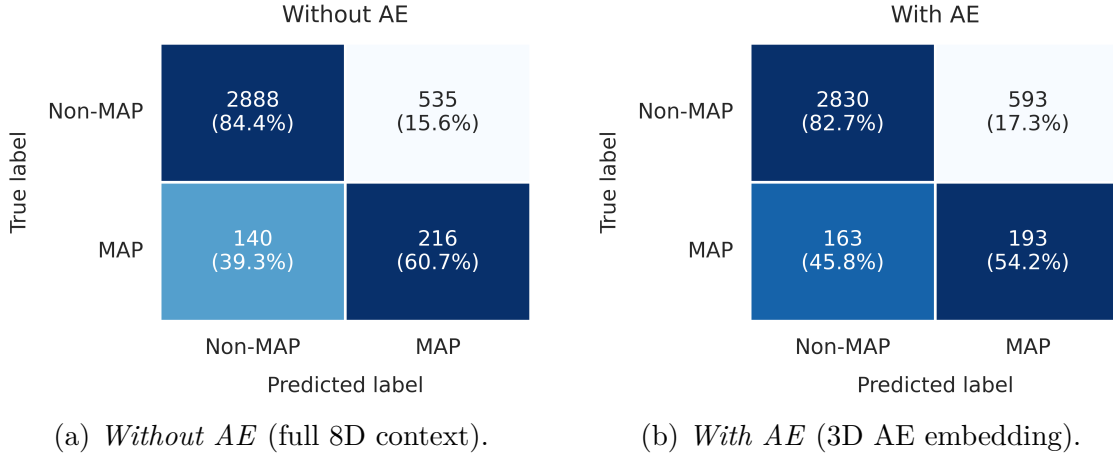
Model	Class	Accuracy	Precision	Recall	F1
With AE	MAP	0.934	0.811	0.385	0.522
	NON-MAP	0.934	0.939	0.991	0.964
Without AE	MAP	0.905	0.496	0.390	0.437
	NON-MAP	0.905	0.938	0.959	0.948

Note. Map = Maps & Navigation taps (positive class); Non-map = all other taps (negative class).

These patterns confirm what we saw in the overall and per-class metrics: both models perform well in abundant non-map events, but the full-vector representation achieves higher accuracy in rare map events and generates fewer false alarms. ROC and precision-recall curves for both configurations are provided in the project’s [GitLab repository](#).

We also assessed both versions on the test set that had been left out and evaluated them at the chosen operating threshold ($p = 0.8$). The model using all eight context features (Without AE) performed the best in all measures: accuracy 0.833, precision 0.356, recall 0.709, F1 0.474, and ROC-AUC 0.864. The autoencoder version was slightly weaker (accuracy 0.817, precision 0.320, recall 0.640, F1 0.427, and ROC-AUC 0.844). Confusion

Figure 4.16.: *Confusion Matrices on the Development Set for the Fusion Head Using (a) the Full Eight-Dimensional Standardized Context Vector (Without AE) and (b) the 3D Autoencoder Embedding (With AE). Counts are raw predictions; map is treated as the positive class. The With AE configuration produced fewer false positives for map predictions than the Without AE configuration, with comparable false negative counts.*



matrices confirm this pattern: the full-context model correctly identified 288 map taps (60.7%) with 118 missed, while the AE version identified 260 (54.2%) with 146 missed. Because the full-context model achieved the strongest performance, we treated it as our final model for prediction.

However, for the explainability step, we used the auto-encoder version. The AE compressed the eight contextual features into three broader factors, and our correlation analysis (see Figure 4.15) revealed that these factors combined urban and personal aspects in meaningful ways. This makes it easier to interpret how three factors influenced the model’s decision-making, rather than reporting SHAP values for eight distinct and partly overlapping features. The SHAP results for the eight raw features are provided in the Appendix Section A.9.

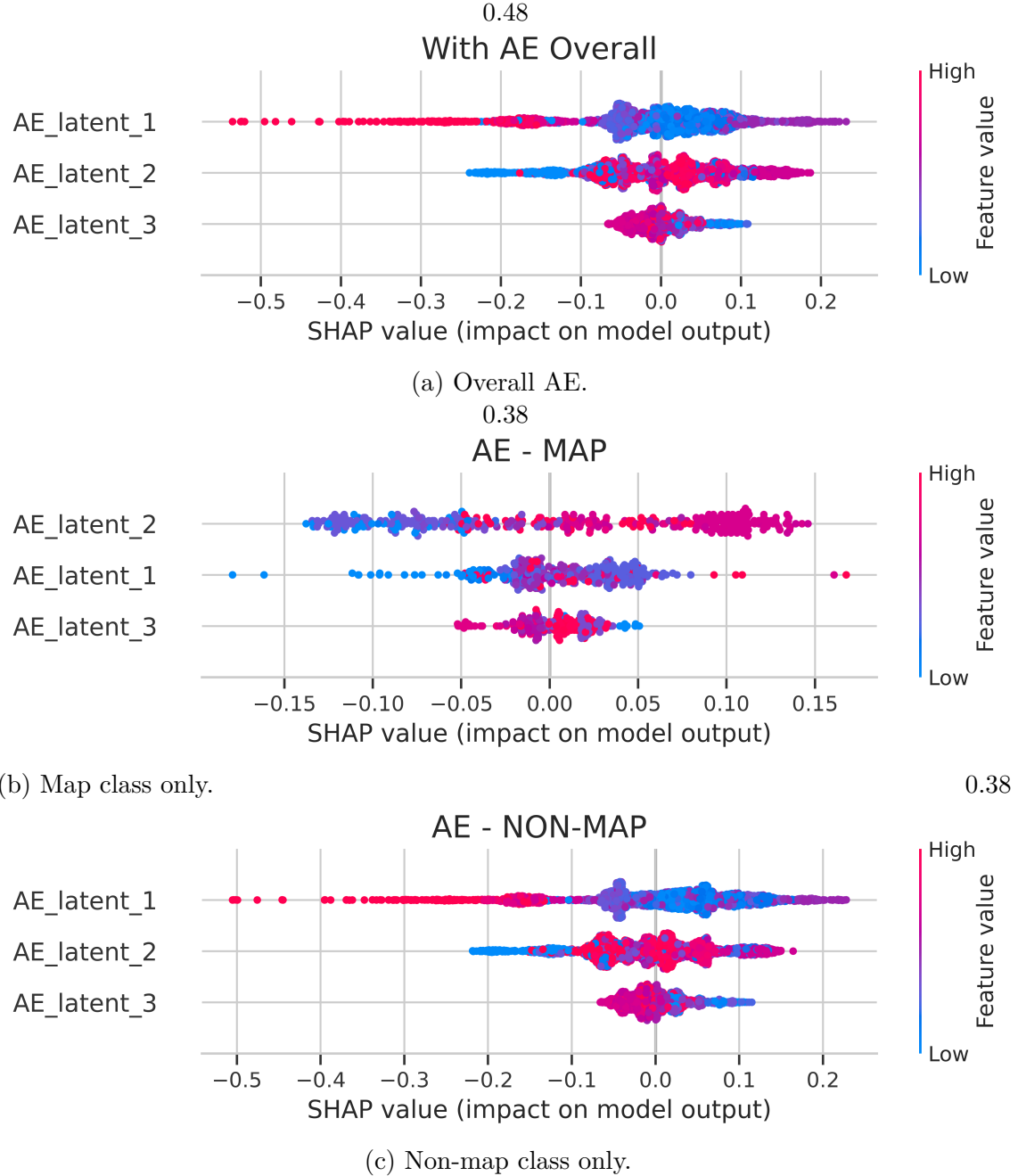
4.3.3 Interpretability of Context Features

Finally, we examined which contextual features were most influential in shaping the model predictions. This step is important because good predictive accuracy alone does not explain *why* the model makes its decisions. Interpretability helps connect model behavior with meaningful aspects of the environment and individual traits, which is crucial for improving mobile map navigation design.

We first examined feature contributions using SHAP values (SHapley Additive exPlanations). SHAP assigns each feature a score for every prediction: positive values push the prediction towards “map tap” while negative values push it towards “non-map tap” (see Section 3.3.1).

Figure 4.17 shows the SHAP *distribution plots* for the model with the autoencoder (With AE). These scores are plotted in a typical beeswarm plot, where each point is a data point arranged in order of decreasing value, with each representing one sample from the development set. The horizontal axis indicates the size and direction of the impact of the

Figure 4.17.: SHAP Summary Plots for Context Inputs in the Fusion Head (With AE configuration). The horizontal axis indicates SHAP values, where positive values increase the probability of predicting a map tap and negative values decrease it. Color encodes the feature value (blue = low, red = high). Features are ordered by average absolute SHAP value, with the most influential at the top. Panel (a) shows all samples, highlighting that AE latent 1 and AE latent 2 carried most of the predictive weight. Panels (b) and (c) show class-specific impacts: AE latent 2 was the strongest positive driver of map predictions, while AE latent 1 more often reinforced non-map predictions.

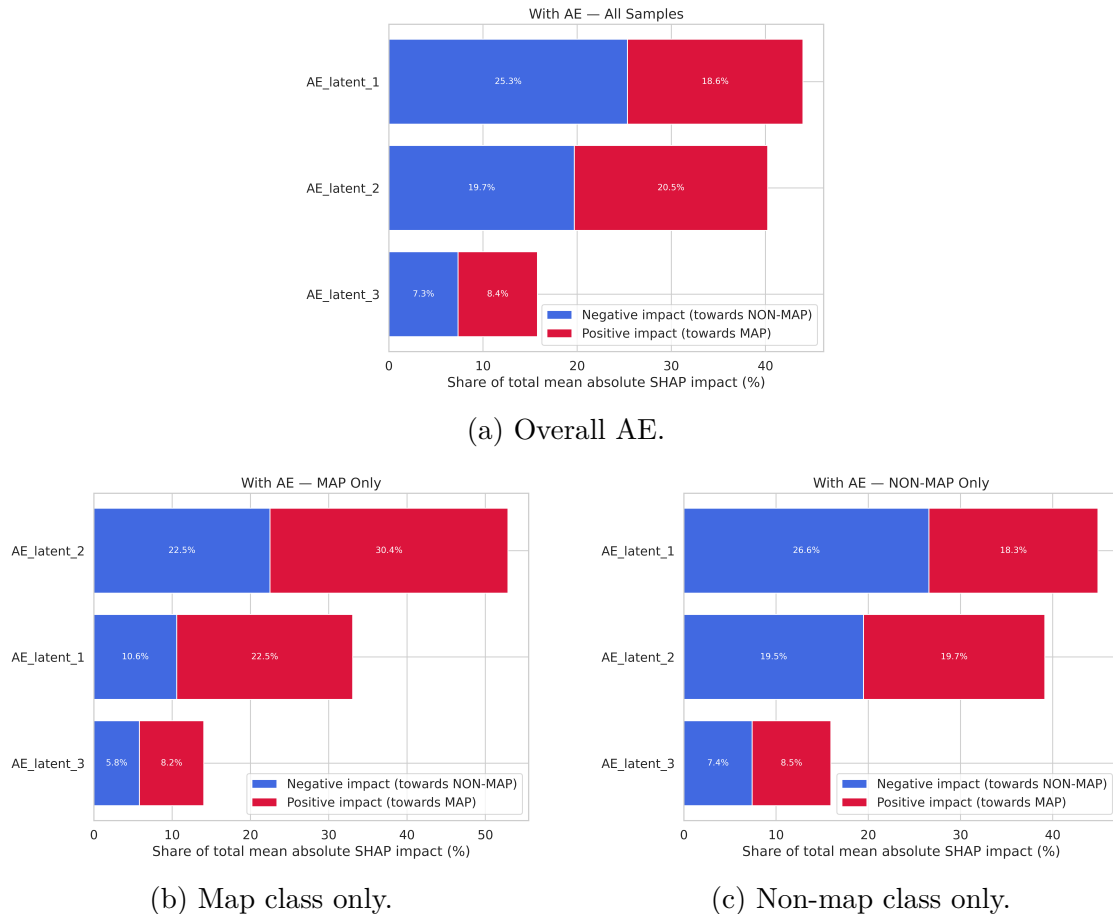


Note. SHAP = SHapley Additive exPlanations; AE = autoencoder; Map = Maps & Navigation taps (positive class); Non-map = all other taps (negative class).

feature, while the color shows whether the original value of the feature was low (blue) or high (red). The features are ordered by their average impact.

Two latent features dominated the overall analysis (Figure 4.17, Panel (a)): AE latent 1 and AE latent 2, which accounted for almost all of the contribution to the model’s predictions. In contrast, AE latent 3 had little effect. Panels (b–c) split the analysis by class: AE latent 2 contributed most strongly to the prediction of map events, whereas AE latent 1 more often reinforced the non-map predictions. These plots highlight the relative importance and direction of each latent feature in the samples.

Figure 4.18.: *Positive and Negative SHAP Contributions for the With AE Fusion Model on the Development Set. Bars show the share of the total SHAP impact, separated into positive contributions (towards predicting map) and negative contributions (towards predicting non-map). Panel (a) shows the aggregation of all samples, where AE latent 1 and AE latent 2 accounted for most of the model’s influence, while AE latent 3 played only a minor role. Panels (b) and (c) split the analysis by class: AE latent 2 was the strongest positive driver for map predictions, whereas AE latent 1 provided the largest negative evidence for non-map predictions.*



Note. SHAP = SHapley Additive exPlanations; AE = autoencoder; Map = Maps & Navigation taps (positive class); Non-map = all other taps (negative class).

To more closely examine the contribution of each latent variable, we dissect the SHAP values into positive and negative components. This shows, in percentage terms, how much

each latent variable was responsible for pushing predictions one way or the other (see Figure 4.18).

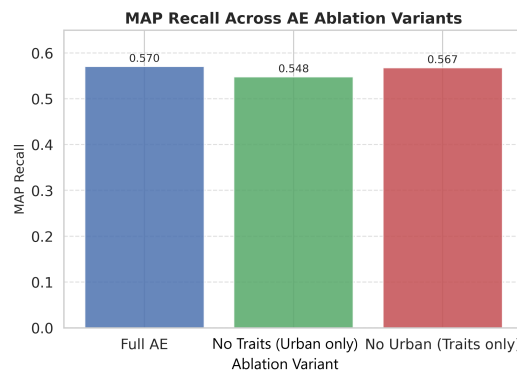
In the **overall analysis** (panel a), *AE latent 1* emerged as the main driver of **non-map predictions**, accounting for 25.3 % of the total negative contributions and 18.6 % of the total positive ones. *AE latent 2* played a **similarly important but more balanced role**, contributing 20.5% positive and 19.7% negative, while *AE latent 3* had **only a minor influence**, never exceeding 8.4% in either direction.

For the **map** (panel b), *AE latent 2* provided the **strongest positive support for predicting map taps**, representing 30.4%, and also made a negative contribution of 22.5 %, pushing predictions away from map taps. *AE latent 1* was the second most important latent, with 22.5% positive and 15.9% negative contributions, and *AE latent 3* remained below 10% in both directions.

For the **non-map-only subset** (panel c), *AE latent 1* was the **strongest driver of non-map predictions**, contributing 26.6 % of the negative influence, while also adding 18.3% positive. *AE latent 2* played a more balanced role with 19.7% positive and 19.5% negative, and *AE latent 3* remained **largely irrelevant**, contributing less than 9 % in either direction.

Finally, we asked whether urban form or individual traits had more influence on predictive performance. To test this, we ran an *ablation analysis*, in which parts of the contextual input were removed to observe how the performance changes.

Figure 4.19.: *Map Recall Across the Three AE Ablation Variants (Full AE, No Traits: Urban only, and No Urban: Traits only) on the Development Set. The figure shows that recall was highest for the Full AE (0.570) and Traits-only (0.567) models, and slightly lower for the Urban-only model (0.548), suggesting that traits were particularly important for sustaining recall of map events. Values above the bars correspond to recall.*



Note. AE = autoencoder; Map = Maps & Navigation taps (positive class).

Three versions of the model were compared in the development set: (i) **Full AE**, which included all features; (ii) **Urban-only model**, where trait features were removed; and (iii) **Traits-only model**, where urban metrics were removed.

All three versions were very good at classifying the frequent non-map taps (precision of about 95%). The differences appeared in the rarer map taps. Recall was highest when all features were included (0.57) or when only traits were included (0.57). The urban-only model achieved a slightly lower recall (0.55) but had the highest overall accuracy (0.84).

Figure 4.19 illustrates these results more clearly. The recall was strongest for the full model and the model with traits only, and weakest for the model with urban only. Although the differences are small, the pattern suggests that removing trait features diminishes recall slightly more than removing urban features.

Collectively, the results show that the LSTM model alone’s prediction of map taps is above chance, reaching 77% accuracy and an ROC–AUC of 0.74 on the test set. At the chosen threshold ($p = 0.8$), it detected 57% of real map events, although the precision was low (22%), reflecting the strong class imbalance. Adding context improved this result. In particular, the version that used all eight features outperformed the auto-encoder version in both development and test sets. In the test set, the model achieved a precision of 0.833, a recall of 0.71, an F1 score of 0.47, and an ROC–AUC of 0.86, with 36 % precision in detecting real map events. For explainability, we used the autoencoder version, which compressed the eight context features into three broader factors. The SHAP analysis demonstrated which factors had the greatest influence on the model, highlighting that signals linked to street network complexity and anxiety often pushed predictions toward non-map predictions, while sense of direction and gender supported map predictions.

5

CHAP-
TER

Discussion

*Vladimir: I don't understand.
Estragon: Use your intelligence,
can't you?
Vladimir uses his intelligence.
Vladimir: (finally) I remain in
the dark.
Samuel Beckett, **Waiting for
Godot***

The primary goal of this thesis was to develop a computational model capable of predicting the next tap a user will make on a mobile map application. This model was based on the integrated framework of three crucial contextual elements: external (environmental factors), internal (cognitive, emotional and motivational traits; e.g., spatial anxiety, sense of direction, cognitive load), and behavioral (interaction patterns that can be observed in real time, including touches, pauses, and recurring gestures). With reference to the related work addressed in Chapter 2, this thesis critically discusses the main results of this research (presented in Section 4) and address its limitations. This chapter is organized according to my three main research questions, outlined in Section 1.4, which are also included below as a reminder:

- **RQ1**: How can mobile map interaction data be collected in situ to capture everyday navigation behavior in real-world contexts?
- **RQ2**: How can mobile map interactions be meaningfully linked to individual differences and spatial context?
- **RQ3**: How can a user's individual traits, prior usage behavior, and spatial context be used to predict their future interactions with mobile map applications during navigation?

5.1 Achieving Ecological Validity in Mobile Map Research

Investigating individuals' everyday use of mobile maps requires methods that accurately reflect the real-world process of navigation. However, controlled studies conducted indoors using VR or outdoors using mobile EEG and ET tend to oversimplify the complex multitasking and sensory experiences of daily navigation and may also influence behavior

due to the intensive equipment involved (Cheng et al., 2022; Dahmani & Bohbot, 2020; Hilton et al., 2023; Kapaj et al., 2024). Additionally, large-scale questionnaires, even if easily administered, rely on participant recall, offer only limited temporal precision, and are susceptible to desirability bias (Ishikawa, 2019; Ruginski et al., 2019, 2022; Topete et al., 2024). Neither approach effectively captures the real-world dynamics of in-the-moment navigation using smartphones, which is often sporadic, brief, and intricately linked to everyday phone activities (Deng et al., 2019; Oulasvirta et al., 2012).

Smartphone sensing provides a systematic method for addressing this issue by recording frequent in situ measurements that seamlessly integrate into daily routines, aligning with the principles of EMA (G. Miller, 2012; Montag & Baumeister, 2023; Shiffman et al., 2008). However, ensuring ecological validity should not compromise trustworthiness. The collection of specific data types, such as screen content, text input, or multimedia, can prompt privacy concerns and decrease the likelihood of participant engagement or compliance (Huckvale et al., 2019; Martinez-Martin et al., 2018). Our strategy for tackling this problem involves adhering to the principle of parsimony (Burnham & Anderson, 2002; Gori et al., 2024), emphasizing the collection of a minimal set of signals that are sufficient to yield meaningful insights, while intentionally avoiding non-essential or sensitive data. Tappigraphy exemplifies this approach by recording only touch timing and the active application to provide a subtle and inconspicuous documentation of interaction patterns. Within the domain of navigation research, tappigraphy offers a behavioral window for studying map use, without the need to record content or private input. However, as noted by Reichenbacher et al. (2022), this simplified approach comes with minimal experimental control and provides limited understanding of the user themselves and their environment. To tackle this limitation, our methodology incrementally incorporated two different contextual elements into our data collection on a voluntary basis: GPS traces to place interactions within their geographical context, and one-time surveys to assess persistent individual characteristics such as sense of direction and spatial anxiety (He & Hegarty, 2020; Hegarty et al., 2006).

5.1.1 Effect of Minimal Enrichment: Evidence from MoT I and MoT II Data Collection

The first MapOnTap campaign (MoT I) adopted a strictly minimalist design (similar to that of (Reichenbacher et al., 2022)), recording tappigraphy and optional GPS logs from 53 participants. Compliance was high, with participants contributing to data collection for durations spanning from 1.9 hours to 868 days. GPS data collection, which was an optional aspect of participation, was provided by 42 out of 53 participants for an average of 103 days. This demonstrated that minimally invasive collection can easily be achieved over extended time frames. However, mobile map-specific data were sparse. In MoT I, map-only sessions represented only $\sim 3\%$ of all phone sessions, and map taps represented, on average, 3.7% of all taps (median of 1.4%). The use of the app also highlighted heterogeneity; while Google Maps was used by almost all participants, other navigation-related apps varied according to personal habits and travel interests. Finally, we reported Google Play Store categories such as Maps & Navigation and Travel & Local to enhance our dataset. Although these labels are designed to help users find applications, they are not based on consistent principles. Developers assign categories when publishing their apps, which can lead to inconsistency. Consequently, the Maps & Navigation category includes apps with little or no navigation relevance, such as atlases, weather overlays, and real estate browsers.

Meanwhile, many apps that incorporate navigation as a core function, such as ride-hailing, delivery, and travel platforms, are classified under different headings. This problem on two sides means that app store categories give a distorted picture of how navigation tools are actually distributed. In MoT I, this created a critical limitation: navigation use appeared to be fragmented across categories, but the dataset lacked the contextual resolution to capture how individual differences shaped these diverse navigation practices.

The exploratory analyses of the MoT dataset (Zingaro & Reichenbacher, 2022; Zingaro et al., 2023) offered limited insights, as they mainly revealed broad behavioral patterns, such as the distinction between the use of local and travel-related maps. These observations are merely descriptive and lack explanatory power concerning the underlying causes of map usage, especially in the absence of context about user activities or situations. Simply put, these initial MoT I studies highlighted the inadequacy of aggregate measures in understanding everyday navigation behavior. However, given the deliberate choice to use tappigraphy as the data collection technique, any content related to phone activity besides temporal information was removed.

To overcome this limitation while retaining a methodologically minimalist approach, MoT II sought to enhance tappigraphy data by incorporating supplementary user information. Sociodemographic factors, such as gender, provided a basic structure for identifying differences in interaction patterns. The Santa Barbara Sense of Direction scale (SBSOD) offered information about participants' self-reported ability to orient and navigate, serving as an indicator of spatial competence that could shape their reliance on or avoidance of GPS guidance. Spatial anxiety captured the degree of stress and worry participants experienced during navigation tasks, which has been shown to heighten dependence on external aids and reduce allocentric processing. Including these measures enabled us to assess not only differences in mobile map usage, but also the extent to which such differences were linked to underlying confidence, navigational abilities, or anxiety. To protect data integrity while safeguarding privacy, we limited the data collection period for each participant to two weeks. Although participants had the option to contribute more data, our analyses revealed that their actual participation tended to be shorter than in the first MoT study, with recording durations ranging from 0.01 up to 33.6 days.

For post hoc analyses of MoT II data, we made a deliberate choice to retain only information that was strictly related to navigation. The analysis focused on 26 navigation-centric applications, defined a priori, to overcome the arbitrary category assignments of the Google Play Store. These included classic map-based tools that provide GPS positioning, routing, and turn-by-turn guidance (e.g., Google Maps, Petal Maps, OsmAnd, Maps.me), as well as applications closely tied to mobility tasks such as public transport planners (SBB, Trenitalia, DBB). For these apps, we retained information on how often and how regularly they were used, how users switched between them, and how they were embedded across phone sessions. By narrowing the scope to applications that directly support wayfinding and mobility, we ensured that the behavioral signals we analyzed reflected genuine navigation demands rather than peripheral activities such as communication or media use. This decision allowed us to interpret tapping dynamics and contextual features more precisely as indicators of cognitive engagement with spatial navigation. Note that, with this filtered set, sessions in which participants only used navigation apps (i.e., sessions that involved apps in the maps categories only) accounted for 18% of all the filtered phone sessions on average and represented just 2–3% of the total time spent using the phone (similar to MoT I).

These enriched datasets facilitated the exploratory analysis of stable individual differences (Zingaro et al., 2023). The findings exhibited gender- and trait-specific interaction tendencies: men with higher SBSOD typically participated in longer map sessions, while women with increased spatial anxiety frequently switched between apps. These outcomes suggest foundational behavioral styles rather than the previously assumed compensatory strategies in gender differences (He & Hegarty, 2020; Lawton, 1994). They correlate with HCI research on fragmented smartphone use (Deng et al., 2019; Oulasvirta et al., 2012) and recent studies that indicate that temporal touch dynamics can serve as behavioral indicators for cognitive faculties, such as visuo-spatial working memory (Ceolini et al., 2022).

5.1.2 Implications

A key methodological implication of these findings is that ecological validity in mobile navigation research does not require heavy instrumentation or intrusive tracking. Instead, it can be achieved by carefully balancing parsimony with minimal enrichment. In our case, this meant combining tappigraphy with voluntarily provided GPS logs and short validated questionnaires that capture intrinsic traits such as sense of direction and spatial anxiety. This combination substantially increases ecological validity while minimizing the burden on participants and safeguarding their privacy. This method presents a significant alternative to conventional approaches for investigating GPS-assisted navigation. It enhances ecological validity beyond that of laboratory research or surveys and minimizes the intrusion typical of outdoor experiments that employ bulky equipment such as ETs, mobile EEGs, or continuous video logging. While these conventional tools provide precise control or detailed sensory data, they can disrupt natural behavior, are not easily scalable, and pose ethical or logistical challenges. In contrast, minimally enriched tappigraphy maintains the non-intrusive nature of digital trace data. It facilitates the connection between behavior and cognitive and emotional traits, as well as environmental structures. Thus, tappigraphy complements rather than replaces traditional methods by offering a scalable and privacy-respecting means to study daily navigation in real-world settings, particularly focusing on reliance on mobile map usage.

5.1.3 Limitations

Despite the methodological and theoretical robustness of the proposed approach, this study has several limitations that must be acknowledged. First, the functional classification of applications based on app store labels (“Maps and Navigation”, “Travel and Location”) was inadequate for capturing their actual use of mapping functionalities. The absence of a taxonomy more closely aligned with real-world use contexts required post hoc filtering to identify meaningful navigation sessions. However, there remains a risk of including geospatially contextualized interactions that are not strictly cartographic in nature.

Second, although tappigraphy provides high-resolution temporal data in ecologically valid settings, it offers limited insight into the underlying cognitive processes of users. Without additional measures, such as behavioral annotations or cognitive tasks, some interaction strategies may be ambiguous or difficult to interpret precisely.

Another limitation is the low saturation of mobile map usage in the dataset; map applications accounted for only around 3% of total smartphone usage. This reduces the number of

relevant events for certain participants and may limit the statistical power of some analyses. Additionally, the relatively short data collection period (two weeks in MoT II) provides a useful snapshot but may not adequately reflect seasonal variations or exceptional situations (e.g., travel or relocation).

Finally, while the emphasis on ecological validity is an advantage of the study, it comes at the cost of reduced control over experimental variables, which limits the ability to draw strong causal inferences. Nonetheless, these limitations do not undermine the validity of the overall approach. Rather, they highlight opportunities for future methodological refinement, such as integrating multimodal data streams or incorporating brief cognitive assessments to enhance the interpretability of observed interaction patterns.

The effective demonstration of tappigraphy, combined with minimal extrinsic (e.g., GPS) and intrinsic (e.g., brief spatial trait questionnaires) contextual layers, fulfills the core methodological ambition of RQ1. This careful balance of parsimony and ecological enrichment is essential for subsequent phases of analysis. The resulting methodological pipeline is, therefore, fully justified. The collection of minimal yet sufficient data supports the translation of raw behavioral traces into theoretically meaningful indicators for performing integrated analyses of intrinsic, extrinsic, and behavioral contexts.

5.2 Mobile Map Interactions are Jointly Shaped by Who Navigates and Where They Navigate

This research specifically responds to the long-standing call for a comprehensive understanding of intrinsic characteristics, the external environment, and real-time interaction data in the context of investigating the effects of mobile map reliance on users and highlighting the need for user-adaptive map systems (Bartling et al., 2022; Fabrikant, 2023; Griffin et al., 2017; Huang et al., 2018; Reichenbacher, 2005; Roth et al., 2023; Ruginski et al., 2022). Earlier studies primarily explored either the relationship between cognition and urban structure (Coutrot et al., 2022; Watts et al., 2015; Xu et al., 2022), the connection between traits and cognitive processes (He & Hegarty, 2020; Hegarty et al., 2006; Lyons et al., 2018), or the dependence on GPS and cognitive function (Dahmani & Bohbot, 2020; Ruginski et al., 2022) in isolation.

5.2.1 Across-Session Mobile Map Interaction Metrics: Stability Over Time

These investigations first required an understanding of the features that could be meaningfully used and extracted for the three levels. For interaction behavior, we extracted across-session measures, such as the inter-map interval (IMI) and the session initiation rate (SIR), at the level of aggregated navigation rhythms. These indices exhibited weak and inconsistent correlations with the sense of direction and spatial anxiety. A borderline negative link between SIR and SBSOD disappeared when controlling for gender, suggesting that short study windows do not allow sufficient time for usage rhythms to stabilize and that coarse person-level summaries cannot capture consistent trait effects. However, neuroscientific studies that leverage tappigraphy (Ceolini & Ghosh, 2023; Ceolini et al., 2022) have demonstrated that general phone inter-touch intervals (ITIs), similar in terms of computation to our focused map-only IMI, can reflect fluctuations in visuospatial

working memory. However, those effects required dense, long-term tappigraphy data to emerge reliably. Similarly, navigation-specific intervals, such as IMI and SIR, may contain meaningful signals; however, to capture stable individual tendencies, they require longer collection periods than our two-week study.

5.2.2 Effects of Individual Traits on Mobile Map Interactions

We then moved to exploring within-session features, extracting data on regularity (CV-ITI), frequency of interaction (taps per second), and switching behavior (transitions per minute). Together, these metrics characterize how actively and consistently participants interacted with navigation apps during a session. To move beyond isolated measures, we reduced these features into a shared behavioral space that highlighted similarities between sessions. We also grouped sessions into recurring engagement styles, revealing six such styles: a dominant cluster that represented the majority of low-intensity, routine sessions, and five smaller clusters that represented more specific patterns, such as heightened tapping intensity, greater irregularity, or frequent switching. We then examined whether the distribution of sessions across these styles could account for differences in sense of direction (SBSOD) or spatial anxiety (SA). We calculated the proportion of each participant's sessions in each group and related these distributions to trait scores. However, the analyses did not reveal consistent associations: neither SBSOD nor SA reliably predicted how participants' sessions were distributed across engagement styles. We then queried whether the distribution of sessions across these groups could explain the stable differences in sense of direction (SBSOD) or spatial anxiety (SA). We calculated the proportion of each participant's sessions that fell into each cluster and regressed these proportions against trait scores. The analyses did not reveal consistent links: neither the high- nor low-trait groups showed a clear preference for particular interaction styles. The lack of stable trait-cluster associations probably reflects both methodological and conceptual factors. Methodologically, our two-week observation windows provided too few navigation sessions per person to robustly estimate individual cluster proportions; with sparse data, distributions may fluctuate more with situational context than with stable dispositions. Conceptually, session-level engagement styles may reflect immediate situational demands (e.g., trip length, multitasking, momentary uncertainty) more than enduring traits. As a result, tapping patterns alone did not strongly differentiate among individuals. Stronger effects only emerged once the environmental structure was added to the models, confirming that behavior in situ is co-shaped by both traits and context rather than by traits in isolation.

5.2.3 Effects of Urban Forms and Individual Traits on Mobile Map Interactions

While the connection between traits and behavior did not exhibit clear patterns, when linking within-phone session dynamics to urban form, we found consistent patterns. Spatial anxiety consistently predicted a reduced tap frequency and more irregular rhythms, indicating a passive reassurance-driven reliance on GPS guidance (He & Hegarty, 2020; Lyons et al., 2018). Beyond these individual traits, features of the urban form shaped interactions in systematic ways. Orientation entropy, which captures the diversity of street orientations, thus indicating the irregularity and characterizing the disorder of a city's layout, predicted fewer touches. This suggests that in visually and structurally complex

environments, users offload demands onto GPS rather than actively engaging with the map, consistent with theories of cognitive load and technological offloading (Montello, 2009; Parush et al., 2007). Degree centrality, which measured how many street segments connect at a given intersection, was likewise associated with reduced engagement. Intersections with many connecting streets present more branching options and increase the complexity of local decisions, which can encourage navigators to rely on automated guidance instead of interacting with the map.

Conversely, closeness centrality, which reflects the accessibility of a location within the wider street network, predicted more regular tapping rhythms. Environments that are more accessible are easier to traverse and provide stronger orientation cues, so users interact with the map in a steadier and less fragmented way, in line with Lynch’s (1960) theory of urban legibility. Finally, circuitry, which expresses the degree to which routes are indirect or tortuous relative to the straight-line distance, amplified the disruptive effect of spatial anxiety on tapping regularity. Indirect street structures impose heavier sequential planning demands, and in such environments, anxious navigators display more fragmented, irregular interaction patterns.

Moderation analyses confirmed that these environmental influences are not uniform but are filtered through individual traits. Low-SBSOD users disengaged more strongly in high-entropy environments, while high-SBSOD users maintained stable engagement. Gender differences in tapping irregularity varied with accessibility, consistent with divergent strategy preferences rather than uniform effects (Lawton, 1994). Combined, these findings demonstrate that navigation behavior is not determined solely by traits, environment, or technology, but by their dynamic interplay in situ.

5.2.4 Implications

Our real-world findings expand upon laboratory-based observations by exploring in-the-wild user behavior. This provides ecological validation of the ways in which the urban structure shapes digital navigation patterns. Urban environments with irregular layouts and high global spatial disorder elicited fewer and more cautious interactions. In contrast, highly accessible areas fostered smoother and more regular rhythms of engagement. These session-level interaction dynamics offer concrete, empirically grounded evidence that urban complexity imposes measurable cognitive constraints, which are clearly reflected in users’ tapping behavior.

From an applied perspective, these insights are critical for designing adaptive, context-aware navigation systems that can address the challenges of urban environments. Such systems must be capable of responding dynamically to varying cognitive loads imposed by diverse spatial configurations and adjusting guidance delivery accordingly. Our analyses clearly demonstrate that characteristics of urban form, including orientation entropy, circuitry, and centrality, significantly influence both real-time interaction patterns with GPS-enabled mobile map applications and individual traits, validating the notion of integrating these elements into mobile map design.

5.2.5 Limitations

While significant progress has been made, certain limitations remain notable. Participant-related features, such as habitual navigation patterns, demand extended observation windows (e.g., several weeks or months) for reliable assessment. Although the two-week span of MoT II was sufficient to effectively identify patterns, it may not fully capture long-term behavior consistency. Additionally, the assessment was selective, focusing on SBSOD, spatial anxiety, and gender, without addressing other spatial abilities such as mental rotation, perspective-taking, and landmark memory. This restricts the comprehensiveness of the cognitive profile and may overlook pertinent moderating effects. Another issue is the complexity associated with the use of multiple applications. Despite rigorous filtering, participants often switched between applications within sessions, complicating the attribution of behaviors solely to navigation tasks. Similarly, the absence of stable trait-style connections to session-level measures likely stems from the limited two-week observation period, which resulted in insufficient navigation sessions per participant to reliably gauge individual cluster proportions. Sparse data may mean that cluster distributions are more influenced by situational circumstances than by stable traits. At a conceptual level, engagement styles during sessions might be shaped by immediate scenarios, including trip length, multitasking, or momentary uncertainty, rather than enduring personal traits. Consequently, tapping patterns alone did not adequately distinguish individuals from one another.

More pronounced associations surfaced when considering environmental structure, emphasizing that navigation behaviors are jointly influenced by traits and situational contexts rather than traits alone. These constraints suggest opportunities for methodological improvements, such as extended participant data collection periods, a broadened range of cognitive assessments, and more detailed logging of application contexts.

However, these limitations do not detract from the dataset’s ecological validity or the robust integrative framework introduced here. The results form an analytical foundation for predictive modeling by transforming raw, ecologically valid tappigraphy and GPS data into interpretable phone session metrics, such as tap rate, CV-ITI, and transitions per minute, directly linking these metrics with intrinsic traits (such as spatial anxiety, sense of direction, and gender) and external spatial context (e.g., urban form metrics). Despite the fact that the predictive models did not directly employ these metrics (RQ3), they were essential for characterizing navigation behavior and illustrating its sensitivity to cognitive-emotional traits and environmental conditions. These metrics captured the intensity, rhythm, and switching behavior of map interactions, serving as a tangible behavioral connection between raw interaction logs and higher-level theoretical concepts.

This intermediate step also guided the design of RQ3’s feature engineering approach. By understanding the variability and associations of these RQ2 metrics, we decided to compute differenced values that were aligned with the overall timeline of phone sessions. This method highlights meaningful short-term changes in interaction behavior rather than static levels. These differenced time-series features, combined with static user traits and urban form descriptors, represent inputs for an LSTM-based neural architecture designed to learn temporal dependencies and context interactions.

5.3 Feasibility of Explainable and Enriched Next Mobile Map Interaction Prediction

Most predictive elements of navigation systems are primarily concentrated on external efficiency indicators such as ETA, route optimization, and traffic avoidance (Cui et al., 2019; Derrow-Pinion et al., 2021; Herrera et al., 2010). While these models perform effectively on a large scale, they fail to consider how **implicit signals from routine phone usage**, particularly the way individuals engage with various applications, could act as **proxy for spatial decision making and navigational needs**. Past initiatives for developing adaptive maps have persistently highlighted that improving mobile map design and making it more user-centric requires the incorporation of user-specific traits in addition to the contextual environment (Ballatore & Bertolotto, 2015; Fabrikant, 2023; Griffin et al., 2017; Montello, 2018; Reichenbacher, 2004a). However, such recommendations remain underutilized in commercial systems, which consistently handle map prediction only as it relates to destinations or routes.

5.3.1 Person–Place Context in Mobile Map Interaction Prediction

This methodology reconceptualized prediction as a behavioral task. Rather than projecting user destinations, the model treated map engagement as a temporally organized stream of taps, aiming to forecast how users interact with navigation applications in real time. This was done by combining recent tap sequences with a context vector that blended intrinsic traits and extrinsic environmental features. Intrinsic traits included the Santa Barbara Sense of Direction (SBSOD), which captures an individual’s confidence in their ability to orient and form spatial representations (Hegarty et al., 2002, 2006), and is typically associated with a reduced dependence on using mobile map applications for navigation (He & Hegarty, 2020), as well as spatial anxiety (SA), which reflects stress or discomfort during navigation tasks (He & Hegarty, 2020; Lawton, 1994; Lyons et al., 2018) and is associated with increased reliance on external aids (He & Hegarty, 2020). Gender was also included, given the documented sex differences in strategy preferences (Kozlowski & Bryant, 1977; Lawton & Kallai, 2002). Extrinsic features characterized the surrounding street network structure for each session: orientation entropy quantified the level of order or disorder in street orientations; circuitry measured the indirectness of routes compared to direct distances; closeness centrality evaluated overall accessibility within the network; degree represented the number of possible directions to be taken at intersections; and betweenness indicated the frequency with which streets were part of the shortest paths. Collectively, these features enabled the model’s consideration of both the navigator’s cognitive-emotional tendencies and the environmental demands.

The objective was not to create a new standard for efficiency but to establish an intermediate behavioral layer: a predictive structure that converts low-level tap data into markers of interaction style, correlating them with navigator identity and the navigation context. By conceptualizing tapping as a structured temporal sequence and modeling the likelihood of subsequent interactions, the system captures engagement patterns influenced by both user characteristics and urban environments. This methodology extends research on next app prediction and context-aware sequence modeling (Greff et al., 2017; Sutskever et al., 2014; Wang et al., 2021) but shifts the emphasis from forecasting the next application to be launched to analyzing the use of navigation apps once they are in operation. In this regard,

it repurposes the app-prediction logic so that it fits into navigation studies, forming a link between HCI techniques for analyzing mobile behavior and spatial cognition methods for examining map interactions.

The results of our three stages of modeling (LSTM only, context in 8D, and context reduced to 3D with AE) consistently demonstrated low precision in all configurations. While LSTM alone achieved discrimination above chance (ROC-AUC = 0.743; recall = 0.567), precision was very low (0.216). Adding the full 8D context vector improved all metrics (ROC-AUC = 0.864; recall = 0.709; precision = 0.356), confirming that traits and urban structure contribute explanatory signals beyond interaction dynamics. Third, compressing context with a 3D autoencoder (AE) provided interpretability but reduced predictive power, with precision (0.320) and recall (0.640) declining compared to the full 8D model. While **8D context fusion performed best overall**, precision remained below 0.50 in all configurations, indicating that the models did not produce **solid predictive precision** for operational deployment.

Despite this limitation, the study provides **proof** of concept for **context-aware, behavior-centric prediction**. Even if the current results are not robust and actionable in terms of accuracy, they illustrate the technical feasibility of building models that integrate **behavioral sequences, intrinsic differences, and environmental complexity** into a unified predictive framework. This aligns with long-standing calls in cartography and HCI for context-aware and explainable adaptivity in mobile maps (Ballatore & Bertolotto, 2015; Bartling et al., 2022; Crease & Reichenbacher, 2011; Fabrikant, 2023; Griffin et al., 2017).

5.3.2 Latent Dimensions in Person–Place Mobile Map Interaction Prediction

While the model’s predictive accuracy was limited, the explainability analyses showed that the autoencoder extracted interpretable blends of traits and spatial context that resonate with the constructs identified in RQ2. These latent dimensions are not direct measurements; they are **unitless coordinates derived from the autoencoder encoding layer**, which compresses the original standardized features into a smaller set of variables. Their meaning only emerges through their correlations with the observed variables, not from any intrinsic scale.

Latent 1 fused *environmental accessibility versus disorder* with *spatial anxiety*. It was positively correlated with the centrality of closeness ($r \approx +.71$) and spatial anxiety ($r \approx +.55$), and negatively associated with the entropy of orientation ($r \approx -.75$). **Higher Latent 1 scores** therefore tended to occur in more accessible and orderly street networks (high proximity, low entropy) and among participants who reported experiencing higher anxiety. **Lower Latent 1 scores** aligned with irregular and disordered layouts and lower reported anxiety. This reflects an “accessible-but-anxious” profile, consistent with evidence that anxious navigators seek reassurance even in legible environments (He & Hegarty, 2020; Lyons et al., 2018).

Latent 2 primarily reflected *individual differences in spatial confidence*. It was positively correlated with SBSOD ($r \approx +.37$) and negatively correlated with spatial anxiety ($r \approx -.50$) and male coding ($r \approx -.76$). **Higher Latent 2 scores** co-occurred with participants reporting a stronger sense of direction and lower anxiety. In this dataset, these profiles were more frequent among those coded as female. **Lower Latent 2 scores** corresponded

to higher anxiety and a weaker self-reported sense of direction, patterns that were more common among the male participants in this sample. This gender association should be treated as descriptive rather than generalizable.

Latent 3 captured *local connectivity in the street network*, showing strong negative correlations with betweenness ($r \approx -.84$), degree ($r \approx -.51$), and closeness ($r \approx -.48$). **Lower Latent 3 scores** corresponded to dense and highly connected local street networks, whereas **higher Latent 3 scores** reflected less connected structures. This aligns with theories linking network density to decision complexity and wayfinding effort (Lynch, 1960; Montello, 2009).

The use of maps was most strongly related to two profiles. First, **higher scores in Latent 2** (greater spatial confidence) accounted for roughly 30% of the predictive influence, aligning with the idea that confident navigators who primarily use the app to confirm their location use maps selectively for verification purposes (Hegarty et al., 2006; Ishikawa et al., 2008).

Second, **higher scores in Latent 1** (easy to-navigate networks with prominent anxiety) contributed to 22%, aligning with the need for assurance in environments that are otherwise easy to interpret (He & Hegarty, 2020). Conversely, a lower tendency to use maps was strongly associated with **low Latent 1 scores** (complex, disordered networks characterized by high entropy and low accessibility), accounting for approximately 27% in a negative capacity. This could be indicative of either anxious individuals withdrawing and passively following due to cognitive overload (Parush et al., 2007) or confident individuals who depend more on inherent spatial knowledge and therefore refer to the map less frequently (Coutrot et al., 2022; Montello, 2009). The impact of **Latent 3** was minimal (less than 10%), indicating that broader complexity and personal tendencies were more significant than simple local connectivity.

Together, the latent dimensions of the AE illustrate that map interaction emerges from **coupled person–place dynamics**. The most influential signals were the mixed dimensions of (i) accessibility/disorder with anxiety (Latent 1) and (ii) spatial confidence (Latent 2). While predictive accuracy was limited, the explainability analysis provides a transparent benchmark, demonstrating how the model weighted these fused factors and generated testable hypotheses about the times at which users interact more or less; for example, fewer touches in irregular, high-entropy networks and check-oriented use among more spatially confident navigators.

5.3.3 Implications

The model was not designed as a deployable detector; instead, it functions as a proof of concept that illustrates (i) the overall approach is feasible, (ii) the relevant signal can be localized within traits and urban form, and (iii) an interpretable, context-sensitive predictive layer can be implemented—an advance consistent with prior work (Coutrot et al., 2022; Greff et al., 2017; McQuire, 2019; Raper et al., 2007b; Wang et al., 2021; Watts et al., 2015; Xu et al., 2022). Before running the models, we formulated three expectations. First, the tap-sequence model should outperform a chance-level baseline. Second, adding contextual features should further improve performance. Third, compressing context with an autoencoder should produce more interpretable features, with a modest reduction in accuracy. The results matched all three expectations. The sequence-only model exceeded the chance baseline (ROC–AUC = 0.743; recall = 0.567; precision = 0.216). Incorporating the full 8-dimensional context vector improved discrimination (ROC–AUC = 0.864; recall

= 0.709; precision = 0.356). Precision remained below 0.50, consistent with the class imbalance in the data (about 9.3% positive cases) and the heterogeneous nature of the “map use” label. Compressing the context into a 3-dimensional autoencoder space increased interpretability but reduced predictive performance (precision = 0.320; recall = 0.640). Thus, the model’s primary value is in establishing the **feasibility and interpretability of human-centered models** that connect user traits, environmental context, and interaction dynamics, rather than in providing high predictive precision. To our knowledge, this work represents the first empirical step toward **explainable context-aware adaptivity** in mobile map systems that jointly incorporate intrinsic, behavioral, and extrinsic factors (Bartling et al., 2023).

5.3.4 Limitations

These findings are subject to several significant limitations. First, the data imbalance had a considerable effect on the model’s performance. In the test dataset, only 494 of 5290 windows (approximately 9.3%) included a map event, while map taps comprised approximately 3% of all phone interactions in the complete dataset. This rarity explains the low precision (0.22–0.36) despite having a recall above random chance. Future investigations should consider employing class-balancing techniques, such as focal loss, cost-sensitive training, or hybrid sampling, to mitigate false positives. Additionally, the model may simply require more data to enhance its performance. Second, our binary classification framework over-simplifies the complexities of navigation behavior. By treating all map interactions as indistinguishable, we obscure significant distinctions between activities such as destination search, rerouting, and position verification. Utilizing a multi-class, multi-label, or time-to-event model would facilitate a more nuanced understanding, offering enhanced predictions concerning both the timing and reasons for map usage. Third, the findings are tailored to specific sessions and contexts. Our study did not assess robustness across different cities or seasons, and the explainability analyzes do not allow for the inference of causality. Latent dimensions represent abstract, unit-free constructs of the model, rather than directly observable variables, and they are neither uniquely identifiable nor suitable for causal inference. Thus, correlations, such as gender loadings, should be viewed from a descriptive perspective. Finally, similar to RQ2, certain input features may require extended observation periods to achieve stability. Although our two-week MoT II dataset offers a solid basis for preliminary modeling, it might not fully capture stable habitual patterns that could enhance sequence learning. In summary, these constraints indicate that the present work should be regarded as a proof of concept rather than an operational system ready for deployment. Its primary value lies in advancing our understanding of how individual traits and spatial context jointly shape mobile map interactions in real-world urban settings.

5.4 Broader Implications of the Research

This research offers a proof-of-concept to demonstrate that mobile map prediction transcends efficiency-focused optimization by incorporating three context dimensions: inherent characteristics, external urban structures, and behavioral interaction patterns. Despite their limited predictive accuracy, the models illustrated that these variables contribute explanatory insight and can be amalgamated into an interpretable framework.

Mobile Cartography and GIScience

The findings suggest that the disparity between traditional theories of context-aware, adaptive cartography and their limited application in commercial apps can be reduced. Urban metrics and traits enhanced predictive discrimination and uncovered stable patterns that align with the principles of geographic information retrieval (De Sabbata, 2013; Raper et al., 2007b; Reichenbacher, 2007).

Human–Computer Interaction (HCI)

This study demonstrates the potential of using a behavioral intermediary layer to serve as a mediator that connects map utilization with personal variability (Ballatore & Bertolotto, 2015). Although not yet ready for widespread implementation, this method emphasizes the potential of utilizing implicit interaction traces and enriches the perspective on context-aware adaptation of the application within the HCI realm (Dey & Häkkinä, 2008; Grifoni et al., 2018; Raza & Ding, 2019; Schilit et al., 1994).

Cognitive Geography and Spatial Cognition

The predictive patterns identified in this study are consistent with the findings of cognitive geography research, indicating that personal characteristics and the intricacy of street networks collectively influence navigational behavior (Coutrot et al., 2022; Xu et al., 2022). These results also refine cognitive offloading theories (Parush et al., 2007), proposing that dependence on mobile maps does not necessarily manifest as frequent user interaction. In certain scenarios, limited interaction may still reflect reliance, as users may simply keep the app open for reassurance or reduce interaction when faced with highly complex environments.

6

CHAP-
TER

Conclusions and Outlook

*“If this is the best of possible
worlds, what then are the others?”*
Voltaire, *Candide*

6.1 Main Findings

Over the last quarter century, research in mobile cartography, GIScience, HCI, and cognitive science has converged on a set of clear principles for mobile map design. These include (i) the retrieval and filtering of geographic information tailored to user needs (De Sabbata, 2013; Reichenbacher, 2007); (ii) visualization strategies that align with cognitive and perceptual principles (Fabrikant et al., 2010; Montello et al., 2018); (iii) mobile map interaction that is designed not merely as a technical channel between the user and the system, but as part of a dynamic exchange involving the user, the environment, and the activity of navigation itself (Ballatore & Bertolotto, 2015; Griffin et al., 2017; Roth, 2013); and (iv) goal-oriented prediction, in which models anticipate user tasks and adapt map content or guidance accordingly (Ballatore & Bertolotto, 2015; Raubal et al., 2004).

However, most existing research in this field still follows a one-size-fits-all, efficiency-driven logic in the design of mobile map applications, especially regarding prediction. Prediction tasks are optimized using a narrow notion of context that is typically limited to readily available smartphone signals such as GPS location, app usage history, or coarse interaction logs. While technically convenient, this approach ignores a broader view of context and overlooks three critical components that shape everyday navigation: the extrinsic (environmental features), the intrinsic (stable individual traits and affective states), and the behavioral (moment-to-moment interaction patterns). As Bartling et al. (2023) argue, theory and design must integrate these dimensions to produce context-sensitive human-centered navigation systems.

This dissertation addressed this disconnect by operationalizing a unified notion of context that involves behavioral (tappigraphy), intrinsic (questionnaires to collect sense of direction and spatial anxiety data), and extrinsic (urban form) features and developing a model to predict the upcoming interaction with mobile map apps during navigation. Specifically, the goal was to illustrate how minimally intrusive, everyday data can be captured (RQ1), integrated to explain interaction dynamics (RQ2), and fused into a sequential predictive model of next-tap behavior (RQ3) (see 1.2). The main findings can be summarized as three key insights:

6.1.1 Parsimonious and Ecologically Valid Data Capture is Possible, but Comes With Trade-Offs

The two MapOnTap campaigns exemplify the delicate balance between collecting *minimal* data and obtaining analyses that are sufficiently robust. MoT I (tappigraphy+GPS) demonstrated that such a lean setup can be scaled to real life, but on its own it offers only limited insights into how individuals actually use mobile maps or which map apps they rely on. MoT II added SBSOD/SA and demographics data, providing richer contextual anchors; however, this came at the cost of shorter observation windows and, consequently, fewer sessions overall. Furthermore, navigation-related interactions were relatively rare, contributing only a small fraction of overall phone use. The key lesson from these data collection efforts is not that parsimonious sensing solves the problem outright, but that it enables a first, privacy-respecting step: it shows how lightweight instrumentation can open a window onto everyday navigation while also exposing the limitations and trade-offs to be addressed by future work.

6.1.2 Mobile Map Interactions are Jointly Shaped by Who Navigates and Where They Navigate

One of the most important insights of this dissertation is that navigation behavior is shaped by the **person, the place, and the interaction between them**. Examined separately, traits such as anxiety or sense of direction, and features of the urban environment, such as the regularity or accessibility of a street network, each explain only a piece of the puzzle. When considered together, the picture is clarified. Participants reporting higher spatial anxiety tapped less frequently and in a less regular rhythm, while those with a better sense of direction exhibited more consistent use patterns. Irregular or overly complex street networks also reduced interaction, whereas more accessible layouts supported more consistent engagement. Crucially, these influences did not each act alone: personal traits could amplify or diminish the impact of the city layout, and the city layout could make individual differences more or less pronounced. In short, the way people tap on a mobile map is co-produced by who they are and where they are, showing that navigation must be understood as a dynamic process shaped by both person and place at the same time.

6.1.3 Context-Aware Prediction is Possible and Interpretable

This dissertation made a first attempt to predict an individual's next interaction with a mobile map by combining recent tapping behavior with information about them and their urban environment. The model was not robust enough to serve as a reliable computational tool, but it demonstrated that such prediction is technically feasible and consistent with that which theories of adaptive cartography and HCI have been calling for over many years. Just as importantly, the model's explainability illustrated why it leaned toward or away from predicting map use: signals related to a street network's complexity and spatial anxiety tended to reduce the likelihood of interaction, while a stronger sense of direction or certain demographic differences tended to support it. This means that we now have the tools to go beyond merely asking *whether* prediction is possible and to begin understanding *what drives it*, laying the groundwork for future systems that adapt to both person and place in real time.

6.2 Design Recommendations

The predictive model developed in this dissertation explicitly anticipates the next user interaction with mobile map applications by fusing real-time behavioral signals (tappigraphy sequences), intrinsic cognitive-emotional traits (spatial anxiety, sense of direction), and extrinsic urban form metrics (orientation entropy, network centrality). The model's predictive capability does more than simply classify taps: it could potentially identify *moments of navigational demand* and *states of vulnerability or confidence*. The following three complementary design directions emerge based on these insights.

Real-Time Contextual Filtering. Our findings show that map interactions frequently cluster at specific moments of need; for example, when anxiety is high or the surrounding street network is complex. When the model predicts a high likelihood of imminent map interaction (map class), this moment should be treated as a *critical window for support*. At this point, the system could proactively suppress irrelevant information (e.g., secondary POIs, background layers) and emphasize only the essentials such as the next turn or immediate path segment. This reduces information overload precisely when users are most cognitively taxed, aligning with theories of limited attentional capacity and prior research on adaptive geographic relevance.

How it could work: modern map SDKs (e.g., Google Maps, Mapbox, Leaflet) already allow developers to dynamically turn on and off layers. The prediction model would simply act as a trigger, instructing the app to load a simplified style sheet or hide secondary layers whenever a high-probability event is detected.

Adaptive Map Visualization. The results also highlight that interaction is jointly shaped by both individual traits and the environment: high spatial anxiety and high orientation entropy predict fewer, more irregular taps, while a strong sense of direction suggests more stable use. These differences motivate adaptive visualization. Users predicted to be struggling in complex layouts could be offered landmark-based or route-centric views, reducing environmental complexity and providing reassurance. In contrast, confident users (low anxiety, high SBSOD) may benefit from richer allocentric maps that encourage spatial learning and exploration. This dual strategy addresses both immediate usability and long-term cognitive sustainability, an issue often overlooked in commercial systems that optimize only for efficiency.

How it could work: adaptive visualization can be achieved by preparing multiple map styles in advance (e.g., a simplified route-only style and a detailed allocentric style) and switching between them at run-time. Existing vector-tile rendering engines already support dynamic style switching, so the model would only need to signal *which* style to activate given the predicted state.

User Interface Adaptation Finally, the model enables interface-level adjustments at precise moments when users are predicted to reconnect with the map. For anxious users, the system could temporarily reduce visual clutter, highlight only

the immediate route segment, and provide lightweight reassurance cues (e.g., “Continue straight, you are on track”). For confident users, the interface could surface optional exploratory features or suggest alternative paths. Importantly, such adaptations should be transparent: brief, user-friendly explanations can link changes to detected traits or environmental conditions, reinforcing trust and user agency instead of obscuring system behavior.

How it could work: interface adaptation can be implemented using conditional UI components. For example, buttons and panels can be hidden or shown dynamically using existing mobile OS APIs, while prompts can be delivered as lightweight notifications, banners, or voice cues. Transparency could be built in by displaying short explanations (e.g., “Simplified view enabled due to complex surroundings”), similar to how apps already provide status updates when switching modes.

6.3 Contributions

This dissertation was motivated by three long-standing gaps, which it seeks to address. **First**, it provides an alternative to the two dominant approaches to research on mobile map use. For example, the methodological approach based on controlled indoor and outdoor studies offers high control but low ecological validity, especially when it is laboratory-based, or high validity but high invasiveness, especially if outdoor studies use tools such as mobile EEG or eye-tracking. Previous work on VR and field experiments has contributed valuable insights into navigation processes (Cheng et al., 2022; Dahmani & Bohbot, 2020; Gardony et al., 2015; Jeung et al., 2023), but these methods struggle to capture the fragmented multitasking character of the use of mobile maps in the real world and often impose artificial tasks or invasive equipment that reduce ecological validity (Oulasvirta et al., 2012). Second, a methodological approach based on telemetry data collected by commercial providers offers high ecological validity, but is rarely accessible to independent research and often raises concerns about intrusiveness and surveillance (Ballatore & Bertolotto, 2015).

The MapOnTap pipeline provides methodological flexibility by achieving **high ecological validity** and **accessibility** while maintaining **low invasiveness**. It does this by adhering to the principle of parsimony, collecting only the information necessary to derive meaningful insights. By focusing on tappigraphy as the core signal and supplementing it with voluntary GPS traces, questionnaires, and demographic information, the pipeline supports multiple levels of granularity to extract behavioral insights. This design ensures that everyday navigation behavior is captured unobtrusively in naturalistic settings, without the need for heavy instrumentation or reliance on opaque commercial telemetry.

Second, this dissertation advances the literature by providing empirical evidence that personal traits and urban form jointly shape mobile map interaction in everyday contexts. Previous studies typically examine the cognitive effects of **GPS reliance** on its own, for example, showing that turn-by-turn navigation can weaken allocentric strategies and long-term spatial learning (Dahmani & Bohbot, 2020; He & Hegarty, 2020; Parush et al., 2007; Ruginski et al., 2022) and that these effects are mediated by individual differences in spatial traits and emotional dispositions, such as sense of direction, spatial anxiety, or gender (Hegarty et al., 2002; Lawton, 1994; Lyons et al., 2018). In parallel, other research has focused exclusively on the effects of **urban form**, linking city legibility, differentiation, and network configuration with spatial cognition at the individual or population scale (Coutrot

et al., 2022; Lynch, 1960; Montello, 2005; Watts et al., 2015; Xu et al., 2022). However, these strands have developed largely in isolation and direct studies of **map interaction itself** remain rare. An exception is the work of Savino et al. (2021), who detected recurring interaction patterns in daily map use; however, the reliance on a browser wrapper limited ecological validity and generalizability compared to full-featured applications.

By integrating tappigraphy, GPS traces, and trait measures, this dissertation demonstrates that interaction rhythms are **co-produced by person and place**: anxious users produce fewer taps and tap less regularly, while urban features such as orientation entropy, degree, and closeness systematically modulate these behaviors. Moderation analyses further show *when* traits matter more, such as the joint effects of spatial anxiety and sense of direction or the interaction between closeness and gender, thus moving beyond trait-only or place-only accounts and advancing the literature toward a contextualized, person-by-place explanation of mobile map use in the wild.

Third, this dissertation advances the literature by shifting the focus of prediction from *where* users will go to *how* they will interact with the map. Current systems primarily optimize efficiency through destination inference, ETA calculation, and route selection (Huang et al., 2018; Rao & Minakakis, 2003; Raper et al., 2007b), employing models based on classic and modern routing and traffic prediction methods (Cui et al., 2019; Derrow-Pinion et al., 2021; Dijkstra, 2022; Hart et al., 1968; Herrera et al., 2010). In contrast, this dissertation introduces a context-aware LSTM model that combines short-term interaction sequences (tappigraphy) with intrinsic traits (e.g., sense of direction, spatial anxiety) and extrinsic urban form metrics (e.g., orientation entropy, network centrality). This operationalizes the long-anticipated *behavioral middle layer* proposed by adaptive cartography and geographic relevance theory (De Sabbata, 2013; Griffin et al., 2017; Reichenbacher, 2001, 2009), providing a model that connects the cognitive-emotional state and environment of the user directly to their use of the map.

Technically, the model employs LSTM sequence learning with encoder–decoder architectures (Greff et al., 2017; Sutskever et al., 2014) and context fusion methods inspired by next-app prediction (Katsarou et al., 2022; Wang et al., 2021) to anticipate future interactions. Essentially, the model produces interpretable outputs via SHAP that identify the roles of specific person–place factors, such as spatial anxiety, sense of direction, and urban complexity, in shaping predictions. This makes it possible to move beyond opaque, one-size-fits-all systems towards a parsimonious, privacy-preserving framework for adaptive navigation that relies only on tappigraphy, voluntary GPS traces, and short trait questionnaires (Balerna & Ghosh, 2018; Ceolini et al., 2022; Reichenbacher et al., 2022).

To summarize, these three contributions collectively demonstrate that mobile map research can move beyond fragmented and efficiency-driven approaches by (1) establishing a parsimonious yet ecologically valid data collection pipeline for mobile map data collection, (2) showing that navigation behavior is jointly shaped by personal traits and urban form in everyday contexts, and (3) introducing an interpretable predictive model that anticipates not only where people go but how they interact with maps, paving the way toward cognitively sustainable and privacy-respecting adaptive navigation systems.

6.4 Outlook

This dissertation shows that real-world everyday navigation can be studied using low-level features to extract interpretable signals. Simultaneously, it highlights where caution is required and how research should proceed. Although the results appear encouraging, they are still restricted by various practical limitations that affect both causal inference and predictive accuracy.

First, the temporal and demographic scope of the data was intentionally limited. The two MapOnTap campaigns demonstrate this trade-off. MoT I offered longer observation periods, sometimes spanning several months, but gathered minimal demographic or trait data beyond tappigraphy and voluntary GPS. In contrast, MoT II enhanced the dataset by employing trait measures (SBSOD, spatial anxiety) and collecting demographic details but had a shorter observation period, typically around two weeks. This balance illustrates the principle of parsimony: gathering only essential data to protect privacy and ensure feasibility. However, this approach restricts long-term and population-wide generalizability. Short observation periods fail to capture how navigation skills and learning evolve over longer durations, and relatively small, conservative samples decrease the ability to conduct detailed subgroup analyses.

Second, the infrequency of navigation-related interactions (i.e., map taps) posed challenges for making inferences throughout the entire pipeline. Findings from both data campaigns revealed that the utilization of maps constitutes a minor portion of total smartphone activity. Specifically, in MoT I and MoT II, map applications accounted for under 4% of all taps. This scarcity has a direct impact on the predictive framework as map taps, representing the positive class, are significantly less common than other types of phone interactions. The imbalance that results from this rarity makes models particularly sensitive to techniques for handling rare events, which in turn restricts the robustness and stability of predictions at the event level.

Third, the research failed to capture detailed in-app semantics, such as specific features accessed in a map application. Consequently, many task-level behaviors remain hidden. The taps and their patterns merely offer a broad and privacy-conscious approximation instead of a comprehensive account of navigation tasks. Finally, the psychological evaluation was intentionally minimal. It included only brief and privacy-sensitive tools such as the Santa Barbara Sense of Direction scale (Hegarty et al., 2002) and measures of spatial anxiety (He & Hegarty, 2020). This approach ensured practicality and participant comfort, but excluded other constructs central to navigation theory, such as mental rotation, perspective taking, and visuospatial working memory (Ruginski et al., 2022).

These limitations do not diminish the value of the MapOnTap pipeline; rather, they outline its current claims and suggest a specific and manageable research agenda. Future research should broaden measurement efforts while prioritizing privacy-first principles. Expanding psychometric breadth is possible with targeted additions, such as short mental rotation tasks, quick perspective-taking assessments, and concise landmark and route learning probes that both align with navigation theory and are practical for field deployment (Ruginski et al., 2022).

In the future, it is essential to clarify the mobility context, as the interpretation of a tap varies depending on whether the user is leisurely walking through a dense street network or traveling swiftly in a car or train. Without this differentiation, the same tap sequence

could be misunderstood. Privacy-friendly proxies, like general accelerometer summaries or step-count categories, can serve as dependable indicators of mode and speed while safeguarding sensitive location information.

The analysis of urban form necessitates consideration across various spatial dimensions. Elements such as the intricacy of junctions, the density of landmarks, and the general configuration of networks influence behavior variably when assessed locally (e.g., at an individual intersection), at the neighborhood scale, or throughout the entire urban area. Previous research indicates that orientation skills and cognitive outcomes vary with scale (Boeing, 2017; Coutrot et al., 2022); thus, evaluating across these levels is essential for comprehensively understanding the full spectrum of interactions between individuals and their environments.

Finally, future work should address the lack of semantic information in tappigraphy by collecting ground truth data using task-based protocols that elicit specific interaction states by design, rather than logging map content or queries. Participants could perform short, scripted micro-tasks that generate identifiable states such as pan and zoom, recenter, search, or route initiation with a clear start/stop marker and expected control sequence, so the associated tap rhythms are labeled by construction. These sessions would provide a compact labeled dataset to train a state classifier that maps tap sequences to interaction states. Once trained, such a classifier could run entirely on a device during natural use, offering low-resolution semantics.

This dissertation extends beyond methodologies and models, providing a pragmatic guide for examining and enhancing everyday navigation in relevant ways across various fields. It demonstrates that theories from cognitive geography and spatial cognition are testable and improvable through the analysis of minimally invasive, interpretable data, facilitating hypothesis-driven research that connects laboratory studies with real-world applications (Lynch, 1960; Montello, 2005). Adaptive maps are highlighted as tools for reducing cognitive overload during crucial instances, improving safety and accessibility for individuals who are older, visually impaired, or have mobility or cognitive challenges, without creating long-term reliance on devices. It shows how privacy-preserving behavioral assessments can pinpoint where urban layouts increase cognitive burdens, supporting targeted, cost-effective measures (such as signage, landmarks, and wayfinding) that enhance navigational clarity on a large scale (Watts et al., 2015). Furthermore, this thesis outlines a viable strategy with which the industry can promote *trust-first* personalization: lightweight, on-device models that offer targeted assistance and concise explanations instead of obscure monitoring, thereby enhancing user experience while also mitigating regulatory and safety concerns (Martinez-Martin et al., 2018; Montag & Baumeister, 2023).

The collective findings suggest a different objective for navigation systems—enhancing rather than replacing human spatial cognition. Navigation is an intrinsic cognitive and social skill influenced by memory, emotion, environment, and interface design. Systems prioritizing only speed may constrain experience, undermine learning, and potentially lead to unquestioning adherence (Dahmani & Bohbot, 2020; Parush et al., 2007). The proposed alternative is straightforward and actionable: craft mobile maps to be minimalist, stratified, and open, amassing minimal yet pertinent data, converting it into understandable signals at appropriate intervals, and modifying interfaces to honor both the user and the environment, all while ensuring user dominance. The adoption of these guidelines could render mobile maps more secure, equitable, and humane, assisting navigation novices in building their

confidence, providing experts with compatible tools, and retaining the map within the realm of human cognition, rather than relegating it to the domain of machines.

Bibliography

- Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact [Place: US Publisher: American Psychological Association]. *Journal of Applied Psychology*, 69(2), 334–345. <https://doi.org/10.1037/0021-9010.69.2.334>
- Apple Inc. (2012). Take flyover tours in maps on iphone [Accessed: 2025-05-06]. <https://support.apple.com/guide/iphone/take-flyover-tours-iph81a3f978/ios>
- Ashbrook, D., & Starner, T. (2003). Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5), 275–286. <https://doi.org/10.1007/s00779-003-0240-0>
- Balerna, M., & Ghosh, A. (2018). The details of past actions on a smartphone touchscreen are reflected by intrinsic sensorimotor dynamics. *npj Digital Medicine*, 1, 4. <https://doi.org/10.1038/s41746-017-0011-3>
- Ballatore, A., & Bertolotto, M. (2015). Personalizing maps [Publisher: ACM]. *Communications of the ACM*, 58(12), 68–74. <https://doi.org/10.1145/2756546>
- Bartling, M., Reichenbacher, T., & Fabrikant, S. I. (2023). Leveraging map use context for advancing cartography in the 21st century [Publisher: Copernicus GmbH]. *Proceedings of the ICA*, 5, 1–7. <https://doi.org/10.5194/ica-proc-5-2-2023>
- Bartling, M., Resch, B., Reichenbacher, T., Havas, C. R., Robinson, A. C., Fabrikant, S. I., & Blaschke, T. (2022). Adapting mobile map application designs to map use context: A review and call for action on potential future research themes [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/15230406.2021.2015720>]. *Cartography and Geographic Information Science*, 49(3), 237–251. <https://doi.org/10.1080/15230406.2021.2015720>
- Bartling, M., Robinson, A. C., Resch, B., Eitzinger, A., & Atzmanstorfer, K. (2021). The role of user context in the design of mobile map applications [Publisher: Taylor & Francis]. *Cartography and Geographic Information Science*, 48(5), 432–448. <https://doi.org/10.1080/15230406.2021.1933595>
- Baum, M., Dibbelt, J., Pajor, T., & Wagner, D. (2016). Dynamic Time-Dependent Route Planning in Road Networks with User Preferences. In A. V. Goldberg & A. S. Kulikov (Eds.), *Experimental Algorithms* (pp. 33–49). Springer International Publishing. https://doi.org/10.1007/978-3-319-38851-9_3

-
- Bennett, J. (2010). *Openstreetmap*. Packt Publishing Ltd. <https://books.google.com/books?id=SZfqRcPXApoC>
- Bhavnani, Y., Rodden, K., Guarnotta, L. C., Lynn, M. T., Chizari, S., & Granka, L. (2017). Understanding mobile phone activities via retrospective review of visualizations of usage data. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–10. <https://doi.org/10.1145/3098279.3119841>
- Bishop, C. M., & Bishop, H. (2023, November). *Deep learning: Foundations and concepts* [Google-Books-ID: 0uTgEAAAQBAJ]. Springer Nature. <https://link.springer.com/book/10.1007/978-3-031-45468-4>
- Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, 126–139. <https://doi.org/10.1016/j.compenvurbsys.2017.05.004>
- Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011). Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, 47–56. <https://doi.org/10.1145/2037373.2037383>
- Borger, J. N., Huber, R., & Ghosh, A. (2019). Capturing sleep–wake cycles by using day-to-day smartphone touchscreen interactions. *npj Digital Medicine*, 2(1), 73. <https://doi.org/10.1038/s41746-019-0147-4>
- Boschmann, E. E., & Cubbon, E. (2014). Sketch maps and qualitative gis: Using cartographies of individual spatial narratives in geographic research. *The Professional Geographer*, 66(2), 236–248. <https://doi.org/10.1080/00330124.2013.781490>
- Brügger, A., Richter, K.-F., & Fabrikant, S. I. (2019). How does navigation system behavior influence human behavior? *Cognitive Research: Principles and Implications*, 4(1), 5. <https://doi.org/10.1186/s41235-019-0156-5>
- Burgess, N. (2006). Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Sciences*, 10(12), 551–557. <https://doi.org/10.1016/j.tics.2006.10.005>
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd). Springer.
- Ceolini, E., & Ghosh, A. (2023). Common multi-day rhythms in smartphone behavior [Publisher: Nature Publishing Group]. *npj Digital Medicine*, 6(1), 1–9. <https://doi.org/10.1038/s41746-023-00799-7>
- Ceolini, E., Kock, R., Band, G. P., Stoet, G., & Ghosh, A. (2022). Temporal clusters of age-related behavioral alterations captured in smartphone touchscreen interactions. *iScience*, 25(8), 104791. <https://doi.org/10.1016/j.isci.2022.104791>
- Cheng, B., Wunderlich, A., Gramann, K., Lin, E., & Fabrikant, S. I. (2022). The effect of landmark visualization in mobile maps on brain activity during navigation: A virtual reality study [Publisher: Frontiers]. *Frontiers in Virtual Reality*, 3. <https://doi.org/10.3389/frvir.2022.981625>

-
- Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review*, 19(1), 1–23. <https://doi.org/10.3758/s13423-011-0182-x>
- Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, R. C., Wiener, J. M., Hölscher, C., Hornberger, M., & Spiers, H. J. (2022). Entropy of city street networks linked to future spatial navigation ability. *Nature*, 604(7904), 104–110. <https://doi.org/10.1038/s41586-022-04486-7>
- Coutrot, A., Schmidt, S., Coutrot, L., Pittman, J., Hong, L., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2019). Virtual navigation tested on a mobile app is predictive of real-world wayfinding navigation performance [Publisher: Public Library of Science]. *PLOS ONE*, 14(3), e0213272. <https://doi.org/10.1371/journal.pone.0213272>
- Crease, P. E., & Reichenbacher, T. (2011). Designing usable cartographic representations of geographic relevance for LBS users. *Proceedings of the 25th International Cartographic Conference*, CO–073. Retrieved June 27, 2025, from https://www.zora.uzh.ch/id/eprint/51496/1/2011_CreaseP_Designing_Usable_Cartographic_Representations_for_Mobile_Information_Seeking_v6.pdf
- Cui, Z., Henrickson, K., Ke, R., & Wang, Y. (2019). Traffic Graph Convolutional Recurrent Neural Network: A Deep Learning Framework for Network-Scale Traffic Learning and Forecasting. *IEEE Transactions on Intelligent Transportation Systems*, PP. <https://doi.org/10.1109/TITS.2019.2950416>
- Cupper, A., Treu, G., & Linnhoff-Popien, C. (2006). TraX: A device-centric middleware framework for location-based services. *IEEE Communications Magazine*, 44(9), 114–120. <https://doi.org/10.1109/MCOM.2006.1705987>
- Dahmani, L., & Bohbot, V. D. (2020). Habitual use of GPS negatively impacts spatial memory during self-guided navigation. *Scientific Reports*, 10(1), 6310. <https://doi.org/10.1038/s41598-020-62877-0>
- De Sabbata, S. (2013). Assessing geographic relevance for mobile information services [Publisher: University of Zurich]. <https://doi.org/10.5167/UZH-89056>
- Deng, T., Kanthawala, S., Meng, J., Peng, W., Kononova, A., Hao, Q., Zhang, Q., & David, P. (2019). Measuring smartphone usage and task switching with log tracking and self-reports. *Mobile Media & Communication*, 7(1), 3–23. <https://doi.org/10.1177/2050157918761491>
- Derrow-Pinion, A., She, J., Wong, D., Lange, O., Hester, T., Perez, L., Nunkesser, M., Lee, S., Guo, X., Wiltshire, B., Battaglia, P. W., Gupta, V., Li, A., Xu, Z., Sanchez-Gonzalez, A., Li, Y., & Velickovic, P. (2021). ETA Prediction with Graph Neural Networks in Google Maps. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, 3767–3776. <https://doi.org/10.1145/3459637.3481916>
- Dey, A. K. (2001). Understanding and using context. *Personal and Ubiquitous Computing*, 5(1), 4–7. <https://doi.org/10.1007/s007790170019>

-
- Dey, A. K., & Häkkinen, J. (2008). Context-awareness and mobile devices. In *Handbook of research on user interface design and evaluation for mobile technology* (pp. 205–217). IGI Global. <https://doi.org/10.4018/978-1-59904-871-0.ch013>
- Dijkstra, E. W. (2022, July). A Note on Two Problems in Connexion with Graphs. In *Edsger Wybe Dijkstra: His Life, Work, and Legacy* (1st ed., pp. 287–290, Vol. 45). Association for Computing Machinery. Retrieved June 25, 2025, from <https://doi.org/10.1145/3544585.3544600>
- Doyle, J., Bertolotto, M., & Wilson, D. C. (2010). Evaluating the benefits of multimodal interface design for compass — a mobile gis. *GeoInformatica*, 14(2), 135–162. <https://doi.org/10.1007/s10707-009-0079-2>
- Ekstrom, A., Spiers, H. J., Bohbot, V. D., & Rosenbaum, R. S. (2018). *Human Spatial Navigation*. Princeton University Press.
- Emmanouilidis, C., Koutsiamanis, R.-A., & Tasidou, A. (2013). Mobile guides: Taxonomy of architectures, context awareness, technologies and applications. *Journal of Network and Computer Applications*, 36(1), 103–125. <https://doi.org/10.1016/j.jnca.2012.04.007>
- Eugster, M. J. A., & Leisch, F. (2009). From Spider-Man to Hero — Archetypal Analysis in R. *Journal of Statistical Software*, 30, 1–23. <https://doi.org/10.18637/jss.v030.i08>
- Fabrikant, S. I. (2023). Neuroadaptive LBS: Towards human-, context-, and task-adaptive mobile geographic information displays to support spatial learning for pedestrian navigation. *Journal of Location Based Services*, 17(4), 340–354. <https://doi.org/10.1080/17489725.2023.2258100>
- Fabrikant, S. I., & Lobben, A. (2009). Introduction: Cognitive Issues in Geographic Information Visualization [Publisher: University of Toronto Press]. *Cartographica*, 44(3), 139–143. <https://doi.org/10.3138/carto.44.3.139>
- Fabrikant, S. I., Rebich Hespanha, S., & Hegarty, M. (2010). Cognitively inspired and perceptually salient graphic displays for efficient spatial inference making [Publisher: Routledge]. *Annals of the Association of American Geographers*, 100(1), 13–29. <https://doi.org/10.1080/00045600903362378>
- Fabrikant, S. I., Rebich-Hespanha, S., Andrienko, N., Andrienko, G., & Montello, D. R. (2008). Novel method to measure inference affordance in static small-multiple map displays representing dynamic processes [Publisher: Taylor & Francis]. *The Cartographic Journal*, 45(3), 201–215. <https://doi.org/10.1179/000870408X311396>
- Falaki, H., Mahajan, R., Kandula, S., Lymberopoulos, D., Govindan, R., & Estrin, D. (2010). Diversity in smartphone usage. *Proceedings of the 8th international conference on Mobile systems, applications, and services*, 179–194. <https://doi.org/10.1145/1814433.1814453>
- Frank, A. U. (1992). Spatial concepts, geometric data models, and geometric data structures. *Computers & Geosciences*, 18(4), 409–417. [https://doi.org/10.1016/0098-3004\(92\)90070-8](https://doi.org/10.1016/0098-3004(92)90070-8)

- Gardony, A. L., Brunyé, T. T., Mahoney, C. R., & Taylor, H. A. (2013). How Navigational Aids Impair Spatial Memory: Evidence for Divided Attention [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13875868.2013.792821>]. *Spatial Cognition & Computation*, 13(4), 319–350. <https://doi.org/10.1080/13875868.2013.792821>
- Gardony, A. L., Brunyé, T. T., & Taylor, H. A. (2015). Navigational aids and spatial memory impairment: The role of divided attention [Place: United Kingdom Publisher: Taylor & Francis]. *Spatial Cognition and Computation*, 15(4), 246–284. <https://doi.org/10.1080/13875868.2015.1059432>
- Gifford, R. (2014). *Environmental psychology: Principles and practice*. Optimal Books.
- Golledge, R. G. (2004). Human wayfinding. In A. Bailly & L. J. Gibson (Eds.), *Applied Geography: A World Perspective* (pp. 233–252). Springer Netherlands. https://doi.org/10.1007/978-1-4020-2442-9_13
- Google. (2020, February). *A look back at 15 years of mapping the world* [Retrieved from Google Blog]. Retrieved May 6, 2025, from <https://blog.google/products/maps/look-back-15-years-mapping-world/>
- Google Inc. (2024). Google maps is piloting generative ai with local guides [Accessed June 2025].
- Gori, M., Betti, A., & Melacci, S. (2024). Chapter 2 - Learning principles. In M. Gori, A. Betti, & S. Melacci (Eds.), *Machine Learning (Second Edition)* (Second Edition, pp. 53–111). Morgan Kaufmann. <https://doi.org/https://doi.org/10.1016/B978-0-32-389859-1.00009-X>
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222–2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Griffin, A. L., Reichenbacher, T., Liao, H., Wang, W., & Cao, Y. (2023). Cognitive issues of mobile map design and use: A collaborative research agenda [Publisher: Copernicus GmbH]. *Abstracts of the ICA*, 6, 1–3. <https://doi.org/10.5194/ica-abs-6-79-2023>
- Griffin, A. L., White, T., Fish, C., Tomio, B., Huang, H., Sluter, C. R., Bravo, J. V. M., Fabrikant, S. I., Bleisch, S., Yamada, M., & Picanço, P. (2017). Designing across map use contexts: A research agenda. *International Journal of Cartography*, 3(sup1), 90–114. <https://doi.org/10.1080/23729333.2017.1315988>
- Grifoni, P., D’Ulizia, A., & Ferri, F. (2018). Context-Awareness in Location Based Services in the Big Data Era. In G. Skourletopoulos, G. Mastorakis, C. X. Mavromoustakis, C. Dobre, & E. Pallis (Eds.), *Mobile Big Data: A Roadmap from Models to Technologies* (pp. 85–127). Springer International Publishing. https://doi.org/10.1007/978-3-319-67925-9_5
- Hägerstrand, T. (1970). What about people in regional science? [Presented at the 1969 European Congress of the Regional Science Association; republished in 1986 and

-
- reprinted in 2013]. In E. de Boer (Ed.), *Transport sociology: Social aspects of transport planning* (pp. 143–158). Franklin Book Company / Pergamon Press.
- Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics*, 4(2), 100–107. <https://doi.org/10.1109/TSSC.1968.300136>
- He, C., & Hegarty, M. (2020). How anxiety and growth mindset are linked to navigation ability: Impacts of exploration and GPS use. *Journal of Environmental Psychology*, 71, 101475. <https://doi.org/10.1016/j.jenvp.2020.101475>
- Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task [Publisher: American Psychological Association]. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(1), 37–53. <https://doi.org/10.1037/a0017683>
- Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, 34(2), 151–176. <https://doi.org/10.1016/j.intell.2005.09.005>
- Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30(5), 425–447. [https://doi.org/10.1016/S0160-2896\(02\)00116-2](https://doi.org/10.1016/S0160-2896(02)00116-2)
- Hejtmánek, L., Oravcová, I., Motýl, J., Horáček, J., & Fajnerová, I. (2018). Spatial knowledge impairment after GPS guided navigation: Eye-tracking study in a virtual town [Place: Netherlands Publisher: Elsevier Science]. *International Journal of Human-Computer Studies*, 116, 15–24. <https://doi.org/10.1016/j.ijhcs.2018.04.006>
- Herrera, J. C., Work, D. B., Herring, R., Ban, X. (, Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The *Mobile Century* field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4), 568–583. <https://doi.org/10.1016/j.trc.2009.10.006>
- Hilton, C., Kapaj, A., & Fabrikant, S. I. (2023). Landmark sequence learning from real-world route navigation and the impact of navigation aid visualisation style. *Journal of Cognition*, 6(1), 1–18. <https://doi.org/10.5334/joc.307>
- Hong, J.-y., Suh, E.-h., & Kim, S.-J. (2009). Context-aware systems: A literature review and classification. *Expert Systems with Applications*, 36(4), 8509–8522. <https://doi.org/10.1016/j.eswa.2008.10.071>
- Huang, H., Gartner, G., Krisp, J. M., Raubal, M., & Van de Weghe, N. (2018). Location based services: Ongoing evolution and research agenda [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/17489725.2018.1508763>]. *Journal of Location Based Services*, 12(2), 63–93. <https://doi.org/10.1080/17489725.2018.1508763>
- Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping: A timely opportunity to consider purpose, quality, and safety [Publisher: Nature

-
- Publishing Group]. *npj Digital Medicine*, 2(1), 88. <https://doi.org/10.1038/s41746-019-0166-1>
- Hund, A. M., & Minarik, J. L. (2006). Getting from here to there: Spatial anxiety, wayfinding strategies, direction type, and wayfinding efficiency [Accessed: 2025-05-13]. *Spatial Cognition and Computation*, 6(3), 179–201. https://doi.org/10.1207/s15427633scc0603_1
- Ishikawa, T. (2019). Satellite Navigation and Geospatial Awareness: Long-Term Effects of Using Navigation Tools on Wayfinding and Spatial Orientation [Publisher: Routledge]. *The Professional Geographer*. Retrieved February 20, 2024, from <https://www.tandfonline.com/doi/abs/10.1080/00330124.2018.1479970>
- Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology*, 28(1), 74–82. <https://doi.org/10.1016/j.jenvp.2007.09.002>
- Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places [Place: Netherlands Publisher: Elsevier Science]. *Cognitive Psychology*, 52(2), 93–129. <https://doi.org/10.1016/j.cogpsych.2005.08.003>
- Jeung, S., Hilton, C., Berg, T., Gehrke, L., & Gramann, K. (2023). Virtual Reality for Spatial Navigation. In C. Maymon, G. Grimshaw, & Y. C. Wu (Eds.), *Virtual Reality in Behavioral Neuroscience: New Insights and Methods* (pp. 103–129). Springer International Publishing. https://doi.org/10.1007/7854_2022_403
- Kapaj, A., Hilton, C., Lanini-Maggi, S., & Fabrikant, S. I. (2024). The influence of landmark visualization style on task performance, visual attention, and spatial learning in a real-world navigation task [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13875868.2024.2328099>]. *Spatial Cognition & Computation*, 0(0), 1–41. <https://doi.org/10.1080/13875868.2024.2328099>
- Kaptelinin, V., & Nardi, B. A. (2009). *Acting with Technology: Activity Theory and Interaction Design*. MIT Press.
- Katsarou, K., Yu, G., & Beierle, F. (2022). WhatsNextApp: LSTM-Based Next-App Prediction With App Usage Sequences. *IEEE Access*, 10, 18233–18247. <https://doi.org/10.1109/ACCESS.2022.3150874>
- Keith, I. (2009, October). Announcing Google Maps Navigation for Android 2.0. Retrieved May 5, 2025, from <http://googlemobile.blogspot.com/2009/10/announcing-google-maps-navigation-for.html>
- Kessler, F., Frankenstein, J., & Rothkopf, C. A. (2024). Human navigation strategies and their errors result from dynamic interactions of spatial uncertainties [Publisher: Nature Publishing Group]. *Nature Communications*, 15(1), 5677. <https://doi.org/10.1038/s41467-024-49722-y>

-
- Kiefer, P., Giannopoulos, I., Anagnostopoulos, V. A., Schöning, J., & Raubal, M. (2017). Controllability matters: The user experience of adaptive maps [Publisher: Springer]. *GeoInformatica*, 21(3), 619–641. <https://doi.org/10.1007/s10707-016-0282-x>
- Kitchin, R. M. (1994). Cognitive maps: What are they and why study them? *Journal of Environmental Psychology*, 14(1), 1–19. [https://doi.org/10.1016/S0272-4944\(05\)80194-X](https://doi.org/10.1016/S0272-4944(05)80194-X)
- Kozlowski, L. T., & Bryant, K. J. (1977). Sense of direction, spatial orientation, and cognitive maps [Place: US Publisher: American Psychological Association]. *Journal of Experimental Psychology: Human Perception and Performance*, 3(4), 590–598. <https://doi.org/10.1037/0096-1523.3.4.590>
- Krajci, I., & Cummings, D. (2013). History and Evolution of the Android OS. In I. Krajci & D. Cummings (Eds.), *Android on x86: An Introduction to Optimizing for Intel® Architecture* (pp. 1–8). Apress. https://doi.org/10.1007/978-1-4302-6131-5_1
- Krumm, J., & Horvitz, E. (2006). Predestination: Inferring Destinations from Partial Trajectories. In P. Dourish & A. Friday (Eds.), *UbiComp 2006: Ubiquitous Computing* (pp. 243–260). Springer. https://doi.org/10.1007/11853565_15
- Kuhn, W. (2001). Ontologies in support of activities in geographical space [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13658810110061180>]. *International Journal of Geographical Information Science*, 15(7), 613–631. <https://doi.org/10.1080/13658810110061180>
- Lawton, C. A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles*, 30(11), 765–779. <https://doi.org/10.1007/BF01544230>
- Lawton, C. A., & Kallai, J. (2002). Gender Differences in Wayfinding Strategies and Anxiety About Wayfinding: A Cross-Cultural Comparison. *Sex Roles*, 47(9), 389–401. <https://doi.org/10.1023/A:1021668724970>
- Lee, W.-M. (2012). *Beginning Android 4 Application Development*. John Wiley & Sons.
- Leroux, P., Roobroeck, K., Dhoedt, B., Demeester, P., & De Turck, F. (2013). Mobile application usage prediction through context-based learning. *Journal of Ambient Intelligence and Smart Environments*, 5(2), 213–235.
- Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., & Rui, Y. (2014). GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 831–840. <https://doi.org/10.1145/2623330.2623638>
- Lin, A. Y., Kuehl, K., Schöning, J., & Hecht, B. (2017). Understanding "Death by GPS": A Systematic Study of Catastrophic Incidents Associated with Personal Navigation Technologies. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1154–1166. <https://doi.org/10.1145/3025453.3025737>
- Low, T. (2022). *Towards combining search and exploration: Escaping the filter bubble through map-based exploration* [Doctoral dissertation, Martin Luther University]

-
- Halle-Wittenberg] [PhD Dissertation. Accepted: 2022-02-24]. <https://doi.org/10.25673/69177>
- Löwen, H., Krukar, J., & Schwering, A. (2019). Spatial Learning with Orientation Maps: The Influence of Different Environmental Features on Spatial Knowledge Acquisition. *ISPRS International Journal of Geo-Information*, 8(3), 149. <https://doi.org/10.3390/ijgi8030149>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.
- Lynch, K. (1960). *The image of the city*. MIT Press.
- Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial Anxiety: A Novel Questionnaire With Subscales for Measuring Three Aspects of Spatial Anxiety. *Journal of Numerical Cognition*, 4(3), 526–553. <https://doi.org/10.5964/jnc.v4i3.154>
- MacEachren, A. M., & Kraak, M.-J. (2001). Research Challenges in Geovisualization [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1559/152304001782173970>]. *Cartography and Geographic Information Science*, 28(1), 3–12. <https://doi.org/10.1559/152304001782173970>
- Martinez-Martin, N., Insel, T. R., Dagum, P., Greely, H. T., & Cho, M. K. (2018). Data mining for health: Staking out the ethical territory of digital phenotyping [Publisher: Nature Publishing Group]. *npj Digital Medicine*, 1(1), 68. <https://doi.org/10.1038/s41746-018-0075-8>
- McKinlay, R. (2016). Technology: Use or lose our navigation skills [Publisher: Nature Publishing Group]. *Nature*, 531(7596), 573–575. <https://doi.org/10.1038/531573a>
- McQuire, S. (2019). One map to rule them all? Google Maps as digital technical object [Publisher: SAGE Publications]. *Communication and the Public*, 4(2), 150–165. <https://doi.org/10.1177/2057047319850192>
- Mei, L., Hu, R., Cao, H., Liu, Y., Han, Z., Li, F., & Li, J. (2020). Realtime mobile bandwidth prediction using LSTM neural network and Bayesian fusion. *Computer Networks*, 182, 107515. <https://doi.org/10.1016/j.comnet.2020.107515>
- Meier, R. (2012). *Professional Android 4 Application Development*. John Wiley & Sons.
- Meng, L., & Reichenbacher, T. (2005). Map-based mobile services. In L. Meng, T. Reichenbacher, & A. Zipf (Eds.), *Map-based Mobile Services: Theories, Methods and Implementations* (pp. 1–10). Springer. https://doi.org/10.1007/3-540-26982-7_1
- Miller, G. (2012). The smartphone psychology manifesto [PMID: 26168460]. *Perspectives on Psychological Science*, 7(3), 221–237. <https://doi.org/10.1177/1745691612441215>
- Miller, H. J., & Bridwell, S. A. (2009). A field-based theory for time geography [Publisher: Routledge]. *Annals of the Association of American Geographers*, 99(1), 49–75. <https://doi.org/10.1080/00045600802471049>

-
- Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Christoph Molnar.
- Montag, C., & Baumeister, H. (Eds.). (2023). *Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-98546-2>
- Montello, D. R. (2002). Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches. *Cartography and Geographic Information Science*, 29(3), 283–304. <https://doi.org/10.1559/152304002782008503>
- Montello, D. R. (2005). Navigation. In *The Cambridge Handbook of Visuospatial Thinking* (pp. 257–294). Cambridge University Press. <https://doi.org/10.1017/CBO9780511610448.008>
- Montello, D. R. (2009). Cognitive Research in GIScience: Recent Achievements and Future Prospects [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-8198.2009.00273.x>]. *Geography Compass*, 3(5), 1824–1840. <https://doi.org/10.1111/j.1749-8198.2009.00273.x>
- Montello, D. R. (2018). *Handbook of Behavioral and Cognitive Geography*. Edward Elgar Publishing.
- Montello, D. R., Fabrikant, S. I., & Davies, C. (2018). Cognitive perspectives on cartography and other geographic information visualizations [ISBN: 9781784717544 Publisher: Edward Elgar Publishing Section: Handbook of Behavioral and Cognitive Geography]. *Handbook of Behavioral and Cognitive Geography*. Retrieved April 13, 2022, from <https://www.elgaronline.com/view/edcoll/9781784717537/9781784717537.00018.xml>
- Montello, D. R., & Sas, C. (2006). Human factors of wayfinding in navigation. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (2nd, pp. 2003–2008). CRC Press/Taylor & Francis, Ltd.
- Münzer, S., Fehring, B. C., & Köhl, T. (2016). Validation of a 3-factor structure of spatial strategies and relations to possession and usage of navigational aids. *Journal of Environmental Psychology*, 47, 66–78. <https://doi.org/10.1016/j.jenvp.2016.04.017>
- Münzer, S., Zimmer, H. D., & Baus, J. (2012). Navigation assistance: A trade-off between wayfinding support and configural learning support. *Journal of Experimental Psychology: Applied*, 18(1), 18–37. <https://doi.org/10.1037/a0026553>
- Münzer, S., Zimmer, H. D., Schwalm, M., Baus, J., & Aslan, I. (2006). Computer-assisted navigation and the acquisition of route and survey knowledge. *Journal of Environmental Psychology*, 26(4), 300–308. <https://doi.org/10.1016/j.jenvp.2006.08.001>
- Murtagh, F., & Contreras, P. (2017). Algorithms for hierarchical clustering: An overview, II [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1219>]. *WIREs Data Mining and Knowledge Discovery*, 7(6), e1219. <https://doi.org/10.1002/widm.1219>

- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1), 1. <https://doi.org/10.1186/s40537-014-0007-7>
- Nori, R., Zucchelli, M. M., Palmiero, M., & Piccardi, L. (2023). Environmental cognitive load and spatial anxiety: What matters in navigation? *Journal of Environmental Psychology*, 88, 102032. <https://doi.org/10.1016/j.jenvp.2023.102032>
- Nyerges, T. L., Mark, D. M., Laurini, R., & Egenhofer, M. (1995). Cognitive Aspects of HCI for GIS: An Introduction. In T. L. Nyerges, D. M. Mark, R. Laurini, & M. J. Egenhofer (Eds.), *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems* (pp. 1–8). Springer Netherlands. https://doi.org/10.1007/978-94-011-0103-5_1
- Ordóñez, F. J., & Roggen, D. (2016). Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition [Number: 1 Publisher: Multidisciplinary Digital Publishing Institute]. *Sensors*, 16(1), 115. <https://doi.org/10.3390/s16010115>
- Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal Ubiquitous Comput.*, 16(1), 105–114. <https://doi.org/10.1007/s00779-011-0412-2>
- Parush, A., Ahuvia, S., & Erev, I. (2007). Degradation in Spatial Knowledge Acquisition When Using Automatic Navigation Systems. In S. Winter, M. Duckham, L. Kulik, & B. Kuipers (Eds.), *Spatial Information Theory* (pp. 238–254). Springer. https://doi.org/10.1007/978-3-540-74788-8_15
- Penn, A. (2003). Space Syntax And Spatial Cognition: Or Why the Axial Line? [Publisher: SAGE Publications Inc]. *Environment and Behavior*, 35(1), 30–65. <https://doi.org/10.1177/0013916502238864>
- Piccardi, L., Risetti, M., & Nori, R. (2011). Familiarity and Environmental Representations of a City: A Self-Report Study [Publisher: SAGE Publications Inc]. *Psychological Reports*, 109(1), 309–326. <https://doi.org/10.2466/01.13.17.PR0.109.4.309-326>
- Prince, S. J. D. (2023, December). *Understanding deep learning* [Google-Books-ID: rvyx-EAAAQBAJ]. MIT Press. <https://mitpress.mit.edu/9780262048644/understanding-deep-learning/>
- Rahmati, A., & Zhong, L. (2013). Studying Smartphone Usage: Lessons from a Four-Month Field Study [Conference Name: IEEE Transactions on Mobile Computing]. *IEEE Transactions on Mobile Computing*, 12(7), 1417–1427. <https://doi.org/10.1109/TMC.2012.127>
- Rao, B., & Minakakis, L. (2003). Evolution of mobile location-based services. *Communications of the ACM*, 46(12), 61–65. <https://doi.org/10.1145/953460.953490>
- Raper, J. (2007). Geographic Relevance. *Journal of Documentation*, 63(6), 836–852.

-
- Raper, J., Gartner, G., Karimi, H., & Rizos, C. (2007a). A critical evaluation of location based services and their potential [Publisher: Taylor & Francis]. *Journal of Location Based Services*, 1(1), 5–45. <https://doi.org/10.1080/17489720701584069>
- Raper, J., Gartner, G., Karimi, H., & Rizos, C. (2007b). Applications of Location-Based Services: A selected review [Publisher: Taylor & Francis]. *Journal of Location Based Services*, 1(2), 89–111. <https://doi.org/10.1080/17489720701862184>
- Raubal, M. (2001). Ontology and epistemology for agent-based wayfinding simulation [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13658810110061171>]. *International Journal of Geographical Information Science*, 15(7), 653–665. <https://doi.org/10.1080/13658810110061171>
- Raubal, M., Miller, H. J., & Bridwell, S. (2004). User-Centred Time Geography for Location-Based Services [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.0435-3684.2004.00166.x>]. *Geografiska Annaler: Series B, Human Geography*, 86(4), 245–265. <https://doi.org/10.1111/j.0435-3684.2004.00166.x>
- Raza, S., & Ding, C. (2019). Progress in context-aware recommender systems — An overview. *Comput. Sci. Rev.*, 31(100), 84–97. <https://doi.org/10.1016/j.cosrev.2019.01.001>
- Reichenbacher, T. (2001). Adaptive concepts for a mobile cartography. *Journal of Geographical Sciences*, 11(1), 43–53. <https://doi.org/10.1007/BF02837443>
- Reichenbacher, T. (2004a). Mobile cartography-adaptive visualisation of geographic information on mobile devices. *Verlag Dr Hut München*.
- Reichenbacher, T. (2004b, February). *Mobile cartography: Adaptive visualisation of geographic information on mobile devices* [Doctoral dissertation, Technische Universität München].
- Reichenbacher, T. (2005). Adaptive egocentric maps for mobile users. In L. Meng, T. Reichenbacher, & A. Zipf (Eds.), *Map-based Mobile Services: Theories, Methods and Implementations* (pp. 141–158). Springer. https://doi.org/10.1007/3-540-26982-7_10
- Reichenbacher, T. (2007). The concept of relevance in mobile maps. In G. Gartner, W. Cartwright, & M. P. Peterson (Eds.), *Location Based Services and TeleCartography* (pp. 231–246). Springer. https://doi.org/10.1007/978-3-540-36728-4_18
- Reichenbacher, T. (2009). Geographic relevance in mobile services. *Proceedings of the 2nd International Workshop on Location and the Web*, 1–4. <https://doi.org/10.1145/1507136.1507146>
- Reichenbacher, T., Aliakbarian, M., Ghosh, A., & Fabrikant, S. I. (2022). Tappigraphy: Continuous ambulatory assessment and analysis of in-situ map app use behaviour. *Journal of Location Based Services*, 16(3), 181–207. <https://doi.org/10.1080/17489725.2022.2105410>
- Reichenbacher, T., & De Sabbata, S. (2011). Geographic relevance: Different notions of geographies and relevancies. *SIGSPATIAL Special*, 3(2), 67–70. <https://doi.org/10.1145/2047296.2047310>

-
- Reid, A. J. (2018). A Brief History of the Smartphone. In A. J. Reid (Ed.), *The Smartphone Paradox: Our Ruinous Dependency in the Device Age* (pp. 35–66). Springer International Publishing. https://doi.org/10.1007/978-3-319-94319-0_2
- Roth, R. E. (2013). Interactive maps: What we know and what we need to know. *Journal of Spatial Information Science*, 6, 59–115. <https://doi.org/10.5311/JOSIS.2013.6.105>
- Roth, R. E., Çöltekin, A., Delazari, L., Denney, B., Mendonça, A., Ricker, B. A., Shen, J., Stachon, Z., & Wu, M. (2024). Making maps & visualizations for mobile devices: A research agenda for mobile-first and responsive cartographic design [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/17489725.2023.2251423>]. *Journal of Location Based Services*, 18(4), 408–478. <https://doi.org/10.1080/17489725.2023.2251423>
- Roth, R. E., Çöltekin, A., Delazari, L., Fonseca Filho, H., Griffin, A., Hall, A., Korpi, J., Lokka, I., Mendonça, A., Ooms, K., & van Elzakker, C. P. (2017). User studies in cartography: Opportunities for empirical research on interactive maps and visualizations [Publisher: Taylor & Francis]. *International Journal of Cartography*, 3(sup1), 61–89. <https://doi.org/10.1080/23729333.2017.1288534>
- Roth, R. E., Griffin, A. L., & Huang, H. (2023). User Experience Design for Mobile Cartography: Update on a ICA Joint Commission Research Agenda [Publisher: Copernicus GmbH]. *Abstracts of the ICA*, 6, 1–2. <https://doi.org/10.5194/ica-abs-6-214-2023>
- Ruginski, I. T., Creem-Regehr, S. H., Stefanucci, J. K., & Cashdan, E. (2019). GPS use negatively affects environmental learning through spatial transformation abilities. *Journal of Environmental Psychology*, 64, 12–20. <https://doi.org/10.1016/j.jenvp.2019.05.001>
- Ruginski, I. T., Giudice, N., Creem-Regehr, S., & Ishikawa, T. (2022). Designing mobile spatial navigation systems from the user’s perspective: An interdisciplinary review. *Spatial Cognition & Computation*. <https://doi.org/10.1080/13875868.2022.2053382>
- Savino, G.-L., Emanuel, N., Kowalzik, S., Kroll, F., Lange, M. C., Laudan, M., Leder, R., Liang, Z., Markhabayeva, D., Schmeißer, M., Schütz, N., Stellmacher, C., Xu, Z., Bub, K., Kluss, T., Maldonado, J., Kruijff, E., & Schöning, J. (2019). Comparing pedestrian navigation methods in virtual reality and real life. *Proceedings of the 2019 International Conference on Multimodal Interaction (ICMI '19)*, 16–25. <https://doi.org/10.1145/3340555.3353741>
- Savino, G.-L., Sturdee, M., Rundé, S., Lohmeier, C., Hecht, B., Prandi, C., Nunes, N. J., & Schöning, J. (2021). MapRecorder: Analysing real-world usage of mobile map applications. *Behaviour & Information Technology*, 40(7), 646–662. <https://doi.org/10.1080/0144929X.2020.1714733>
- Schilit, B., Adams, N., & Want, R. (1994). Context-Aware Computing Applications. *Proceedings of the 1994 First Workshop on Mobile Computing Systems and Applications*, 85–90. <https://doi.org/10.1109/WMCSA.1994.16>

-
- Schiller, J., & Voisard, A. (2004, May). *Location-Based Services*. MORGAN KAUFMANN PUBL Incorporated.
- Schmidt, A., Beigl, M., & Gellersen, H.-W. (1999). There is more to context than location. *Computers & Graphics*, 23(6), 893–901. [https://doi.org/10.1016/S0097-8493\(99\)00120-X](https://doi.org/10.1016/S0097-8493(99)00120-X)
- Schwering, A., Wang, J., Chipofya, M., Jan, S., Li, R., & Broelemann, K. (2014). SketchMapia: Qualitative Representations for the Alignment of Sketch and Metric Maps. *Spatial Cognition & Computation*, 14(3), 220–254. <https://doi.org/10.1080/13875868.2014.917378>
- SettanniMichele, AzucarDanny, & MarengoDavide. (2018). Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis [Publisher: Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA]. *Cyberpsychology, Behavior, and Social Networking*. <https://doi.org/10.1089/cyber.2017.0384>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Silverman, B. W. (2018). *Density estimation for statistics and data analysis* (1st). Routledge. <https://doi.org/10.1201/9781315140919>
- Singhal, M., & Shukla, A. (2012). Implementation of location based services in android using gps and web services. *International Journal of Computer Science Issues (IJCSI)*, 9(1), 237.
- Smith, B., & Mark, D. (1998). Ontology With Human subjects Testing: *The American Journal of Economics and Sociology*, 58(2), 245–312. <https://doi.org/10.1111/j.1536-7150.1998.tb03471.x>
- Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones [Publisher: Proceedings of the National Academy of Sciences]. *Proceedings of the National Academy of Sciences*, 117(30), 17680–17687. <https://doi.org/10.1073/pnas.1920484117>
- Sugimoto, M., Kusumi, T., Nagata, N., & Ishikawa, T. (2022). Online mobile map effect: How smartphone map use impairs spatial memory [Place: United Kingdom Publisher: Taylor & Francis]. *Spatial Cognition and Computation*, 22(1-2), 161–183. <https://doi.org/10.1080/13875868.2021.1969401>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, 2, 3104–3112.

-
- Swienty, O., Reichenbacher, T., Reppermund, S., & Zihl, J. (2008). The role of relevance and cognition in attention-guiding geovisualisation [Publisher: Taylor & Francis]. *The Cartographic Journal*, 45(3), 227–238. <https://doi.org/10.1179/000870408X311422>
- Tolman, E. C. (1948). Cognitive maps in rats and men [Place: US Publisher: American Psychological Association]. *Psychological Review*, 55(4), 189–208. <https://doi.org/10.1037/h0061626>
- Topete, A., He, C., Protzko, J., Schooler, J., & Hegarty, M. (2024). How is GPS used? Understanding navigation system use and its relation to spatial ability. *Cognitive Research: Principles and Implications*, 9(1), 16. <https://doi.org/10.1186/s41235-024-00545-x>
- Tversky, B. (2003). Structures of mental spaces: How people think about space [Place: US Publisher: Sage Publications]. *Environment and Behavior*, 35(1), 66–80. <https://doi.org/10.1177/0013916502238865>
- United Nations Department of Economic and Social Affairs. (2019). *World urbanization prospects 2018: Highlights* [Accessed: 2025-06-23]. United Nations. <https://doi.org/10.18356/6255ead2-en>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010.
- Viswanath, V. K., Hartogenesis, W., Dilchert, S., Pandya, L., Hecht, F. M., Mason, A. E., Wang, E. J., & Smarr, B. L. (2024). Five million nights: Temporal dynamics in human sleep phenotypes [Publisher: Nature Publishing Group]. *npj Digital Medicine*, 7(1), 150. <https://doi.org/10.1038/s41746-024-01125-5>
- Wang, H., Li, Y., Du, M., Li, Z., & Jin, D. (2021). App2Vec: Context-Aware Application Usage Prediction. *ACM Trans. Knowl. Discov. Data*, 15(6), 112:1–112:21. <https://doi.org/10.1145/3451396>
- Watts, A., Ferdous, F., Moore, K. D., & Burns, J. M. (2015). Neighborhood Integration and Connectivity Predict Cognitive Performance and Decline. *Gerontology & Geriatric Medicine*, 1, 2333721415599141. <https://doi.org/10.1177/2333721415599141>
- Weakliam, J., Wilson, D., & Bertolotto, M. (2008). Personalising Map Feature Content for Mobile Map Users. In L. Meng, A. Zipf, & S. Winter (Eds.), *Map-based Mobile Services: Design, Interaction and Usability* (pp. 125–145). Springer. https://doi.org/10.1007/978-3-540-37110-6_7
- Wiener, J. M., Büchner, S. J., & Hölscher, C. (2009). Taxonomy of Human Wayfinding Tasks: A Knowledge-Based Approach. *Spatial Cognition & Computation*, 9(2). <https://doi.org/10.1080/13875860902906496>
- Wilson, D., Bertolotto, M., & Weakliam, J. (2010). Personalizing map content to improve task completion efficiency. *International Journal of Geographical Information Science*, 24(5), 741–760. <https://doi.org/10.1080/13658810903074490>

-
- Winter, S. (2001). Ontology: Buzzword or paradigm shift in GI science? *International Journal of Geographical Information Science*, 15(7), 587–590. <https://doi.org/10.1080/13658810110061207>
- Worboys, M. F. (2001). Nearness relations in environmental space [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/13658810110061162>]. *International Journal of Geographical Information Science*, 15(7), 633–651. <https://doi.org/10.1080/13658810110061162>
- Xu, J., Liu, X., Li, Q., Goldblatt, R., Qin, W., Liu, F., Chu, C., Luo, Q., Ing, A., Guo, L., Liu, N., Liu, H., Huang, C., Cheng, J., Wang, M., Geng, Z., Zhu, W., Zhang, B., Liao, W., ... Schumann, G. (2022). Global urbanicity is associated with brain and behaviour in young people [Publisher: Nature Publishing Group]. *Nature Human Behaviour*, 6(2), 279–293. <https://doi.org/10.1038/s41562-021-01204-7>
- Ye, M., Yin, P., Lee, W.-C., & Lee, D.-L. (2011). Exploiting geographical influence for collaborative point-of-interest recommendation. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, 325–334. <https://doi.org/10.1145/2009916.2009962>
- Zhang, J.-D., & Chow, C.-Y. (2015). GeoSoCa: Exploiting Geographical, Social and Categorical Correlations for Point-of-Interest Recommendations. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 443–452. <https://doi.org/10.1145/2766462.2767711>
- Zhao, S., Ramos, J., Tao, J., Jiang, Z., Li, S., Wu, Z., Pan, G., & Dey, A. K. (2016). Discovering different kinds of smartphone users through their application usage behaviors. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 498–509. <https://doi.org/10.1145/2971648.2971696>
- Zheng, Y., Zhang, L., Xie, X., & Ma, W.-Y. (2009). Mining interesting locations and travel sequences from GPS trajectories. *Proceedings of the 18th international conference on World wide web*, 791–800. <https://doi.org/10.1145/1526709.1526816>
- Zingaro, D., Bartling, M., & Reichenbacher, T. (2023). Exploring Map App Usage Behaviour Through Touchscreen Interactions (Short Paper). <https://doi.org/10.4230/LIPIcs.GIScience.2023.95>
- Zingaro, D., & Reichenbacher, T. (2022). Exploratory analysis of mobile app usage in relation to distance from home. *Zingaro, Donatella; Reichenbacher, Tumasch (2022). Exploratory analysis of mobile app usage in relation to distance from home. In: International Conference on Location Based Services (LBS2022), Munich / Augsburg, Germany, 12 September 2022 - 14 September 2022, 59-67., 59–67.* <https://doi.org/10.5167/uzh-224336>
- Zingaro, D., Reichenbacher, T., Bartling, M., & Fabrikant, S. I. (2024). Exploring the relation between sense of direction and spatial anxiety in everyday mobile map app use (B. Adams, A. L. Griffin, S. Schneider, & G. McKenzie, Eds.) [ISBN: 9783959773300 Number: 315 Place: Saarbrücken Publisher: Schloss Dagstuhl]. *LIPIcs : Leibniz International Proceedings in Informatics*, (315), 15:1–15:8. <https://doi.org/10.4230/LIPIcs.COSIT.2024.15>

A

CHAP-
TER



Appendix

A.1 Questionnaires

The following pages provide the original versions of the questionnaires used in this study.



Participants' Instructions and Questionnaires

Dear Participant,

Welcome to the first phase of the MapOnTap study!

As a participant, you will be asked to complete three questionnaires. Please read the instructions carefully, and if you have any questions, feel free to ask the research team for assistance.

QUESTIONNAIRE 1: SANTA BARBARA SENSE-OF-DIRECTION SCALE

In the following statements, we want to know about your spatial and navigational abilities, preferences, and experiences. For each statement, choose a number between 1 and 7 to show how much you agree. Use '1' if you strongly agree, '7' if you strongly disagree, or some number in between if your agreement is intermediate. If you feel neutral about a statement, you can circle '4'.

QUESTIONNAIRE 2: SPATIAL ANXIETY

In the following questionnaire, we are interested in how you might feel in different imagined navigation scenarios you might encounter in your everyday life. Please rate your level of anxiety on a 5-point scale for each of the thirteen scenarios. Use '5' if you would feel a great deal of anxiety in that circumstance, '4' if you would feel a lot of anxiety, '3' if you would feel a moderate amount, '2' if you would feel just a little, and '0' if you would not feel anxious at all.

Before proceeding with the questionnaires, kindly answer the following two questions by checking the appropriate box.

1) What is your age range?

- ☐ 18-24
- ☐ 25-34
- ☐ 35-44
- ☐ 45-54
- ☐ 55-64
- ☐ 65 or older
- ☐ I prefer not to say

2) Which of the following best describes your gender identity?

- ☐ Male
- ☐ Female
- ☐ Non-binary
- ☐ Transgender
- ☐ I prefer not to say
- ☐ Other (please specify) _____



SANTA BARBARA SENSE-OF-DIRECTION SCALE Participant: _____

The following statements ask you about your spatial and navigational abilities, preferences, and experiences. After each statement, you should circle a number to indicate your level of agreement with the statement. **Circle “1”** if you **strongly agree** that the statement applies to you, **“7”** if you **strongly disagree**, or some number in between if your agreement is intermediate.
Circle “4” if you **neither agree nor disagree**.

1. I am very good at giving directions.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

2. I have a poor memory for where I left things.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

3. I am very good at judging distances.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

4. My “sense of direction” is very good.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

5. I tend to think of my environment in terms of cardinal directions (N, S, E, W).
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

6. I very easily get lost in a new city.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

7. I enjoy reading maps.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

8. I have trouble understanding directions.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

9. I am very good at reading maps.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

10. I don't remember routes very well while riding as a passenger in a car.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

11. I don't enjoy giving directions.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

12. It's not important to me to know where I am.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

13. I usually let someone else do the navigational planning for long trips.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

14. I can usually remember a new route after I have traveled it only once.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree

15. I don't have a very good “mental map” of my environment.
strongly agree 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ strongly disagree



SPATIAL ANXIETY

Participant: _____

Start of Block: 13-item Spatial Anxiety He & Hegarty 2020

SA Instruction Please rate your level of anxiety on a 5-point scale in the following scenarios.

SA_Q1

How anxious do you feel when:

Leaving a store that you have been to for the first time and deciding which way to turn to get to a destination.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q2

How anxious do you feel when:

Finding your way out of a complex arrangement of offices that you have visited for the first time.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q3

How anxious do you feel when:

Pointing in the direction of a place outside that someone wants to get to and has asked you for directions when you are in a windowless room.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**



- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q4

How anxious do you feel when:

Locating your car in a very large parking lot or parking garage.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q5

How anxious do you feel when:

Trying a new route that you think will be a shortcut without the benefits of a map.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q6

How anxious do you feel when:

Finding your way back to a familiar area after realizing you have made a wrong turn and become lost while traveling.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**



SA_Q7

How anxious do you feel when:

Finding your way around in an unfamiliar mall.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q8

How anxious do you feel when:

Finding your way to an appointment in an area of a city or town with which you are not familiar.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q9 How anxious do you feel when:

Finding your way back to your hotel after becoming lost in a new city.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**



SA_Q10

How anxious do you feel when:

Asked to follow directions to a location across town without the use of a map.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q11

How anxious do you feel when:

Trying to get somewhere you have never been to before in the middle of an unfamiliar city.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

SA_Q12 How anxious do you feel when:

Asked to do the navigation planning for a long car trip.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**



SA_Q13 How anxious do you feel when:

Memorizing routes and landmarks on a map for an upcoming event.

- ☐ **A great deal (5)**
- ☐ **A lot (4)**
- ☐ **A moderate amount (3)**
- ☐ **A little (2)**
- ☐ **Not at all (1)**

End of Block: 13-item Spatial Anxiety He & Hegarty 2020

A.2 Classification Categories according to the Google Play Store

Table A.1.: *Application Categories in the Google Play Store ($n = 33$).*

Category	Description
-1	Uncategorized or undefined apps
ART_AND_DESIGN	Creativity tools: drawing, painting, design
AUTO_AND_VEHICLES	Car info, parking, auto-buy/sell, repairs
BOOKS_AND_REFERENCE	E-books, reading apps, dictionaries, references
BUSINESS	Business tools, office, productivity, collaboration
COMICS	Comic readers, manga, graphic novels
COMMUNICATION	Messaging, calls, social chat, email
DATING	Dating and relationship-focused apps
EDUCATION	Learning, language, student, and teaching apps
ENTERTAINMENT	Streaming, movies, fun, and leisure content
EVENTS	Event planning, ticketing, local events
FINANCE	Banking, personal finance, investments, payments
FOOD_AND_DRINK	Restaurant, delivery, cooking, recipe apps
GAME	Games (all genres, cross-category)
HEALTH_AND_FITNESS	Fitness tracking, workouts, nutrition, health
HOUSE_AND_HOME	Home automation, real estate, gardening
LIBRARIES_AND_DEMO	App libraries, SDK demos, sample apps
LIFESTYLE	Life, hobby, daily routine, home
MAPS_AND_NAVIGATION	Maps, GPS, travel navigation, traffic
MEDICAL	Medical info, patient records, telemedicine
MUSIC_AND_AUDIO	Music players, streaming, recording, radio
NEWS_AND_MAGAZINES	News sources, magazines, news aggregators
PARENTING	Parental control, baby/childcare, family tracking
PERSONALIZATION	Wallpapers, launchers, themes, widgets
PHOTOGRAPHY	Cameras, photo editing, gallery apps
PRODUCTIVITY	Calendar, notes, to-do, work tools
SHOPPING	E-commerce, stores, deals, coupons
SOCIAL	Social networks, communities, sharing
SPORTS	Live scores, teams, sports tracking, fantasy
TOOLS	Utilities, file managers, system tools
TRAVEL_AND_LOCAL	Trip planning, hotels, local guides, transport
VIDEO_PLAYERS	Video players, streaming, editing
WEATHER	Weather forecasts, radar, weather widgets

Note. Categories are adapted from the [Google Play Store developer documentation](#). The value -1 indicates applications that the MapOnTap application was unable to categorize.

A.3 Applications Used for Navigation-only Filtered Dataset

Table A.2.: *Mobile Map Applications used when filtering MoT II Dataset for Navigation only*

Description	Generic Name
Advanced Battery Range Planner for EVs	ABRP
South Tyrol Public Transport	AltoAdige To Go
Bike Navigation	Geovelo
Comprehensive Mapping Service	Google Maps
Swiss Federal Railways	SBB Mobile
Swiss National Topography	SwissTopo
Switzerland Mobility Solutions	Switzerland Mobility
German Railway Navigator	DB Navigator
Paris Public Transport	RATP
Netherlands Railways	NS
Italian Train Timetable	ProntoTreno
Budapest Public Transportation	BKK FUTÁR
Austrian Federal Railways	ÖBB
Austrian Railways (HAFAS)	ÖBB (HAFAS)
Czech Republic Mapping Service	Mapy.cz
Zurich City Plan	ZüriPlan
Global Navigation	HERE Maps
Zurich Public Transport	ZVV
High-Speed Train Service in Italy	Italo Treno
UK and International Train Tickets	Trainline
OpenStreetMap Navigation	OsmAnd
Pakistan Railways Official	Pakistan Railways
Public Transport Finder	Öffi
South Tyrol Hiking	Sentres
Milan Public Transport	ATM Milano
Research Center Jülich Campus App	FZJ App

Note. Applications were selected from the *Maps & Navigation* and *Travel & Local* categories of the Google Play Store and restricted to those directly related to navigation or travel planning.

A.4 Supplementary Descriptive Statistics (MoT II, Raw Data)

This section reports descriptive statistics for raw tap and GPS data in MoT II (before filtering for navigation-only sessions).

GPS Data (MoT II, Raw Sample)**Table A.3.:** *Participant-Level GPS Descriptive Statistics* (filtered sample, $N = 61$ participants)

Variable	Mean	SD	Mdn	Min	Max
Sessions per participant	1,210.92	1,218.23	852.00	3	5,552
Points per participant	181,881.70	194,828.00	165,720.00	76	1,416,351
Session length (s)	15,420.12	74,096.85	1,155.71	16.52	505,057.97

Note. Statistics summarize participant-level distributions ($N = 61$). Session length is expressed in seconds. On average, participants had a mean session length of about 15,420 s (≈ 257 min or 4.3 hr). The median session length was 1,156 s (≈ 19 min), while the maximum observed value corresponds to about 505,058 s (≈ 140 hr or 5.8 days). Values are rounded to two decimal places.

Tappigraphy Data (MoT II, Raw Sample)**Table A.4.:** *Participant-Level Tappigraphy Descriptive Statistics* (MoT II, $N = 61$ participants)

Variable	Mean	SD	Mdn	Min	Max
Sessions per participant	2,366.59	3,820.54	1,652.00	4.00	30,068.00
Mean taps per session	52.16	33.23	46.53	0.00	213.19
Mean session length (s)	121.75	89.05	104.24	0.00	529.61

Note. Each row summarizes the distribution of participant-level values ($N = 61$). Sessions per participant refers to the number of sessions recorded per participant. Mean taps per session is the participant-level mean of taps per session. Mean session length (s) is the participant-level mean of session length in seconds. Values are rounded to two decimals.

A.5 GPS Data Alignment Supplementary Information

This section provides supplementary information on the alignment of GPS and tap data. For each participant, we report the total number of GPS points used in alignment (data available for participants $N = 60$). After alignment, we summarize the number of tap events that were successfully matched to GPS points. All computations were performed on the raw GPS dataset before filtering for navigation-only sessions. More details on the alignment procedure are available in the project report [GitLab repository](#).

Table A.5.: *GPS Data Descriptive Statistics Before and After Alignment ($N = 60$ participants)*

Variable	Mean	SD	Mdn	Min	Max
GPS points	142,966.00	104,664.00	144,339.00	66.00	504,364.00
Earliest date	2023-04-04	—	—	2023-04-04	2023-04-04
Latest date	2023-08-02	—	—	2023-08-02	2023-08-02
Matched taps	105,470.00	128,590.00	77,630.00	248.00	949,245.00

Note. Statistics are computed across $N = 60$ participants. Dates are reported in ISO format (YYYY-MM-DD). One participant without valid GPS records was excluded, explaining the reduction from $N = 61$ to $N = 60$. Values are rounded to two decimals where applicable.

A.6 Urban Form: Additional Analyses

This appendix provides additional information on the characterization of urban form in the study in Section 4.2.2. Specifically, it reports: (i) the distribution of participants by country, (ii) the classification of GPS points in urban versus rural contexts using the Settlement Model (SMOD), and (iii) sensitivity analyses of buffer zones used for the extraction of urban form metrics, including descriptive statistics for the final 500 m buffer selected for the main analyses.

Distribution of Participants by Country

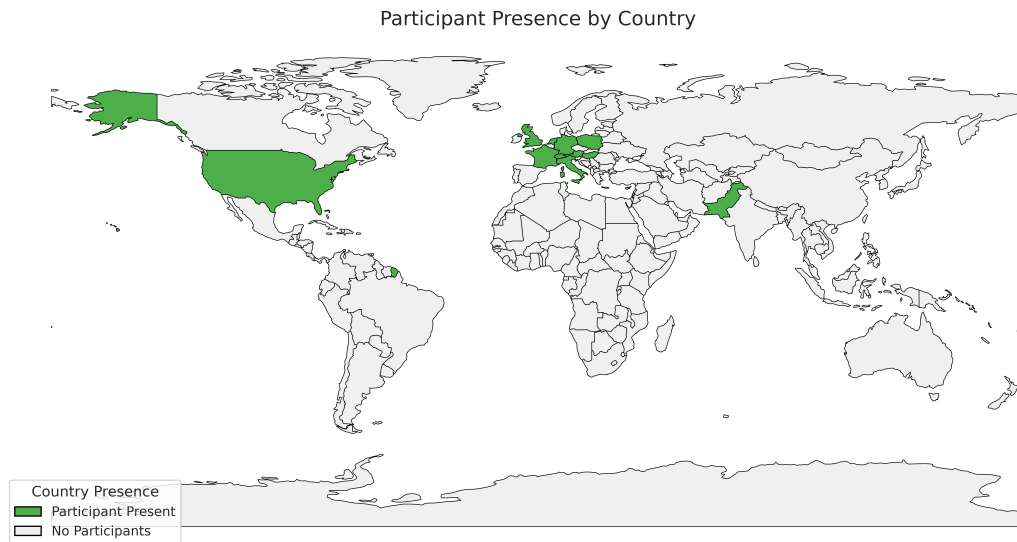


Figure A.1.: *Distribution of participants by country. The dataset includes participants from 11 countries: Switzerland (23), Italy (9), France (3), Austria (3), United Kingdom (2), United States (1), Germany (1), Hungary (1), Pakistan (1), Poland (1), and the Netherlands (1).*

To ensure regional comparability, participants outside Europe (United States and Pakistan) were excluded. This exclusion did not alter the direction of the observed effect: the relationship between orientation entropy and tap rate remained negative ($\beta = -0.060 \rightarrow -0.024$). This confirms that the reported association is not driven by regional outliers and is robust under conservative conditions.

Urban vs. Non-Urban Point Classification

To contextualize GPS locations within settlement environments, we used the **Settlement Model (SMOD)** from the Global Human Settlement Layer (GHSL), available at: <https://ghsl.jrc.ec.europa.eu/download.php?ds=SMOD>.

SMOD provides a globally consistent classification of land into settlement typologies, derived from population density and built-up area data. The version employed here corresponds to the year 2020, with a spatial resolution of 1 km. Each raster cell is assigned a categorical code that distinguishes rural and urban contexts. In this study, we aggregated these codes into two broad categories:

- **Urban BROAD (≥ 21):** urban centers, dense urban clusters, or semi-dense urban clusters.
- **Rural (12–20):** rural clusters or dispersed rural settlements.

Cells corresponding to water bodies, undefined areas, or missing values were grouped as **Unknown/NaN**.

Table A.6.: *SMOD (2020, 1 km) classification of participant session locations.*

Category	Count	Percentage
Total rows scanned	65,603	—
Rows with coordinates	43,645	66.53%
Rows classified (of coord rows)	43,645	100.00%
Urban BROAD (≥ 21)	39,441	90.37%
Rural (12–20)	3,004	6.88%
Unknown / water / NaN	1,200	2.75%

Sensitivity Analysis for the Buffer Zone

Because centrality values are scale-dependent and sensitive to the definition of the underlying street network, we extracted localized subgraphs from OpenStreetMap by buffering the median GPS coordinate of each tile. We tested three radii (200 m, 500 m, 1000 m) to ensure that the results were not sensitive to buffer size. These thresholds were selected to reflect different scales of interaction with the built environment: 200 m captures the immediate street frontage around a location, 500 m approximates a typical pedestrian catchment radius and is widely used in urban morphology research, and 1000 m represents a broader neighborhood scale. By comparing across these three levels, we ensured that our centrality metrics were robust to reasonable variations in the definition of the localized road network.

In all metrics, the values were highly stable on all scales ($\rho \geq 0.99$, $\Delta\text{means} \leq 0.03\%$). For the main analyses, we adopted a buffer of 500 m, since this scale corresponds to a typical pedestrian catchment and is widely used in urban morphology research.

Table A.7.: *Sensitivity of urban form metrics to buffer size. Descriptive statistics are reported for 200 m, 500 m, and 1000 m buffers.*

Buffer (m)	Metric	Mean	SD	Mdn	IQR
200	Betweenness	0.0658	0.0727	0.0428	0.0465
200	Circuitry	1.0772	0.1217	1.0587	0.0386
200	Closeness	0.0928	0.0642	0.0751	0.0413
200	Degree	5.1474	0.4299	5.1843	0.4297
200	Orientation entropy	0.9474	0.0794	0.9695	0.0787
500	Betweenness	0.0658	0.0727	0.0429	0.0463
500	Circuitry	1.0772	0.1217	1.0587	0.0386
500	Closeness	0.0928	0.0642	0.0751	0.0413
500	Degree	5.1474	0.4299	5.1843	0.4297
500	Orientation entropy	0.9474	0.0794	0.9695	0.0787
1000	Betweenness	0.0658	0.0727	0.0431	0.0467
1000	Circuitry	1.0772	0.1217	1.0587	0.0386
1000	Closeness	0.0928	0.0642	0.0751	0.0413
1000	Degree	5.1474	0.4299	5.1843	0.4297
1000	Orientation entropy	0.9474	0.0794	0.9695	0.0787

A.7 Additional Notes for the Exploratory Analyses Results

This section provides supplementary material for the exploratory analyses reported in Section 4.2.1.

Note on bin cut-points for distance-from-home analyses.

As described in Zingaro and Reichenbacher (2022), the near-home cut-off point at 3.5 km was chosen based on the first local maximum in the pooled participant-level KDEs, which typically occurs within 0–3.5 km of the estimated home location. This range also corresponds to distances that can be walkable in ~ 30 to 45 minutes, consistent with prior definitions of local-scale activity spaces.

The mid-range bin of 100 to 200 km was selected to capture the secondary KDE peak observed for many participants, typically corresponding to inter-city travel within the same country. Distances in this range often involve different transport modes (e.g., regional train, highway travel) and align with transitions in navigation context reported in mobility

studies. These bin definitions were applied consistently for statistical comparisons between participants.

Table A.8.: *Correlation Coefficients between Santa Barbara Sense of Direction scores (SBSD), Spatial Anxiety (SA) scores, and the extracted variable linked to interactivity behaviour. Significant correlations are highlighted in bold, with asterisks denoting p-value thresholds (* $p < 0.05$).*

Variable	SBSOD score	SA Score
switches_all	-0.10 (0.482)	0.15 (0.270)
switches_from	0.07 (0.615)	0.06 (0.661)
switches_to	0.04 (0.752)	0.12 (0.384)
switches_map_all	0.13 (0.351)	0.08 (0.538)
prop_switch_maps	-0.02 (0.898)	0.03 (0.813)
switches_from_prop	0.09 (0.509)	-0.24 (0.070)
switches_to_prop	-0.03 (0.798)	0.05 (0.736)
total_time_spent	0.11 (0.414)	0.17 (0.203)
total_time_spent_on_map	0.14 (0.297)	-0.07 (0.604)
prop_time_on_map	0.02 (0.859)	-0.03 (0.800)
prop_map_count	-0.02 (0.865)	-0.16 (0.242)
SBDS_score	1 (-)	-0.33 (0.0131*)
SA_Score	-0.33 (0.0131*)	1 (-)

Table A.9.: *Correlation Coefficients for the Male Dataframe, between Santa Barbara Sense of Direction scores (SBSOD), Spatial Anxiety (SA) scores, and the extracted variable linked to interactivity behavior. Significant correlations are highlighted in bold, with asterisks denoting p-value thresholds (* $p < 0.05$).*

Variable	Correlation with SBSOD	P-value
switches_all	-0.20	(0.28)
switches_from	-0.04	(0.83)
switches_to	0.38	(0.0369*)
switches_map_all	0.14	(0.468)
prop_switch_maps	-0.11	(0.548)
switches_from_prop	0.03	(0.855)
switches_to_prop	0.20	(0.288)
total_time_spent	0.03	(0.868)
total_time_spent_on_map	0.31	(0.09)
prop_time_on_map	0.01	(0.96)
SA_Score	-0.30	(0.0987)
Variable	Correlation with SA Score	P-value
switches_all	0.03	(0.88)
switches_from	0.13	(0.482)
switches_to	0.21	(0.265)
switches_map_all	0.24	(0.193)
prop_switch_maps	0.14	(0.444)
switches_from_prop	-0.11	(0.573)
switches_to_prop	0.18	(0.329)
total_time_spent	0.14	(0.444)
total_time_spent_on_map	0.03	(0.87)
prop_time_on_map	0.03	(0.86)

Table A.10.: *Correlation Coefficients for the Female Dataframe, between Santa Barbara Sense of Direction scores (SBSD), Spatial Anxiety (SA) scores, and the extracted variable linked to interactivity behaviour. Significant correlations are highlighted in bold, with asterisks denoting p-value thresholds (* $p < 0.05$).*

Variable	Correlation with SBSD score	P-value
switches_all	0.09	0.683
switches_from	0.11	0.586
switches_to	-0.23	0.259
switches_map_all	-0.01	0.949
prop_switch_maps	0.09	0.676
switches_from_prop	0.14	0.494
switches_to_prop	-0.31	0.131
total_time_spent	0.28	0.180
total_time_spent_on_map	-0.16	0.449
prop_time_on_map	-0.001	0.996
SA Score	-0.29	0.164
Variable	Correlation with SA_score	
switches_all	0.48	0.103
switches_from	0.02	0.921
switches_to	-0.10	0.624
switches_map_all	0.06	0.788
prop_switch_maps	0.05	0.806
switches_from_prop	-0.40	0.0492*
switches_to_prop	-0.18	0.400
total_time_spent	0.13	0.539
total_time_spent_on_map	-0.11	0.603
proportion_time_on_map	-0.10	0.637

A.8 Additional Notes for the Main Analyses

This section provides supplementary details for the results reported in Section 4.2.2.

U-map Cluster Summary

The nonparametric Kruskal–Wallis test indicated robust differences between clusters for traits and all session metrics (see Table 4.14 in Section 4.2.2).

Table A.11.: *Kruskal–Wallis tests across DBSCAN clusters (clusters 0–5; noise excluded).*

Feature	H-statistic	<i>p</i> -value
SA score	82.419	2.62×10^{-16}
SBSOD score in raw tables;	27.253	5.09×10^{-5}
CV-ITI (log)	1,247.322	1.65×10^{-267}
Taps/s (log)	951.836	1.61×10^{-203}
Transitions/min (log)	534.669	2.62×10^{-113}

Regression Analyses for Inter-Map Interval (IMI)

For SBSOD, the regression model explained only 2.2% of the variance ($R^2 = 0.022$, $p = 0.285$). IMI did not significantly predict the SBSOD scores ($\beta = 0.800$, 95% CI $[-0.685, 2.284]$).

For SA, the model explained 5.6% of the variance ($R^2 = 0.056$, $p = 0.083$). The IMI coefficient was negative ($\beta = -0.818$, 95% CI $[-1.745, 0.110]$), but this effect did not reach significance.

A priori power analysis ($N = 55$, $m = 1$) indicated approximately 0.79 power to detect a moderate effect ($f^2 = 0.15$, $\alpha = 0.05$). Thus, moderate effects would likely have been detected. The null findings suggest that the true effects of IMI on SBSOD or SA are small.

Including gender did not materially change the interpretation. For SBSOD, the model explained 6.0% of the variance ($R^2 = 0.060$, $p = 0.199$), with neither IMI ($p = 0.222$) nor gender ($p = 0.150$) significant. For SA, the model explained 12.1% of the variance ($R^2 = 0.121$, $p = 0.035$). IMI showed a borderline negative effect ($\beta = -0.909$, 95% CI $[-1.817, 0.000]$, $p = 0.050$). Gender also approached significance ($\beta = -0.317$, $p = 0.055$), with men tending to report lower SA.

For two predictors ($N = 55$, $m = 2$), the power was approximately 0.68, below the conventional threshold of 0.80. Consequently, the results should be interpreted with caution.

Regression Analyses Session Initiation Rate (SIR)

For SBSOD, the regression model explained 7.1% of the variance ($R^2 = 0.071$, $p = 0.050$). The SIR coefficient was negative ($\beta = -56.91$, 95% CI $[-113.81, -0.02]$), suggesting that participants who initiated navigation sessions reported a lower sense of direction. Given the borderline value p , this result should be interpreted with caution.

For SA, the model explained only 0.2% of the variance ($R^2 = 0.002$, $p = 0.730$). The coefficient was positive but imprecise ($\beta = 6.49$, 95% CI $[-31.00, 43.97]$), providing no evidence of an association between SIR and spatial anxiety.

Including gender did not change the interpretation. For SBSOD, the model explained 8.2% of the variance ($R^2 = 0.082$, $p = 0.107$), with SIR no longer significant ($p = 0.099$). The gender was not related to SBSOD ($p = 0.419$). For SA, the model explained 5.3% of the

variance ($R^2 = 0.053$, $p = 0.243$). SIR was not significant ($p = 0.885$). The gender showed a borderline negative effect ($\beta = -0.29$, $p = 0.101$), with men tending to report lower SA.

For two predictors ($N = 55$, $m = 2$), the power was about 0.68, suggesting that the models were underpowered for moderate to small effects.

Equations: Moderation Models

The full equations for models A.1–A.5 (moderation analyses) are provided below. These correspond to the results reported in Section 4.2.2 and illustrated in Figures ??.

Model 3 – SA × SBSOD (Interaction Frequency)

$$\begin{aligned} \text{tap_rate}_z = & \beta_0 + \beta_1 \cdot \text{SA}_z + \beta_2 \cdot \text{SBSOD}_z + \beta_3 \cdot (\text{SA}_z \cdot \text{SBSOD}_z) \\ & + \beta_4 \cdot \text{orientation_entropy}_z + \beta_5 \cdot \text{circuitry_log}_z + \beta_6 \cdot \text{degree_mean}_z + \beta_7 \cdot \text{Gender_M} + \varepsilon \end{aligned} \quad (\text{A.1})$$

Model 4 – Entropy × SBSOD (Interaction Frequency)

$$\begin{aligned} \text{tap_rate}_z = & \beta_0 + \beta_1 \cdot \text{orientation_entropy}_z + \beta_2 \cdot \text{SBSOD}_z + \beta_3 \cdot (\text{orientation_entropy}_z \cdot \text{SBSOD}_z) \\ & + \beta_4 \cdot \text{SA}_z + \beta_5 \cdot \text{Gender_M} + \beta_6 \cdot \text{degree_mean}_z + \beta_7 \cdot \text{circuitry_log}_z + \varepsilon \end{aligned} \quad (\text{A.2})$$

Model 5 – Circuitry × SA (Interaction Regularity)

$$\begin{aligned} \text{cv_iti}_z = & \beta_0 + \beta_1 \cdot \text{circuitry_log}_z + \beta_2 \cdot \text{SA}_z + \beta_3 \cdot (\text{circuitry_log}_z \cdot \text{SA}_z) \\ & + \beta_4 \cdot \text{SBSOD}_z + \beta_5 \cdot \text{Gender_M} + \beta_6 \cdot \text{orientation_entropy}_z + \beta_7 \cdot \text{degree_mean}_z + \varepsilon \end{aligned} \quad (\text{A.3})$$

Model 6 – Closeness × SBSOD (Interaction Regularity)

$$\begin{aligned} \text{cv_iti}_z = & \beta_0 + \beta_1 \cdot \text{closeness_log}_z + \beta_2 \cdot \text{SBSOD}_z + \beta_3 \cdot (\text{closeness_log}_z \cdot \text{SBSOD}_z) \\ & + \beta_4 \cdot \text{SA}_z + \beta_5 \cdot \text{orientation_entropy}_z + \beta_6 \cdot \text{degree_mean}_z + \beta_7 \cdot \text{Gender_M} + \varepsilon \end{aligned} \quad (\text{A.4})$$

Model 7 – Closeness × Gender (Interaction Regularity)

$$\begin{aligned} \text{cv_iti}_z = & \beta_0 + \beta_1 \cdot \text{closeness_log}_z + \beta_2 \cdot \text{Gender_M} + \beta_3 \cdot (\text{closeness_log}_z \cdot \text{Gender_M}) \\ & + \beta_4 \cdot \text{SA}_z + \beta_5 \cdot \text{SBSOD}_z + \beta_6 \cdot \text{orientation_entropy}_z + \beta_7 \cdot \text{degree_mean}_z + \varepsilon \end{aligned} \quad (\text{A.5})$$

A.9 Context-Aware LSTM Fusion Model

Behavioral Interaction Branch

The preprocessing of tappigraphy data for LSTM-based behavioral modeling involved a multistep pipeline, designed to maximize data quality, relevance, and suitability for sequential learning. Each step is described in the following, with explicit reference to data transformations and quality control measures.

Session Filtering and Data Cleaning

- **Removing empty sessions:** Of the initial 144,362 phone sessions, all without recorded user interaction events were excluded ($n = 83,056$), leaving 61,306 valid sessions for analysis.
- **Excluding MapOnTap sessions:** Interactions originating from the MapOnTap application itself were removed to avoid circularity to focus the analysis on user-initiated app participation ($n = 2,654$), leaving ($n = 58,652$) sessions.
- **Exclusion of underrepresented categories:** The original tappigraphy logs included 33 distinct application categories. To reduce data sparsity and enhance category-level learning, underrepresented categories were excluded, reducing the final set to 27 categories and 56,467 sessions ($n = 2,185$ removed). A full list of excluded categories is provided in the Appendix.

Temporal Aggregation and Sequence Construction

- **Aggregation to 1-minute intervals:** The raw data consisted of time-stamped tap events annotated with app categories. For each session, the events were aggregated into 1-minute intervals. This granularity was chosen after systematic comparison with alternatives (5s, 10s, 30s, 2min, 5min), with the 1-minute window providing the best trade-off between temporal resolution and noise reduction.
- **Conversion to structured session tensors:** For each session, a tensor was constructed with rows representing consecutive 1-minute intervals ($t = 1, 2, \dots, n$) and columns corresponding to the 27 behavioral categories. The (i, j) -th entry of a tensor denotes the number of tap events in the category j during the minute i . This yielded a set of variable length tensors, each encoding the temporal distribution of category-specific app use within a session.
- **Dimension of session data:** The preprocessed dataset comprised 56,438 session tensors (one per phone session), each with a variable number of rows (depending on the duration of the session) and 27 columns (one per category).

Binary Labeling

For each session tensor, a binary label was assigned:

$$y = \begin{cases} 1 & \text{if the session includes a map-related tap} \\ 0 & \text{otherwise} \end{cases}$$

Feature Normalization and Differencing

To prepare the session tensors for sequential modeling, we applied two preprocessing steps: feature normalization and temporal differencing.

Global normalization. To ensure comparability across categories and stabilize training, all tensors were normalized feature-wise. The global mean (μ_j) and standard deviation (σ_j) for each category j (column) were calculated in all time intervals of all sessions. Then each value x_{ij} in a session tensor was standardized as

$$x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j},$$

where x'_{ij} denotes the normalized tap count for category j in the interval i .

Temporal differencing. To highlight changes in activity over time rather than absolute levels, we applied a differencing step across consecutive intervals. For each session tensor with more than one interval, the difference between each row and its predecessor was calculated:

$$x''_{t,j} = \begin{cases} x'_{1,j}, & \text{if } t = 1, \\ x'_{t,j} - x'_{t-1,j}, & \text{if } t > 1, \end{cases}$$

where $x''_{t,j}$ denotes the differenced normalized value for category j in the interval t . The first interval retained its normalized value.

Train/Validation/Test Splitting

Data splitting protocol: To ensure robust evaluation and prevent information leakage, the dataset was randomly split at the session level into training sets (80%), validation (10%), and test (10%) sets. The splitting was performed after all the preprocessing and shuffling steps and the splits were stored for reproducibility.

Window Extraction and Data Augmentation

To construct inputs for the LSTM model and increase sample diversity, we employed a **sliding window** approach with systematic data augmentation applied **only to the training set**. For each behavioral sequence (session tensor) of length n , all possible contiguous windows of length $T = 10$ (the LSTM input length) were extracted using stride $= 1$. Each window is denoted as

$$X_{\text{seq}} \in \mathbb{R}^{T \times F},$$

where F denotes the number of behavioral features (that is, the number of taps in the app category).

To improve generalization and robustness to noise in real-world interaction patterns, we applied three types of stochastic augmentation during training.

- **Time-step masking:** Tap counts for randomly selected time steps were set to zero to simulate missing activity, with the probability of masking $\text{mask_prob} \in \{0.05, 0.10, 0.15, 0.20\}$.
- **Gaussian noise injection:** Each feature value was perturbed by additive Gaussian noise,

$$\tilde{X}_{\text{seq}} = X_{\text{seq}} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2),$$

with $\sigma \in \{0.05, 0.10, 0.15, 0.20, 0.30\}$.

- **Feature dropout:** All channels in the app categories were randomly dropped within a window, with $\text{feat_drop_prob} \in \{0.00, 0.05, 0.10, 0.15\}$.

The full factorial design ($4 \times 5 \times 4$) yielded 80 unique augmentation configurations, each trained and evaluated under the same optimization regime. The configuration achieving the highest F1 validation score was selected as the final model, with parameters $\text{mask_prob} = 0.20$, $\sigma = 0.15$, and $\text{feat_drop_prob} = 0.00$. All test set metrics reported for the augmented model correspond to this configuration. Additional details of the implementation are provided in [GitLab repository of the project](#).

Ablation: Hidden Size (oversampling only)

To isolate the effect of model capacity, we conducted an ablation experiment that varied the hidden size of the LSTM while keeping all other training settings fixed. In this analysis, models were trained using oversampling only (no augmentation), in order to avoid the confounding influence of stochastic data transformations.

Let N_{pos} and N_{neg} denote the number of positive and negative windows in the training data, and let $N_{\text{max}} = \max(N_{\text{pos}}, N_{\text{neg}})$. Each window i with label $y_i \in \{0, 1\}$ received a sampling weight:

$$w_i = \begin{cases} \frac{N_{\text{max}}}{N_{\text{pos}}}, & \text{if } y_i = 1, \\ \frac{N_{\text{max}}}{N_{\text{neg}}}, & \text{if } y_i = 0. \end{cases}$$

Then a `WeightedRandomSampler` was used to generate training batches, with the total number of samples per epoch set to $1.25 \times N_{\max}$ (capped at 8,000,000 samples per epoch) to avoid excessive duplication.

Oversampling was applied strictly to the training set, while the **validation and test sets** preserved the empirical class ratio to provide a realistic estimate of predictive performance.

Ablation: Hidden size (oversampling only)

In this analysis, models were trained using oversampling only (no augmentation), in order to isolate the impact of hidden size without the confounding influence of stochastic data transformations. To examine the effect of LSTM capacity on classification performance, we performed an ablation experiment that varied the hidden size while keeping all other training settings fixed. In this analysis, models were trained using oversampling only (no augmentation), in order to isolate the impact of hidden size without the confounding influence of stochastic data transformations.

Table A.12 reports test-set results for hidden sizes of 16, 32, and 64 units, evaluated at a decision threshold of $p = 0.5$.¹

The test set comprised 5,290 windows (494 positives; 4,796 negatives). As expected, increasing the hidden size improved recall, reaching 0.97 for 64 units; however, this came at the cost of lower precision, as larger models generated more false positives. ROC-AUC values remained stable in all configurations (0.76–0.77), suggesting that the model capacity influenced the trade-off between precision and recall rather than overall separability.

Table A.12.: *Hidden-size comparison (oversampling only; threshold $p = 0.5$). Test set $N = 5,290$; positives = 494, negatives = 4,796.*

Hidden	Precision	Recall	F1	ROC-AUC	FP	TP	TN	FN
16	0.1200	0.8725	0.211	0.760	3,160	431	1,636	63
32	0.1161	0.8968	0.206	0.770	3,373	443	1,423	51
64	0.1056	0.9717	0.191	0.770	4,064	480	732	14

Urban Context and Trait Branch

The urban context of each session was characterized by graph- and information-theoretic features of the surrounding street network, reflecting environmental structures known to affect navigation and spatial behavior. In parallel, cognitive and emotional traits were included as predictors to capture inter-individual variability in navigational strategy and app reliance. Data sources included:

- GPS-aligned tappigraphy logs linked with urban form metrics, extracted using OSMnx-based pipelines.
- The responses of the participant questionnaire providing the SBSOD and SA scores, and sex (binary coded).

¹These runs are for the oversampling-only condition and are not directly comparable to the augmented `best_0.2_0.15_0.0.pt` model.

Context Vector Construction

For each session, urban metrics were averaged across GPS-aligned events, then concatenated with participant-level traits, producing an eight-dimensional context vector.

$$\mathbf{x}_{\text{ctx}}^{(j)} = [\text{betweenness}, \text{circuitry}, \text{closeness}, \text{degree}, \text{orientation entropy}, \text{SBSOD}, \text{SA}, \text{gender}]_j.$$

Only sessions with complete context data were retained ($N = 35,089$).

Autoencoder Compression

To reduce redundancy and dimensionality, a fully connected autoencoder was trained to compress $\mathbf{x}_{\text{ctx}} \in \mathbb{R}^8$ into a three-dimensional latent code $\mathbf{z}_{\text{ctx}}$. The encoder was a two-layer MLP (8 16 3, ReLU activations), mirrored by a decoder (3 16 8). Training minimized reconstruction MSE with Adam (lr = 10^{-3} , batch size 32, 100 epochs). Only encoder weights were retained for downstream use.

Fusion Head

Each session window was represented by (a) the behavioral embedding derived from LSTM $\mathbf{h}_{\text{seq}} \in \mathbb{R}^{16}$ and (b) the context embedding $\mathbf{z}_{\text{ctx}} \in \mathbb{R}^3$ (or the raw 8D context vector). These were concatenated:

$$\mathbf{z}_{\text{fused}} = [\mathbf{h}_{\text{seq}}; \mathbf{z}_{\text{ctx}}] \in \mathbb{R}^{19} \quad (\text{AE case}),$$

or $\mathbb{R}^{24}$ in the raw case.

The fusion head was a two-layer MLP (19/24 8 4, ReLU) with a final sigmoid output:

$$\hat{p} = \sigma(\mathbf{w}_{\text{out}}^\top \text{ReLU}(\mathbf{W}_2 \text{ReLU}(\mathbf{W}_1 \mathbf{z}_{\text{fused}} + \mathbf{b}_1) + \mathbf{b}_2) + b_{\text{out}}).$$

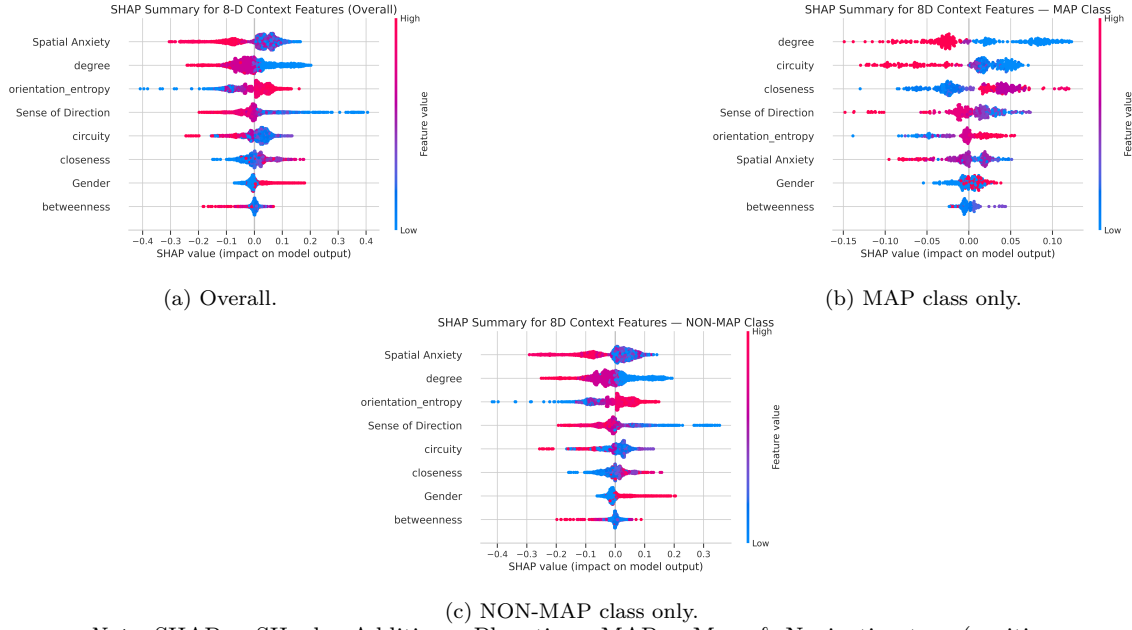
Training Procedures

- **Autoencoder:** MSE loss, Adam (10^{-3}), batch size 32, 100 epochs.
- **LSTM sequence encoder:** Trained with BCEWithLogitsLoss, Adam (10^{-4}), batch size 32, early stopping of validation loss.
- **Fusion head:** BCEWithLogitsLoss, Adam (10^{-3}), batch size 32. LSTM and AE encoders were frozen during training.

Explainability Models

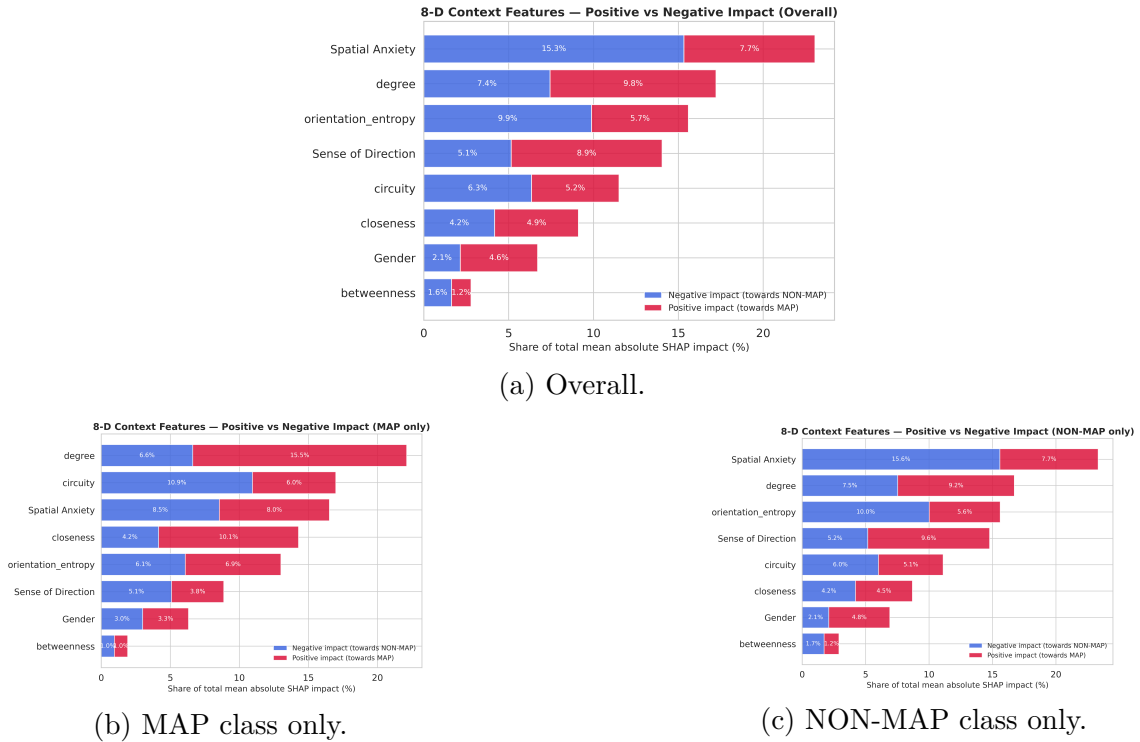
Finally, we report explainability results for the 8D Fusion Model (Without AE) and examine the contribution of contextual elements through ablation.

Figure A.2.: *SHAP summary plots for context inputs in the fusion head (Without AE configuration). Features are ordered by mean absolute SHAP value, with the most influential at the top.*



Note. SHAP = SHapley Additive exPlanations; MAP = Maps & Navigation taps (positive class); NON-MAP = all other taps (negative class). Positive SHAP values increase the probability of predicting a map tap; negative values decrease it. Colors encode feature values (blue = low, red = high).

Figure A.3.: Positive versus negative SHAP contributions for context features in the fusion head (Without AE configuration).



Note. SHAP = SHapley Additive exPlanations; MAP = Maps & Navigation taps (positive class); NON-MAP = all other taps (negative class). Positive SHAP values increase the probability of predicting a map tap, while negative values decrease it. Colors encode feature values (blue = negative impact, red = positive impact).

Fusion Head Ablation Analyses: With-AE vs. 8D (Without-AE)

We conducted two sets of ablation experiments to examine the contribution of contextual inputs. First, we compared the configurations of the fusion head with autoencoder compression (With-AE): (i) **Full AE** (traits and urban features compressed), (ii) **No Traits** (urban form features only), and (iii) **No Urban** (individual traits only). Second, we ran parallel ablations using the full 8D feature set without compression (Without-AE): (i) **Full 8D**, (ii) **No Traits**, and (iii) **No Urban**.

Table A.13 and Table A.14 report the overall and per-class performance of the With-AE ablations. Table A.15 and Table A.16 present the corresponding results for the 8D model without AE. Figure A.4 highlights the recall differences between the 8D ablations.

Table A.13.: Overall development set performance for fusion head ablations (With-AE)

Configuration	Accuracy	Precision	Recall	F1	ROC-AUC
Full AE	0.805	0.258	0.570	0.356	0.754
No Traits (Urban only)	0.836	0.299	0.548	0.387	0.768
No Urban (Traits only)	0.794	0.244	0.567	0.341	0.785

Note. Results are based on ablation experiments with autoencoder-compressed contextual inputs. Values are rounded to three decimals.

Table A.14.: *Per-class development set performance for fusion head ablations (With-AE)*

Configuration	Class	Accuracy	Precision	Recall	F1
Full AE	MAP	0.805	0.258	0.570	0.356
	NON-MAP	0.805	0.949	0.830	0.885
No Traits (Urban only)	MAP	0.836	0.299	0.548	0.387
	NON-MAP	0.836	0.949	0.866	0.906
No Urban (Traits only)	MAP	0.794	0.244	0.567	0.341
	NON-MAP	0.794	0.948	0.817	0.878

Note. Metrics reported separately for MAP and NON-MAP classes. Values are rounded to three decimals.

Table A.15.: *Overall model performance in ablation analyses (8D, Without-AE)*

Model	Accuracy	Precision	Recall	F1	ROC-AUC
Full (8D)	0.788	0.255	0.652	0.366	0.792
Urban only	0.849	0.323	0.553	0.408	0.792
Traits only	0.804	0.247	0.528	0.337	0.794

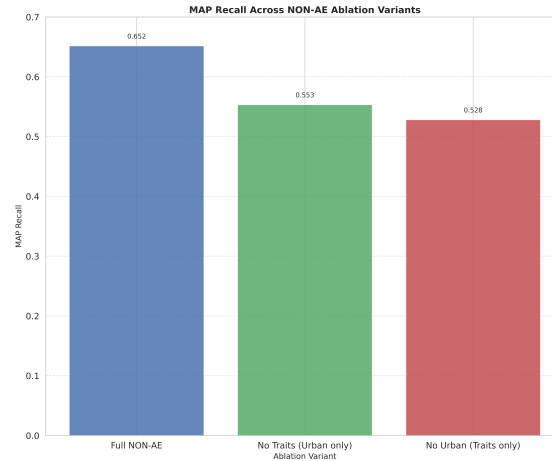
Note. Results are based on ablation experiments with the full 8D feature set (no autoencoder compression). Values are rounded to three decimals.

Table A.16.: *Per-class model performance in ablation analyses (8D, Without-AE)*

Model	Class	Accuracy	Precision	Recall	F1
Full (8D)	MAP	0.788	0.255	0.652	0.366
	NON-MAP	0.788	0.957	0.802	0.872
Urban only	MAP	0.849	0.323	0.553	0.408
	NON-MAP	0.849	0.950	0.879	0.913
Traits only	MAP	0.804	0.247	0.528	0.337
	NON-MAP	0.804	0.944	0.833	0.885

Note. Metrics reported separately for MAP and NON-MAP classes. Values are rounded to three decimals.

Figure A.4.: MAP recall across the three 8D ablation variants on the development set: Full model, No Traits (Urban only), and No Urban (Traits only). Recall was highest for the Full model (0.65), followed by the Urban-only model (0.55), with the Traits-only model slightly lower (0.53). This pattern indicates that urban features contributed more to recall of map events than trait features.



Note. MAP = Maps & Navigation taps (positive class). Values above the bars correspond to recall.