Individual Differences in Human Navigation on the Information Network

Manran Zhu

Supervisor: János Kertész

A Dissertation Submitted in fulfillment of the requirements
for the Degree of Doctor of Philosophy in Network Science

Central European University
Vienna, Austria

2024

Manran Zhu: *Individual Differences in Human Navigation on the Information Network*

2024

Researcher declaration

I Manran Zhu certify that I am the author of the work Individual Differences of Human Navigation on the Information Network. I certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, the contributions of others. The thesis contains no materials accepted for any other degrees in any other institutions. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright.

Statement of inclusion of joint work

I confirm that Chapter 4 is based on a paper which was written in collaboration with Taha Yasseri and János Kertész [1]. Dr. Kertész and I conceived of the idea to conduct an experiment online to investigate human navigation behavior and their individual differences on the information network. I carried out the experiment, collected and analyzed the data. Dr. Yasseri conceived of the idea to study the interplay between success and uniqueness of the participants navigation trajectories. All authors contributed to the writing of the paper on which the chapter is based and gave final approval for publication. Dr. Kertész and

Dr. Yasseri endorses this statement with their signatures below.

I confirm that Chapter 5 is based on a working paper which was written in collaboration with János Kertész. Dr. Kertész and I conceived of the idea to study the categorization and individual preferences of the participants' navigation strategies. Dr. Kertész and I collaborated on developing and improving the methods used in the paper. I designed and conducted the online experiment, collected and analyzed the data. Dr. Kertész and I and both contributed to the writing of the paper on which the chapter is based and gave final approval for submission. Dr. Kertész endorses this statement with his signature below.

Signature of PhD Candidate:

March, 2024

Signature of Dr. János Kertész, endorsing statement of joint work:

March, 2024

Signature of Dr. Taha Yasseri, endorsing statement of joint work:

March, 2024

ABSTRACT

With the rapid accumulation of online information, efficient web navigation has grown vital yet challenging. To create an easily navigable cyberspace catering to diverse demographics, understanding how people navigate differently is paramount. Previous research has discovered multiple patterns in how individuals navigate in the geographic, social, and information spaces, yet individual differences in navigation performance and strategies in the knowledge space has remained largely unexplored. To bridge the gap, we conduct an online experiment where participants played a navigation game on Wikipedia and completed questionnaires about their personal information. Our analysis shows that age negatively affects knowledge space navigation performance, while multilingualism enhances it. Under time pressure, participants' performance improves across trials and males outperform females, an effect not observed in games without time pressure. In our experiment, successful route-finding is usually not related to abilities of innovative exploration of routes. Utilizing a graph embedding trained on the English Wikipedia, our study identifies distinctive strategies that participants adopt: when the target is a famous person, participants typically use the geographical and occupational information of the target to navigate, reminiscent of hub-driven and proximity-driven

approaches respectively. We discovers that many participants playing the same game exhibit a ″wisdom of the crowd″ effect: The set of strategies provide a good estimate for the information landscape around the target indicating that the individual differences complement each other.

ACKNOWLEDGEMENTS

I would like to acknowledge several individuals who have significantly influenced this work.

First, I must thank my advisor, János Kertész, for his endless support and patience throughout my PhD studies, without which this thesis would not have been possible. He has shown me, not in words, but in action, how to stay positive when stuck in difficulties, how to dive deep when floating is easy, and most importantly, where to look among the garbage and the flowers.

I am grateful to the members of the Department of Network and Data Science at CEU: our coordinator, Olga Peredi, the fellow Ph.D. students, postdocs, and visitors, who have provided a lively environment in Budapest for us to learn and grow. I'm particularly thankful for the help of Dr. Milán Janosov, Dr. Johannes Wachs, Dr. Srebrenka Letina, Dr. Orsolya Vásárhelyi, Dr. Dávid Deritei, Dr. Tamer Khraisha, Prof. Ancsa Hannák, Prof. Roberta Sinatra, Prof. Federico Battiston, Prof. Tiago Peixoto, Prof. Elisa Omodei, Prof. Márton Pósfai, and Prof. Balazs Vedres. Last but not least, the Ph.D. students of my cohort and my friends, Dr. Rebeka O. Szabó, Matteo Neri, Dr. Luis Guillermo Natera Orozco, Dr. Júlia Perczel, Felipe Vaca, Maria Gabriela Juncosa Calahorrano, Teodoro Criscione, and Hao Cui.

I am grateful to my colleagues at CSSH, RWTH Aachen, and Prof. Markus Strohmaier, who hosted me there. They have broadened my view of the field of computer science and directly inspired my PhD research. I am also deeply grateful to Prof. Taha Yasseri, who hosted me in the Geary Institute at UCD and supervised me generously for many hours, both when I was in Dublin and afterwards. I'm also grateful to my collaborators at the Joint Research Center, European Commission, Dr. Mafini Dosso, Dr. Simone Sasso, Dr. Angela Sarcina, and especially Dr. Fatime Barbara Hegyi, who started my journey as an external expert at the European Commission and provided me great support during the process. Lastly, I am thankful to Miklos Orban J.D., who has collaborated with me on multiple projects that are both interesting and rewarding.

I'm lucky to have the encouragement and support from my dear friends, Milán Janosov, Casey Tompkins, James Parsley, Kevin Caffrey, Qian Feng, Chuanqi Xiao, Noémi Éltető, Dongge Han, Xi Liu and many others. Their friendship has made the PhD journey much more enjoyable and meaningful.

Most importantly, I'd like to thank my father, Daichun Zhu, and mother, Cuiping Hao, who have always been there for me in good times and bad. My mother has shown me the warmest love I could imagine. I dedicate this thesis to her.

Contents

List of Tables

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The ability to move from place to place is essential for animals to find critical resources such as food, mate, and places to live [3]. This quest for resources isn't limited to just the tangible world; it extends into more conceptual areas, such as seeking out individuals who can provide help within our social networks [4], or looking for answers to our questions online in the knowledge space [5]. The rapid growth of information available online has led to an era of information overload, posing a significant challenge to effectively finding our way through the digital landscape [6]. Addressing this issue requires an in-depth understanding of how we navigate through the abundance of information.

An early exploration into the process of navigating the information landscape traces back to 1929 when the Hungarian writer Frigyes Karinthy wrote about an thought experiment to "...select any person from the 1.5 billion inhabitants of the Earth", to "...contact the selected individual using nothing except the network of personal acquaintances" [7]. Witnessing the then-recent advancements in communication technologies that seemingly shrank distances between people, Karinthy conjectured that fewer than five intermediaries would be needed to complete this task. Forty years later, Stanley Milgram implemented the thought experiment in what became known as the small world experiment [8]. In this study, participants in Omaha, Nebraska, and Wichita, Kansas, were asked to forward a letter to a designated recipient in Boston, Massachusetts, by passing it through friends in a relay. Despite the considerable geographical and social distances involved, a considerable number of letters reached their destination within an average of six steps. This experiment was replicated in 2003 with approximately 100,000 participants attempting to send an email to 18 individuals in 13 countries, and the new experiment largely reproduced the results of Milgram's experiment [9]. These outcomes suggest a remarkable human capacity for navigating the social space to reach specific targets. But what underlies this ability? What strategies do we em¬¬ploy to navigate our way?

Various studies have approached the issue from different angles. At the neuronal level, it has been discovered that navigation in conceptual space is closely linked to spatial navigation, a subject extensively explored by O'Keefe and Moser et al. in their Nobel Prize-winning works [10, 11, 12, 13, 14]. Spatial navigation studies reveal that we possess a map-like representation of the physical world, often referred to as the cognitive map, during navigation. Neurons in our brain hold a representation of location in physical space (place cells) and encode a variety of information such as the metrics of space (grid cells), the head direction of the individual (head direction cells), and so on, enabling one to plan a route towards a target [3, 15, 16]. Recent research has indicated that our brains also maintain a cognitive map for abstract concepts [4, 17, 18]. For example, when seeking assistance within social spaces, studies have shown that neurons in our hippocampus encode the affiliation and power of the individuals we interact with on a two dimensional map [19].

Going beyond the neuronal level, Watts et. al. [20] and Kleinberg [21] independently discovered that our ability to successfully navigate to a target is linked to the structure of the information space. Specifically, if the space exhibits a hierarchical structure, a simple greedy decentralized search strategy suffices for efficient navigation. This theory was later empirically confirmed by Adamic et al. [22] who demonstrated that given the email logs at HP Labs, a greedy decentralized search could effectively leverage the organizational hierarchy to find short paths to the target. In fact, different hierarchies can be utilized: in the context of social navigation, individuals typically rely on either geographical or occupational hierarchies to facilitate their navigation [9, 23].

To extend the study of social navigation to the broader context of information spaces, researchers have focused on Wikipedia [24], given its extensive range of topics and high user engagement. This interest led to the development of online navigation games on Wikipedia, such as Wikispeedia [25] and the Wiki Game [26]. In these games, players aim to navigate from one Wikipedia article (the source) to another (the target) by clicking through hyperlinks of other Wikipedia articles. Studies of navigation on the Wikipedia network have revealed interesting behavioral patterns of the players: they typically start by moving towards broader and more well-known articles before honing in on content that is more closely related to the target in terms of meaning [27]; the navigation behavior of players is non-Markovian, indicating that the choice made in the current step of navigation is influenced by previous decisions [28]; players do not adhere strictly to a greedy strategy, they sometimes choose the next page to click on at random, particularly in the early stages of navigation [29].

While previous research has shed light on various aspects of knowledge navigation, a comprehensive understanding of how individual differences affect this process remains elusive. Studies have shown that factors such as age, gender, and place of origin can significantly influence navigational performance in physical spaces [30, 31, 32, 33, 34]. This raises the question: do similar disparities exist in navigating knowledge spaces? Milgram's experiment highlighted that people rely on either geographical or occupational information for navigation, but what are the reasons for these preferences and their implications? How does the interplay between the structure of knowledge networks and individual strategies shape the navigation patterns across different demographic groups? Addressing these questions is crucial in today's online environment, which is theoretically designed to be accessible and egalitarian. However, in practice, access to information is not uniformly available to all. Individual factors like technology familiarity, digital literacy [35], education level [36], and even personality traits [37], all impacted one's ability to utilize the vast information resources available online [38]. Understanding the diverse ways individuals navigate knowledge spaces is key to addressing informational inequality. By gaining insight into these variations, we can develop personalized support mechanisms tailored to meet each user's unique needs.

This thesis investigates individual differences in navigation performance and strategies within the information space using a game setting. Games have been widely utilized to study our economic [39], social [40] and mobility [34] behavior patterns, owing to their intuitive, fun, and scalable nature [41]. In this thesis, we conducted an online experiment where we hired 802 participants from the US who played nine rounds of Wikipedia navigation games and subsequently completed a survey to provide their demographic information. Our experiment overcomes the common issue of small sample sizes typically associated with laboratory experiments and captures detailed demographic data of participants, which is often challenging to gather in large-scale online studies. Thus, this research yields a unique dataset for examining individual navigation behavior patterns.

This thesis examines the impact of individual characteristics on participants' performance and strategies in navigating through data, drawing on comprehensive data from our experiment. To analyze the semantic connections between Wikipedia articles, we trained a 64-dimensional embedding for each article in the English Wikipedia. Our findings indicate that older age groups tend to perform worse in navigating the knowledge space, whereas multilingual individuals exhibit improved performance. Additionally, under time constraints, participants generally enhance their performance over successive trials, with males surpassing females —- a discrepancy not observed in non-timed tasks. Moreover, successful navigation does not necessarily correlate with innovative route exploration. Through clustering analysis of the navigation paths, we identified three predominant strategies employed by participants: reliance on geographical cues, occupational details of the target, or a combination of both. These strategies reflect a broader navigation approach, characterized by either a hub-driven or proximity-driven methodology. Interestingly, the choice of strategy appears to be influenced more significantly by the surrounding knowledge landscape than by personal attributes like age or gender, displaying a 'wisdom of the crowd' effect.

The thesis is structured as follows:

- Chapter 2: Reviews the previous research on human navigation pertinent to this thesis.
- Chapter 3: Describes the online experiment we conducted, forming the foundation of this thesis.
- Chapter 4: Analyzes how individual characteristics affect participants' navigation performance.
- Chapter 5: Examines the navigation strategies employed by participants, along with their personal preferences.
- Chapter 6: Concludes the thesis, discussing the significance of our findings and suggesting avenues for future research.

CHAPTER 2

Related Work

In this chapter, I review prior research on human navigation behavior across physical, social, and informational spaces. These studies have examined human navigation behavior from diverse perspectives: some focus on how places or concepts are represented in our brain at the neuronal level, others investigate navigation tasks within the social networks at the global level; some research delves into navigation algorithms derived from synthetic networks, while others assess navigation behavior on empirical networks. This body of work has significantly inspired my research, offering a rich view of how humans navigate the world.

2.1 Cognitive maps and cognitive graphs

The idea of a cognitive map dates back to Tolman [42] in the 1940s. In an experiment studying how rats navigate the maze to find food, Tolman discovered that unlike the behaviorist perspective prevalent at the time, which emphasized stimulus-response relationships, rats show a more flexible navigation behavior: when the usual paths to food learned through stimulus-response are blocked, the rats could find new paths to the food that it had never traveled before, as though they have a map-like internal representation of the maze's layout. This idea, referred to as the cognitive map, was later supported by the Nobel Prize winning discovery of the place cells and grid cells in the animal brain [10, 11, 12, 13, 14]. in the early 1970s John O'Keefe and Jonathan Dostrovsky discovered that certain neurons in the rat hippocampus were activated only when the animal was in a specific location in its environment [10, 11]. These neurons, dubbed "place cells," fired in different locations, effectively creating a map of the environment within the brain. Several decades later, in 2005, May-Britt Moser and Edvard I. Moser, along with their colleague Torkel Hafting, discovered grid cells in the entorhinal cortex, a region of the brain closely linked to the hippocampus [12, 13, 14]. Grid cells fire in a unique hexagonal pattern, representing multiple locations that form a grid-like structure in the environment. This discovery showed the brain is able to calculate location and distance, providing a coordinate system that complements the place-specific information provided by place cells.

Alongside the concept of a Euclidean-style cognitive map, researchers have observed that a map-like representation is not enough to capture all the information encoded in our brain: we also maintain representations of the physical environment's topological structure, including the routes connecting locations and the hierarchical organization of locations which forms regions [43, 44]. Later research developed multiple models of graph-like representation of the physical space: Kuipers et. al. [45] proposed that the topological map we have of the physical world could be constructed from a skeleton of major routes, represented as a bipartite graph of places and paths, where a path is a one-dimensional ordered set of places; Warren [46] observed that truly novel shortcuts in animals, and human performance is highly unreliable and biased by environmental features, and that our spatial knowledge can be better characterized as a labeled graph consisting of a network of paths between places augmented with local metric information. Additionally, the hippocampus was later shown to encode relational position of locations (e.g., A is east of B) [47], segment space and time [48, 49], and even hierarchical structures of experiences [50]. These all point to a cognitive graph type of representation of the physical space.

Recently it was shown that the hippocampus and entorhinal cortex, which contain cells that encode spatial information and enable spatial navigation, also play essential roles in other neural processes such as social cognition [19], inference [51], imagination [52] and memory [53] in the broader knowledge domain [3, 17, 18, 54, 55]. In an experiment where participants were asked to play a role-playing game in the fMRI scanner in which they were moving to a new town, and that their goal was to find a job and a place to live by getting to know the town's people, Tavares et. al [19] found that a two dimensional geometric model of social relationships, a "social space" framed by power and affiliation, predicted hippocampal activity. In another study, where the participants were shown demorphisms of a bird with varying neck and leg length, researchers found that humans navigating conceptual two-dimensional knowledge showed the same hexagonal signal in a set of brain regions markedly similar to those activated during spatial navigation.

2.2 Navigation in the social space

The concept of navigating the social space can be traced to a 1929 thought experiment by the Hungarian author Frigyes Karinthy. Witnessing the burgeoning advancements in communication technologies that seemed to shrink the distances between people, Karinthy suggested a thought experiment in which one could choose any one of the Earth's 1.5 billion people (at the time) and attempt to connect with them solely through a chain of personal acquaintances. He speculated that it would take no more than five intermediaries to make the connection. Four decades later Stanley Milgram implemented the thought experiment, where he asked individuals in Omaha, Nebraska, and Wichita, Kansas, to try and send a letter to a person in Boston, Massachusetts, by handing it off through a chain of friends. Despite significant geographical and social separation, many of the letters successfully reached their destination within about six steps on average. This experiment was repeated in 2003 through an email-based version involving roughly 100,000 participants and 18 targets across 13 countries, which essentially replicated Milgram's findings. These experiments collectively underscore the profound ability of humans to navigate through social networks to connect with the target person.

Previous research have found that our efficient navigation ability is linked to the structure of the social network. Watts et al. [20] observed that the way we are connected within the social network is highly structured: we all possess different identities and belong to groups characterized by specific social attributes. These group structures naturally form hierarchies, akin to the departmental organization in universities or companies, where individuals belong to groups, which in turn belong to larger groups. They demonstrated that social networks created from this hierarchical structure through simple linking rules are navigable: utilizing a greedy search algorithm with only local knowledge of the social network, one can navigate to any target person in a few steps. Independently, Kleinberg[21] proved that for a network formed from a tree graph under certain linking conditions, a greedy decentralized search algorithm could reach any target in polylogarithmic time. This theory was later empirically confirmed by Adamic et al. [22] who demonstrated that given the email logs at HP Labs, a greedy decentralized search could effectively leverage the organizational hierarchy to find short paths to the target. In fact, different hierarchies can be utilized: in the context of social navigation, individuals typically rely on either geographical or occupational hierarchies to facilitate their navigation [9, 23].

2.3 Navigation in the knowledge space

Extending social navigation to the realm of knowledge spaces, attention has been directed towards online information-seeking behaviors, especially on the Wikipedia [24] platform, noted for its wide topic range and significant user interaction. Recently a type of Wikipedia navigation games, such as the Wikispeedia [25] and the Wiki Game [26] gained popularity online, with over 500,000 users and 1.5 million click trails. In each game, participants are challenged to move from one Wikipedia entry (the starting point) to another (the destination) by clicking on the hyperlinks within the Wikipedia articles. Research into player behavior within these Wikipedia-based games has uncovered intriguing patterns: players often initially steer towards more general and familiar articles before narrowing their focus to content more directly related to their goal [27]; their navigation decisions are non-Markovian, meaning that choices at any given step are influenced by past actions [28]; and players' strategies are not purely greedy, as they sometimes randomly select their next move, particularly in the early stages of the game [29].

West et. al. [27] observed that players adopt a certain strategies during navigation: they often first navigate to pages well-connected to others in the early phase, and then prioritize pages closer to the target in the homing-in phase. Comparing human navigation paths to those generated by automatic navigation agents equipped with simple numerical features, researchers found that navigating the complex Wikipedia network efficiently does not require sophisticated background knowledge or high-level reasoning [56]. Furthermore, our navigation patterns are not only unsophisticated but also predictable in their failures: trained on a large number of players' navigation paths, a statistical model can identify patterns of failure and predict whether a user will complete or abandon a navigation task [57]. Additionally, these navigation paths can aid in developing methods to automatically identify and add missing hyperlinks to Wikipedia, thereby improving the navigability of the Wikipedia network [58].

Applying statistical inference methods to the extensive set of navigation trails left by users in the Wikipedia navigation game, researchers identified various preferences and biases in our navigation patterns [28, 59, 29, 60, 61]. For instance, by categorizing Wikipedia articles into topics, it was found that our current navigation decisions are influenced by past navigation choices [28]. Compared to a greedy navigation algorithm, which selects the closest page to the target at each step, humans do not always make the optimal choice: sometimes we make random moves, especially in the early stages of navigation [29]. Our navigation is not only random at times but also biased. Studies have shown that the structure of Wikipedia pages influences our navigation decisions, with players more frequently clicking on links located near the top of an article [59]. These insights into our navigation biases highlight the challenges and promising directions for designing information environments that are easier to navigate for everyone.

2.4 Individual differences in navigation

Several individual differences have been noted in spatial navigation [30, 31]. Age, linked with declining cognitive abilities such as fluid intelligence, perceptual speed, memory, and vocabulary [62], has been found to negatively affect navigation performance in physical spaces [63, 64, 34]. Studies suggest that men often use a Euclidean strategy involving cardinal directions and distances, whereas women tend to prefer a landmark strategy, relying on a sequence of turns and proximal cues [65, 66, 67]. A meta-analysis [32] indicates that, on average, men outperform women in navigation tasks with a small to medium effect size, influenced significantly by the task type, dependent variable, and testing conditions, such as timing. Recent research through the video game Sea Hero Quest shows a near-linear decline in navigation ability from the early 20s, with a consistent male advantage across countries, although this varied and was partially predicted by gender inequality [33]. Additionally, individuals reporting urban upbringing were found to navigate worse than those from non-urban environments.

In the realm of online information-seeking, distinct cognitive patterns and abilities have been identified, with significant factors influencing performance. Research indicates that effective information-seeking relies not just on internet knowledge but also on critical cognitive abilities, disadvantaging older adults [68, 69]. Gender also influences information-seeking behaviors, with males typically showing higher confidence in web navigation than females [70]. Ethnic and cultural backgrounds affect online information-seeking approaches, with a survey revealing that international students prefer starting searches on the internet over university electronic resources, unlike American graduate students [71]. Other factors, such as ideology [72] and personality [37], shape cognitive processes and, consequently, information-seeking patterns. While not directly linked to information-seeking, bi/multilingualism is associated with cognitive advantages [73], hinting at possible impacts on information-seeking efficacy.

CHAPTER 3

THE EXPERIMENT

3.1 Introduction

Wikipedia [24], the world's most visited online encyclopedia, offers vast knowledge across a wide spectrum of topics and languages. Unlike traditional encyclopedias, which are crafted by expert groups, Wikipedia serves as a collaborative platform allowing anyone to edit articles within specific guidelines [74]. In recent years, the Wikipedia navigation game, known as Wikispeedia [25] or the Wiki Game [26], has become increasingly popular online. Players start with two Wikipedia articles, designated as the source and target, and must navigate to the target page from the source using hyperlinks within Wikipedia articles. Several studies have analyzed players' navigation paths, uncovering various insights into navigation patterns [27, 28, 29]. However, there remains a lack of understanding, however, as the navigation patterns identified have not considered personal information such as age and gender, thus failing to highlight the behaviors and preferences of different demographic groups. Consequently, further research is necessary to gain a deeper comprehension of how these factors might affect navigation behaviors. Additionally, while previous studies have used different information-seeking tasks, there has yet to be a comprehensive analysis that integrates these elements within a singular information-seeking framework.

To bridge this gap, we carried out an online experiment where we hired 802 participants from the crowd-sourcing platform Prolific [75]. These participants engaged in nine rounds of the Wikipedia navigation game and subsequently completed a survey providing their demographic details. In this chapter, I will detail the experiment's setup, as well as the data collection and cleaning processes.

3.2 Experiment setup

Our longitudinal study comprises two rounds of online experiments, the first conducted in January 2020 and the second in October 2023. Participants were sourced from Prolific [75], a well-regarded crowdsourcing platform for behavioral studies [76]. The experiments were conducted on the Qualtrics [77] platform, where we embedded Wikipedia navigation games into the Qualtrics survey using custom JavaScript, followed by a survey. We utilized the 20190820 English Wiki Dump [78] for the navigation games in both experiment rounds. This Wikipedia snapshot includes 5.9 million nodes and 133.6 million edges.

In our experiment, each participant engages in nine rounds of the Wikipedia navigation game and completes a survey afterward. To test the effect of different constraints on the game, participants can choose between a Speed-race game or a Least-clicks game challenge. To win, they must navigate to the target page within 150 seconds for Speed-race games or in 7 steps for Least-clicks games. During each game, the interface displays pages visited earlier in the current game on the left margin, allowing players to backtrack to any of those pages (screenshots of the game environment in Figure 3.1). Following the game session, participants fill out a survey regarding their demographics, take a Big Five personality test, and answer questions about their experience with Wikipedia, computer games, and their self-assessed spatial navigation skills. Upon completing the experiment, each participant receives a base payment of 5 pounds, plus a bonus of 0.5 pounds for each game won. The entire survey takes approximately 60 minutes and needs to be completed in 90 minutes. After 90 minutes the survey will be inactive and you will not be able to submit your answer and get paid. Participants in our experiment won on average 4 games, adding up to an average payment of 7 pounds per participant, which is beyond the minimum hourly rate on Prolific (6 pounds per hour).

3.2.1 The participants

Our participants, recruited from the US, were selected based on the following prescreening criteria. For the first round of the experiment, the criteria were: i) participants are US citizens, ii) an equal number of female and male participants, iii) the sample composition by ethnicity to be approximately $\sim$50% White, $\sim$17% Asian, $\sim$17% Hispanic, and $\sim$17% African American. In the second round of the experiment, the prescreening criteria were: i) participants are

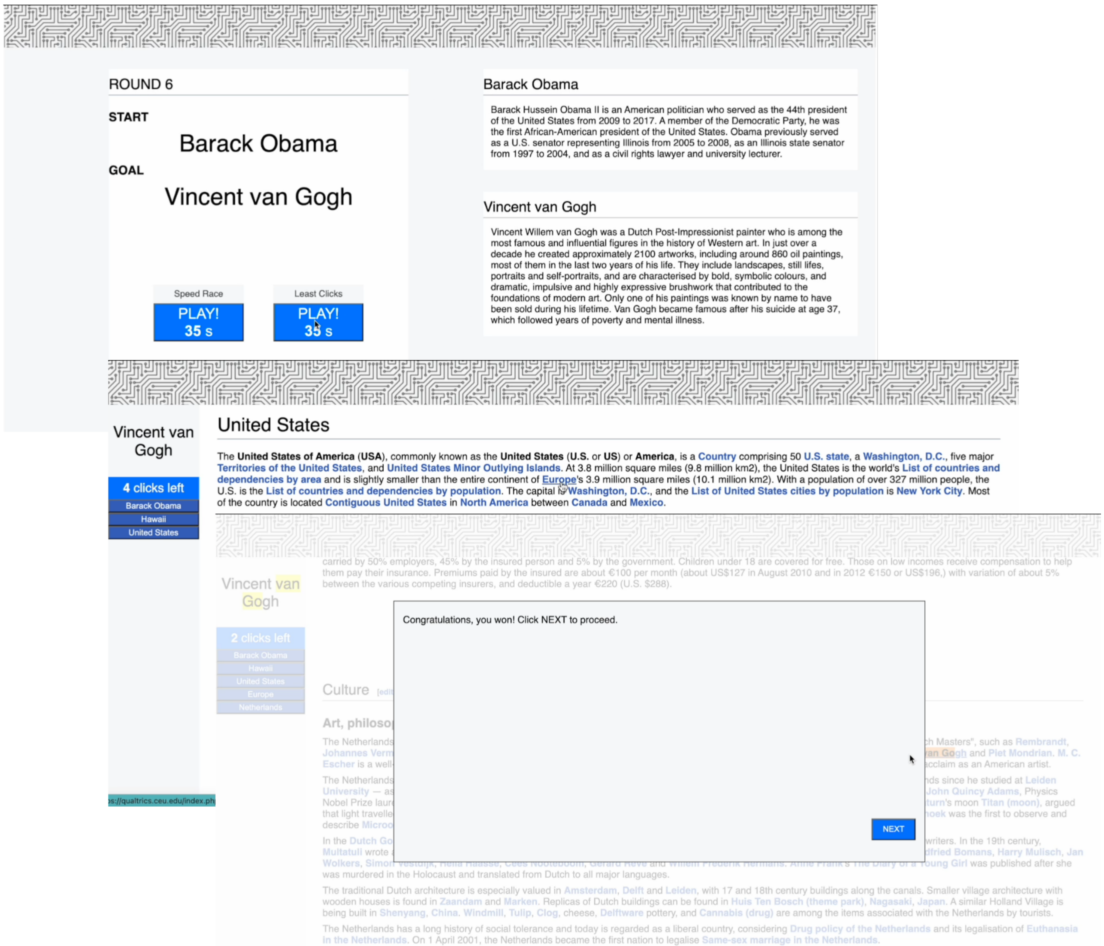

Figure 3.1. *The figure shows the screenshots of the game environment for the Least-clicks game with the source page "Barack Obama" and target page "Vincent van Gogh". In the beginning of the game, participants are shown the source and target page of the game and a short introduction of the pages. Participants have 60 seconds to choose to play either the Speed-race or Least-clicks game by clicking the respective "play" button under the source and target game.*

US citizens, ii) an equal number of female and male participants, iii) half of the participants having English as their sole native language and the other half with a non-English native language aside from English, and iv) the ethnic composition of the sample to be roughly ∼33% White, ∼33% Asian, and ∼33% African American.

3.2.2 The game sessions

The experiment begins with an introduction explaining the navigation game, including the two types of games participants can choose from and the potential bonus they can earn. Participants are then given two trial games: a Speed-race game starting from the "William Shakespeare" page to the "Oprah Winfrey" page, and a Least-clicks game from "Bob Dylan" to "Abraham Lincoln". Following the trial games, there are three game sessions, A, B, and O, separated by one-minute breaks. Each session involves playing three rounds of the navigation game. The specific games included in each session are detailed in Figure 3.2. To reduce the disparities in prior knowledge among the participants, the source and target pages are chosen to be similarly distanced (2 or 3 steps away on the Wikipedia network) pages about renowned individuals from various domains such as artists, directors, scientists, and politicians, spanning different historical periods and encompassing both genders. To minimize the influence of play order, we randomized both the order of games within each session and the sequence of the three sessions. After each game, participants answer several questions about the game they just played, such as such as how much did you know about the source/target page before the game? Did you find this round of the game difficult? To go from the source page to the target page, at least how many steps are needed (your estimation)? As well as several other questions regarding the steps participants took in the game. Once all three game sessions are finished, the survey session starts.

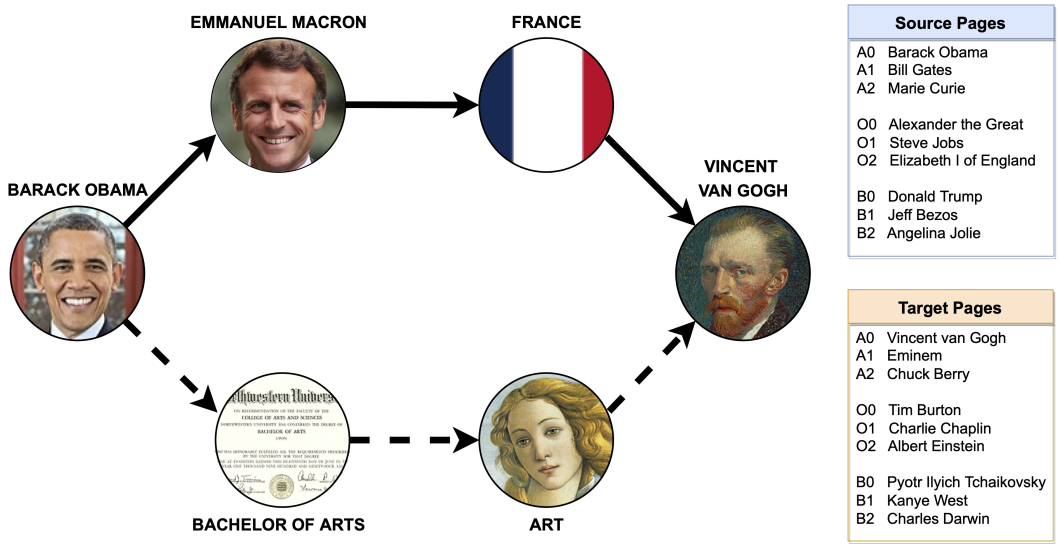

Figure 3.2. *In the Wikipedia navigation game, players need to go from one Wikipedia article (source page) to another (target page) through the links of other Wikipedia articles on the current page in 7 steps (Least-clicks game) or 150 seconds (Speed-race game). The figure shows two possible navigation paths from the source page* BARACK OBAMA *to the target page* VINCENT VAN GOGH*: 1)* BARACK OBAMA *to* EMMANUEL MACRON *to* FRANCE *to* VINCENT VAN GOGH *(solid arrows); and 2)* BARACK OBAMA *to* BACHELOR OF ARTS *to* ART *to* VINCENT VAN GOGH *(dotted arrows). Participants each played nine rounds of games whose source page and target page are shown in the figure. The games are divided into three sessions A, O, and B, with three games in each session. The order of the games is randomized in each game session, and the order of sessions A and B are randomized to reduce the effect of the games' order on performance. Attributions to the images used in the figure are included in the references [79, 80, 81, 82, 83, 84].*

3.2.3 The survey sessions

The survey sessions commence with a Big Five personality test [85], assessing participants' five personality traits: openness to experience, conscientiousness, extroversion, agreeableness, and neuroticism. Following this, we pose six categories of questions to gather information about participants' i) employment status, ii) educational background, iii) spatial navigation habits, and their previous experience with iv) the Wikipedia navigation game, v) the Wikipedia website, and vi) computer games. We also inquire about demographic details, including age, gender, ethnicity, political affiliation, and language skills. An attention check question is included at the survey's end, requiring participants to slide a bar to the left. In the second round of the experiment, participants complete a digit span memory test [86] at the end of the survey, earning an additional 0.25 pounds for each extra digit remembered, starting from a base of three digits.

3.3 Data

The experiment provided us with rich information about the participants. Beyond the answers to the survey questions mentioned earlier, we also recorded the navigation paths of each participant in each game, detailing the pages visited, the timestamp of each visit, whether the step was a backclick, and the type of game (Speed-race or Least-clicks) chosen by the participants. The complete list of survey questions is summarized in the Appendix.

3.3.1 Cleaning

Inevitably, our dataset includes some noise. With a large number of participants taking part in the experiment simultaneously, there were errors in data recording, such as some game types being logged multiple times. Moreover, as common in many online and offline studies, the presence of participants with low engagement resulted in biased samples. To mitigate these issues, we implemented the following data cleaning steps:

- Removed navigation paths where the recorded number of game types did not align with the total number of games played.
- Excluded data from participants who failed the attention check.
- Excluded data from participants who took part in the experiment more than once.
- Removed data from participants with missing values in their survey responses.

Following these cleaning procedures, we were left with 6687 navigation paths from 743 participants over two rounds of the experiment.

CHAPTER 4

Navigation Performance

4.1 Introduction

Digital advancements have transformed the way society accesses and disseminates information. Although the internet is perceived as a democratic space for information dissemination, allowing easy access to everyone, the reality proves more complex. The complexity of the web's networks complicates the search for information [38], and not all individuals have the same level of access to digital resources. Disparities in access to high-speed internet, smart devices, and differences in individual characteristics like familiarity with technology, digital literacy [35], education [36], and personality traits [37] impact one's ability to utilize online information effectively [38]. Moreover, broader issues such as information censorship by governments further obstruct the free flow of information online [87]. The process of seeking information online, which includes both searching and navigating, involves skills that are interrelated yet distinct [88]. Our strategy aims to address these disparities by acknowledging the varied ways individuals access knowledge, tailoring support to meet each user's unique needs.

Research has shown that individual cognitive styles and skills significantly influence online information search effectiveness, highlighting the importance of not only internet knowledge but also specific cognitive abilities, thereby placing older adults at a disadvantage [68, 69]. Gender differences have been observed, with men typically more confident in web navigation and women more likely to use landmarks for guidance [70]. Cultural and ethnic backgrounds also affect online search strategies, with differences noted between international and American students in their preferences for starting searches on the internet or using academic resources [71]. Other factors, including personal beliefs [72] and personality traits [37], influence information search behavior by shaping cognitive processes. Additionally, being bilingual or multilingual may offer cognitive benefits that enhance information-seeking skills [73]. Despite these findings, earlier research has not consistently used the same tasks to study information seeking, and a detailed analysis that includes all these factors in a single study setup is missing. As a result, how these factors rank in importance in affecting information-seeking success is still to be fully understood.

Our investigation is also inspired by the relationship between navigation in the physical spaces and knowledge domains. Prior studies highlight that certain brain regions, notably the hippocampus and entorhinal cortex, essential for spatial processing and navigation, play a significant role in cognitive abilities such as social cognition and memory [19, 53]. There are observed disparities in individuals' spatial navigation capabilities: spatial skills progressively decline with age [63, 30]; in general, men surpass women in tasks requiring spatial navigation [32, 30]; and people from rural environments exhibit superior spatial navigation abilities compared to their urban counterparts [33]. Given these disparities and the connection between physical and knowledge-based navigation, it is vital to explore if the differences in navigating the realm of knowledge reflect those in physical space navigation.

In this chapter[1], I investigate the interplay between individual characteristics and navigation performance in the information space using the data obtained from the first round of experiment discussed in Chapter 3. In particular, we looked at two aspects of the participants' performance: if they have won the game or not (success) and how unique their navigation paths are.

[1]This chapter is based on the article "Individual differences in knowledge network navigation" [1].

4.2 Methods

4.2.1 Individual characteristics

After processing the responses to the survey questions outlined in the Appendix, we identified 18 control variables reflecting the participants' characteristics as described by the six question groups previously mentioned, plus an additional 5 control variables related to game specifics, such as game type (Speed-race or Least-clicks), round number, and participants' familiarity with the Wikipedia articles designated as source and target in the games. Furthermore, we incorporated 11 independent variables to represent participants' Big Five personality traits, age, gender, ethnic background, political orientation, and language skills, be it foreign or native. To address the issue of strong correlations and anti-correlations among these control variables, we employed principal components analysis (PCA) [89] for each question group, capturing 80% of the variance through a reduced set of variables (principal components). This approach led to a streamlined collection of 13 control variables, with their loadings presented in Table 4.1. Descriptive statistics detailing participant demographics are included in Table 4.2. The data shows that male participants tend to be younger and less politically liberal than female participants and show greater ethnic diversity. Additionally, they are more likely to speak a foreign language and have previous experience with the Wikipedia navigation game. In contrast, female participants showed a preference for the Least-clicks game, which is not time-bound, as opposed to males who preferred the Speed-race game. In terms of Big Five personality traits, the study observed minimal gender differences

(Maximum t-value = 1.75).

Table 4.1. *The table displays encoded variables (first column) and their corresponding loadings on the primary principal components in each question category, retaining at least 80% of the variance within each category. Loadings quantify the extent to which original variables contribute to specific principal components. Larger values, regardless of sign, indicate a stronger association between the original variable and the principal component. The sign of the loading indicates whether the correlation between the variable and component is positive or negative.*

Variables	Principle Components and Factor Loadings			
	Wikipedia$_1$	**Wikipedia$_2$**		
$W_{purpose}$	0.71	-0.71		
$W_{frequency}$	0.71	0.71		
	Computer$_1$	**Computer$_2$**		
$C_{frequency}$	0.58	-0.53		
C_{good}	0.54	0.82		
C_{like}	0.61	-0.21		
	Spatial$_1$	**Spatial$_2$**	**Spatial$_3$**	**Spatial$_4$**
S_{good}	0.58	-0.13	-0.22	-0.44
S_{learn}	0.44	-0.48	-0.30	0.69
$S_{unknown}$	0.59	0.19	0.10	-0.33
S_{known}	0.33	0.61	0.46	0.47
S_{left}	0.06	-0.59	0.80	-0.08
	Education$_1$	**Education$_2$**		
ED_{years}	0.71	-0.71		
$ED_{highest}$	0.71	0.71		
	Employment$_1$	**Employment$_2$**	**Employment$_3$**	
EM_{status}	0.30	-0.55	0.77	
EM_{mental}	0.60	-0.06	-0.18	
$EM_{physical}$	0.06	0.74	0.56	
$EM_{intensive}$	0.54	0.37	-0.08	
$EM_{creative}$	0.51	-0.09	-0.22	

Table 4.2. *Description characteristics of the study group.*

Characteristic	Female (N = 203)		Male (N = 192)		Total (N = 397)	
Age						
Mean ± S.D.	34.6 ± 11.1		32.0 ± 9.6		33.2 ± 10.5	
Minimum	19		19		19	
Maximum	64		77		77	
Political Position (N, %)						
Liberal	125	(61.6)	109	(56.8)	235	(59.2)
Moderate	41	(20.2)	50	(26.0)	91	(22.9)
Conservative	25	(12.3)	21	(10.9)	47	(11.8)
Other	12	(5.9)	12	(6.2)	24	(6.0)
Ethnicity Background (N, %)						
White	103	(50.7)	81	(42.2)	185	(46.6)
African	36	(17.7)	38	(19.8)	74	(18.6)
Asian	34	(16.7)	37	(19.3)	71	(17.9)
Hispanic	27	(13.3)	35	(18.2)	63	(15.9)
Other	3	(1.5)	1	(0.5)	4	(1.0)
Language (Speaks a foreign language) (N, %)						
No	98	(48.3)	103	(53.6)	201	(50.6)
Yes	105	(51.7)	89	(46.4)	196	(49.4)
Language (Has a foreign native language) (N, %)						
No	164	(80.8)	161	(83.9)	327	(82.4)
Yes	39	(19.2)	31	(16.1)	70	(17.6)
Big Five Personality Traits (Mean ± S.D.)						
Agreeableness	20.1 ± 4.6		20.7 ± 4.1		20.3 ± 4.4	
Conscientiousness	20.0 ± 4.5		20.1 ± 4.4		20.0 ± 4.5	
Extraversion	20.0 ± 4.6		20.0 ± 4.4		20.0 ± 4.5	
Neuroticism	19.3 ± 4.6		19.3 ± 4.6		19.4 ± 4.6	
Openness to Experience	21.0 ± 4.5		20.3 ± 4.1		20.6 ± 4.3	
Prior Experience with the Wikigame (N, %)						
I have never heard of the game and never played it before	157	(77.3)	119	(62.0)	278	(70.0)
I have heard of the game but never played it before	35	(17.2)	51	(26.6)	86	(21.7)
I have played the game (or similar game) several times before	10	(4.9)	15	(7.8)	25	(6.3)
I have played the game (or similar game) many times before	1	(0.5)	4	(2.1)	5	(1.3)
Other	0	(0)	3	(1.6)	3	(0.8)
Number of Games Won						
Mean ± S.D.	4.0 ± 2.6		4.4 ± 2.6		4.2 ± 2.6	
Number of Games Played in Each Type (Mean ± S.D.)						
Speed-race (game with time constraint)	3.8 ± 3.3		4.7 ± 3.3		4.3 ± 3.3	
Least-clicks (game with distance constraint)	5.2 ± 3.3		4.3 ± 3.3		4.7 ± 3.3	

4.2.2 Navigation paths

A participant's navigation path is defined as the series of Wikipedia articles, or Wikipages, that the participant selects during a game. By modeling the hyperlink structure of the English Wikipedia as a directed graph $G = (V, E)$, where $V = \{a_k\}$ represents all the Wikipages a_k and $E = \{H_{kl}\}$ signifies all the hyperlinks H_{kl} connecting a_k to a_l, we can describe the navigation path with N steps for the nth participant in the ith game g_i as a sequence $P_n^i = (a_k)_{k=0}^N$ within the graph G. Here, i ranges from 1 to 8, and n from 1 to 397. If the game g_i's source and target Wikipages are A_s^i and A_t^i respectively, then the navigation path $P_n^i = (a_k)_{k=0}^N$ for the nth participant in game i is deemed successful if it initiates at the source and concludes at the target, meaning $a_0 = A_s^i$ and $a_N = A_t^i$. Conversely, it's considered unsuccessful if $a_0 = A_s^i$ but $a_N \neq A_t^i$. Across all games and participants, the success of the nth participant in the ith game is evaluated by a binary variable s_n^i, assigned 1 if the path P_n^i is successful, and 0 otherwise.

4.2.3 Quantifying the uniqueness of the navigation paths

To analyze the variation in navigation paths, we first developed a 64-dimensional vector representation for each Wikipedia page a_i within the English Wikipedia network G through the DeepWalk [90] algorithm. Graph embedding, a method that maps each graph node to a numerical vector in a multidimensional space, places similar nodes closer together, facilitating the measurement of node dissimilarity as the vector distance. In our case, each Wikipedia page a_i is assigned a 64-dimensional vector $\vec{v}_i$, enabling us to define a semantic distance between any two Wikipedia pages as:

$$d(a_i, a_j) = 1 - \frac{\vec{v}_i \cdot \vec{v}_j}{\|\vec{v}_i\| \|\vec{v}_j\|}, \tag{4.1}$$

where $d(a_i, a_j)$ represents the cosine distance between the embeddings $\vec{v}_i$ and $\vec{v}_j$ of Wikipedia pages a_i and a_j. To validate our embedding's effectiveness, we assessed it using the WikipediaSimilarity 353 Test [91], an adaptation of the WordSimilarity 353 Test [92] designed to evaluate semantic similarities between words. Our embedding achieved a Spearman rank correlation of 0.667 with this test, indicating a performance on par with the latest measures of semantic relatedness for Wikipedia articles [93].

Considering two navigation paths $P^i_m = (a_k)^M_{k=0}$ and $P^i_n = (a_l)^N_{l=0}$ from the mth and nth participants in game g_i, we measure the distance between these paths using the Hausdorff distance [94, 95] across the sets of Wikipedia pages:

$$D_H(P^i_m, P^i_n) = \max\{ \sup_{a_k \in P^i_m} d(a_k, P^i_n), \sup_{a_l \in P^i_n} d(P^i_m, a_l)\} \tag{4.2}$$

with $d(a_k, P^i_n) = \inf_{a_l \in P^i_n} d(a_k, a_l)$ estimating the shortest distance from the Wikipedia page a_k to any page a_l within path P^i_n. Utilizing these distances, we determined the uniqueness of a successful navigation path P^i_n for the nth participant in game g_i as the average distance to all other successful paths in the same game:

$$u_n^i = \frac{1}{K_i - 1} \sum_{s_m^i = 1, m \neq n} D_H(P_n^i, P_m^i) \tag{4.3}$$

$$\tilde{u}_n^i = \frac{u_n^i - \mu}{\sigma} \tag{4.4}$$

Here, K_i counts the successful paths in game i. By normalizing the uniqueness of paths using the average score μ and standard deviation σ from all successful paths in the game, we acquire the standardized uniqueness $\tilde{u}_n^i$ for the nth participant's path in game i. The uniqueness scores for each successful path are depicted in Fig. 4.1.

4.2.4 Regression models

To explore the effects of individual traits on navigation success and creativity, we utilized four regression models. We carried out logistic regression analyses for navigation success across games with either time or distance constraints, using the binary outcome s_n^i, indicating whether the navigation by the nth participant in the ith game was successful or not, as the dependent variable. For evaluating creativity in navigation, linear regression was applied for each type of game, with the dependent variable being the standardized uniqueness score. Ethnicity was coded into two binary variables for identifying Asian/African American backgrounds, and political orientation was coded as a binary variable for a liberal viewpoint, as these factors (Asian/African American background and liberal orientation) were found to be significant ($p < 0.01$) predictors of navigation success, whereas other ethnic and political categories were not. Control

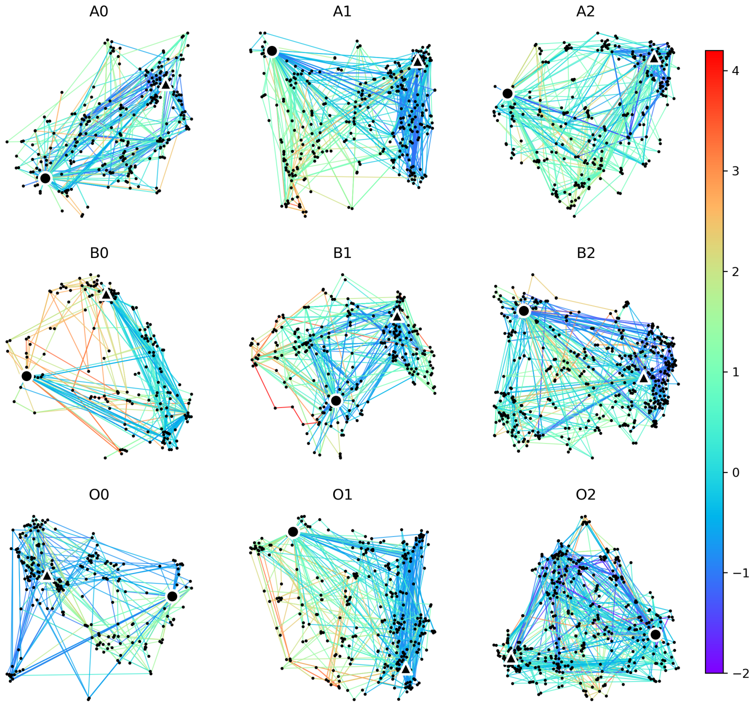

Figure 4.1. *The figure displays uniqueness scores for successful navigation paths across nine distinct games. The black circle and triangle denote the source and target Wikipedia pages of the respective game, and black dots represent visited Wikipedia articles, and their positions reflect their new two-dimensional coordinates derived from reducing the original 64-dimensional embeddings using the TSNE technique [96]. Lines indicate successful navigation paths within the games, with line color corresponding to the uniqueness scores of these paths.*

variables included a dummy variable for the game index to address varying game difficulties and a numeric variable for game order to control for potential shifts in attention during the study. The regression findings are summarized in Table 4.3 (dummy variables for game index are omitted for simplicity in presentation. For comprehensive regression results, refer to Table 4.4). Figure 4.2 shows the correlation among significant predictors ($p < 0.01$) of navigation success, and Table 4.5 lists the variance inflation factors (VIF) for all independent variables, indicating minimal collinearity concerns with the highest VIF at 2.38. Additional logistic regressions including interactions among variables were conducted to verify the persistence of main effects on navigation success, with details on interaction term selection and findings presented in Table 4.4). The analysis and VIF for the variables in these logistic regression models are detailed in Tables 4.4-4.5. Significant predictors of navigation success remained influential in both Speed-race and Least-clicks games after adding interaction terms, except for Wikipedia_1 in the Least-clicks games, which still showed significance at $p < 0.05$ with the interaction terms included. While factors such as employment status, proficiency in computer games, and interaction terms became significant, their effect on the primary observed effects was marginal. Thus, our study mainly concentrates on the main effects of personal attributes. Further, we performed a set of logistic regressions separately for Speed-race and Least-clicks games, incrementally including predictors and presenting the results in Tables 4.6-4.7 to evaluate each predictor's distinct impact on navigation success.

Table 4.3. *The table presents logistic regression results for navigation route success (first two columns) and linear regression results for route uniqueness (last two columns) in Speed-race and Least-clicks games. Coefficients are highlighted in bold when their corresponding variables significantly predict the dependent variable ($p < 0.01$). The dummy variables indicating the eight games are omitted in the results for simplicity.*

	Dependent variable: Success		*Dependent variable: Uniqueness*	
	logistic		*OLS*	
	Speed-race games	Least-clicks games	Speed-race games	Least-clicks games
Age	−**0.053***** (0.008)	−**0.021***** (0.006)	−**0.018**** (0.007)	−0.002 (0.004)
Female	−**0.362**** (0.136)	0.086 (0.118)	−0.127 (0.100)	−0.087 (0.070)
Asian American	**0.812***** (0.181)	0.168 (0.159)	0.215 (0.113)	0.022 (0.089)
African American	−0.452* (0.186)	−0.293* (0.138)	−0.024 (0.146)	−0.021 (0.085)
Foreign Language (Native)	−**0.601**** (0.190)	−0.130 (0.155)	−0.129 (0.127)	0.045 (0.092)
Foreign Language	**0.721***** (0.141)	**0.457***** (0.120)	0.156 (0.096)	0.066 (0.072)
Liberal	0.086 (0.137)	**0.411***** (0.117)	0.125 (0.099)	−0.026 (0.071)
Agreeableness	−0.012 (0.015)	−0.002 (0.013)	0.011 (0.011)	−0.002 (0.007)
Conscientiousness	0.026 (0.015)	0.020 (0.013)	0.007 (0.010)	0.001 (0.007)
Extroversion	−0.031* (0.014)	−0.013 (0.013)	0.0005 (0.010)	−0.004 (0.008)
Neuroticism	−0.007 (0.014)	0.021 (0.012)	0.001 (0.009)	−0.001 (0.007)
Openness	0.018 (0.015)	−0.017 (0.013)	−0.002 (0.010)	0.009 (0.008)
Wikipedia_1	**0.206***** (0.060)	**0.130**** (0.050)	0.024 (0.041)	0.028 (0.029)
Wikipedia_2	0.081 (0.085)	−0.108 (0.078)	0.079 (0.058)	0.028 (0.047)
Spatial_1	0.097 (0.053)	**0.160***** (0.048)	−0.034 (0.035)	0.017 (0.028)
Spatial_2	0.004 (0.059)	0.067 (0.057)	0.014 (0.038)	−0.026 (0.033)
Spatial_3	0.081 (0.068)	−0.111* (0.053)	**0.154**** (0.048)	−0.022 (0.033)
Spatial_4	−0.030 (0.075)	0.120 (0.065)	0.017 (0.048)	−0.049 (0.040)
Employment_1	0.033 (0.048)	−0.014 (0.042)	0.056 (0.034)	0.001 (0.025)
Employment_2	−0.045 (0.058)	−0.107* (0.049)	−0.003 (0.039)	0.009 (0.028)
Employment_3	−0.116 (0.079)	−0.023 (0.067)	−0.018 (0.058)	−0.036 (0.041)
Education_1	0.063 (0.060)	−0.083 (0.048)	−0.091* (0.041)	0.017 (0.029)
Education_2	−0.047 (0.098)	−0.023 (0.081)	0.012 (0.076)	0.063 (0.050)
Computer_1	−0.102* (0.047)	0.051 (0.045)	0.005 (0.034)	0.008 (0.027)
Computer_2	−0.020 (0.084)	−0.072 (0.077)	0.009 (0.056)	−0.030 (0.046)
Prior (Wikigame)	**0.562***** (0.107)	**0.394***** (0.110)	0.041 (0.062)	0.022 (0.057)
Prior (Source Page)	−0.087 (0.068)	−0.021 (0.057)	−0.009 (0.050)	−0.056 (0.034)
Prior (Target Page)	**0.192**** (0.068)	**0.264***** (0.055)	0.011 (0.045)	0.016 (0.033)
Order	**0.103***** (0.023)	0.020 (0.019)	0.028 (0.015)	0.006 (0.011)
Constant	−0.192 (0.745)	−0.489 (0.683)	−0.140 (0.512)	0.095 (0.424)
Observations	1,479	1,662	695	899
R^2			0.105	0.040
Adjusted R^2			0.056	0.00003
Pseudo R^2	0.237	0.117		
Deviance	1644.4	2085.3		
Null Deviance	2045.0	2292.9		
Log Likelihood	−822.205	−1,042.631		
Akaike Inf. Crit.	1,718.410	2,159.261		
Residual Std. Error			1.044 (df = 658)	0.910 (df = 862)
F Statistic			2.140*** (df = 36; 658)	1.001 (df = 36; 862)

$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table 4.4. *The table shows the regression results for navigation performance in the Speed-race games and Least-clicks games, where interactions are excluded or included in the model. The interaction terms are chosen in the following way: firstly, from all the pairwise interaction terms we selected those that are significant predictors of the players' navigation performance ($p < 0.01$) and do not have colinearity issues ($VIF < 5$) when added alone to the logistic regression model without interaction terms. Then we add all the selected interaction terms to the logistic regression model and filtered out those that are not significant ($p < 0.01$) at predicting the players' navigation performance. The interaction terms chosen this way are shown in the table for the Speed-race games and Least-clicks games respectively.*

	Dependent variable: navigation performance			
	Speed-race Games		*Least-clicks Games*	
	Without Interactions	With Interactions	Without Interactions	With Interactions
Age	-0.053^{***} (0.008)	-0.062^{***} (0.009)	-0.021^{***} (0.006)	-0.024^{***} (0.006)
Female	-0.362^{**} (0.136)	-0.405^{**} (0.141)	0.086 (0.118)	0.102 (0.120)
Asian American	0.812^{***} (0.181)	0.585^{**} (0.189)	0.168 (0.159)	0.200 (0.163)
African American	-0.452^{*} (0.186)	-0.505^{**} (0.193)	-0.293^{*} (0.138)	-0.223 (0.139)
Foreign Language (Native)	-0.601^{**} (0.190)	-0.524^{**} (0.194)	-0.130 (0.155)	-0.182 (0.158)
Foreign Language	0.721^{***} (0.141)	0.665^{***} (0.146)	0.457^{***} (0.120)	0.473^{***} (0.122)
Liberal	0.086 (0.137)	-0.062 (0.145)	0.411^{***} (0.117)	0.415^{***} (0.119)
Agreeableness	-0.012 (0.015)	-0.017 (0.015)	-0.002 (0.013)	-0.010 (0.013)
Conscientiousness	0.026 (0.015)	0.036^{*} (0.016)	0.020 (0.013)	0.012 (0.013)
Extroversion	-0.031^{*} (0.014)	-0.027 (0.015)	-0.013 (0.013)	-0.013 (0.013)
Neuroticism	-0.007 (0.014)	-0.001 (0.015)	0.021 (0.012)	0.023 (0.012)
Openness	0.018 (0.015)	0.032^{*} (0.016)	-0.017 (0.013)	-0.021 (0.013)
Wikipedia_1	0.206^{***} (0.060)	0.223^{***} (0.063)	0.130^{**} (0.050)	0.126^{*} (0.050)
Wikipedia_2	0.081 (0.085)	-0.074 (0.097)	-0.108 (0.078)	-0.111 (0.079)
Spatial_1	0.097 (0.053)	0.123^{*} (0.054)	0.160^{***} (0.048)	0.170^{***} (0.048)
Spatial_2	0.004 (0.059)	-0.047 (0.061)	0.067 (0.057)	0.073 (0.058)
Spatial_3	0.081 (0.068)	0.091 (0.072)	-0.111^{*} (0.053)	-0.113^{*} (0.054)
Spatial_4	-0.030 (0.075)	0.065 (0.078)	0.120 (0.065)	0.008 (0.071)
Employment_1	0.033 (0.048)	-0.049 (0.054)	-0.014 (0.042)	0.031 (0.045)
Employment_2	-0.045 (0.058)	-0.092 (0.060)	-0.107^{*} (0.049)	-0.105^{*} (0.049)
Employment_3	-0.116 (0.079)	-0.475^{***} (0.144)	-0.023 (0.067)	-0.049 (0.068)
Education_1	0.063 (0.060)	0.066 (0.062)	-0.083 (0.048)	-0.065 (0.049)
Education_2	-0.047 (0.098)	-0.058 (0.106)	-0.023 (0.081)	-0.043 (0.083)
Computer_1	-0.102^{*} (0.047)	-0.169^{***} (0.050)	0.051 (0.045)	0.031 (0.046)
Computer_2	-0.020 (0.084)	-0.038 (0.087)	-0.072 (0.077)	-0.122 (0.078)
Prior (Wikigame)	0.562^{***} (0.107)	0.712^{***} (0.113)	0.394^{***} (0.110)	0.371^{**} (0.115)
Prior (Source Page)	-0.087 (0.068)	-0.094 (0.071)	-0.021 (0.057)	-0.010 (0.058)
Prior (Target Page)	0.192^{**} (0.068)	0.222^{**} (0.071)	0.264^{***} (0.055)	0.280^{***} (0.056)
Order	0.103^{***} (0.023)	0.107^{***} (0.023)	0.020 (0.019)	0.016 (0.020)
Game Index A1	1.059^{***} (0.251)	1.107^{***} (0.258)	0.442^{*} (0.221)	0.458^{*} (0.223)
Game Index A2	0.746^{**} (0.263)	0.814^{**} (0.270)	0.389 (0.232)	0.449 (0.235)
Game Index B0	1.598^{***} (0.262)	1.757^{***} (0.270)	0.575^{**} (0.214)	0.616^{**} (0.216)
Game Index B1	0.947^{***} (0.264)	1.033^{***} (0.272)	-0.053 (0.218)	-0.026 (0.220)
Game Index B2	0.528^{*} (0.254)	0.603^{*} (0.261)	-0.101 (0.219)	-0.081 (0.221)
Game Index O1	0.757^{**} (0.247)	0.823^{**} (0.253)	0.115 (0.212)	0.128 (0.214)
Game Index O2	0.658^{*} (0.276)	0.742^{**} (0.285)	0.316 (0.238)	0.334 (0.240)
African American:Employment_1		0.401^{**} (0.126)		
Foreign Language (Native):Wikipedia_2		0.762^{***} (0.225)		
Liberal:Employment_3		0.553^{**} (0.181)		
Wikipedia_1:Employment_1		0.209^{***} (0.038)		
Wikipedia_2:Computer_2		0.391^{**} (0.119)		
Asian American:Employment_1				-0.302^{**} (0.112)
Wikipedia_2:Education_2				0.318^{**} (0.116)
Spatial_4:Prior (Wikigame)				0.463^{***} (0.127)
Constant	-0.192 (0.745)	-0.579 (0.785)	-0.489 (0.683)	-0.118 (0.698)
Observations	1,479	1,479	1,662	1,662
Log Likelihood	-822.205	-786.682	$-1,042.631$	$-1,027.605$
Akaike Inf. Crit.	1,718.410	1,657.364	2,159.261	2,135.210
Note:				*p<0.05; **p<0.01; ***p<0.001

Table 4.5. *The table shows the Variance Inflation Factors (VIF) for all the independent variables in the logistic regression models for the navigation performance of the participants in the Speed-race games and Least-clicks games, with and without interactions terms respectively.*

	Variance Inflation Factors			
	Speed-race Games		*Least-clicks Games*	
	Without Interactions	With Interactions	Without Interactions	With Interactions
Age	1.313	1.388	1.530	1.561
Female	1.275	1.305	1.249	1.266
Asian American	1.315	1.360	1.295	1.312
African American	1.211	1.223	1.185	1.195
Foreign Language (Native)	1.467	1.457	1.322	1.332
Foreign Language	1.381	1.399	1.310	1.329
Liberal	1.235	1.316	1.224	1.253
Agreeableness	1.100	1.117	1.135	1.159
Conscientiousness	1.128	1.154	1.119	1.140
Extroversion	1.163	1.169	1.192	1.207
Neuroticism	1.176	1.178	1.074	1.099
Openness	1.111	1.155	1.150	1.150
Wikipedia_1	1.302	1.399	1.348	1.342
Wikipedia_2	1.121	1.387	1.135	1.145
Spatial_1	1.227	1.244	1.236	1.252
Spatial_2	1.147	1.172	1.148	1.155
Spatial_3	1.147	1.192	1.085	1.090
Spatial_4	1.180	1.216	1.111	1.321
Employment_1	1.381	1.584	1.311	1.512
Employment_2	1.233	1.279	1.157	1.172
Employment_3	1.150	3.662	1.165	1.193
Education_1	1.471	1.497	1.415	1.435
Education_2	1.156	1.224	1.122	1.135
Computer_1	1.363	1.430	1.394	1.425
Computer_2	1.171	1.212	1.242	1.255
Prior (Wikigame)	1.307	1.390	1.251	1.253
Prior (Source Page)	1.834	1.853	1.759	1.757
Prior (Target Page)	1.769	1.800	1.536	1.534
Order	1.038	1.043	1.014	1.016
Game Index A1	1.977	1.973	1.822	1.828
Game Index A2	2.280	2.295	2.080	2.089
Game Index B0	1.971	1.984	1.902	1.909
Game Index B1	1.943	1.939	2.040	2.042
Game Index B2	2.008	2.009	1.819	1.820
Game Index O1	1.928	1.931	1.807	1.807
Game Index O2	2.378	2.383	2.233	2.235
Foreign Language (Native):Wikipedia_2		1.499		
Liberal:Employment_3		3.953		
Wikipedia_1:Employment_1		1.185		
Wikipedia_2:Computer_2		1.155		
African American:Employment_1		1.236		
Wikipedia_2:Education_2				1.122
Spatial_4:Prior (Wikigame)				1.270
Asian American:Employment_1				1.324

Table 4.6. *Logistic regression results for the success of the navigation routes in the Speed-race games.*

	Dependent variable: Success								
	Speed-race Games								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Age				−0.046***	−0.043***	−0.046***	−0.045***	−0.046***	−0.046***
				(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Female									−0.217
									(0.121)
Asian American					0.711***	0.791***	0.829***	0.963***	0.942***
					(0.156)	(0.159)	(0.160)	(0.168)	(0.169)
Foreign Language (Native)								−0.501**	−0.493**
								(0.179)	(0.179)
Foreign Language			0.921***	0.802***	0.740***	0.700***	0.662***	0.767***	0.784***
			(0.115)	(0.117)	(0.119)	(0.120)	(0.121)	(0.127)	(0.127)
Wikipedia_1							0.181***	0.206***	0.204***
							(0.053)	(0.054)	(0.054)
Prior (Wikigame)		0.988***	1.019***	0.824***	0.777***	0.739***	0.687***	0.640***	0.621***
		(0.089)	(0.092)	(0.096)	(0.096)	(0.097)	(0.098)	(0.099)	(0.100)
Prior (Target Page)						0.245***	0.205***	0.200***	0.189**
						(0.058)	(0.059)	(0.059)	(0.059)
Order	0.071***	0.081***	0.085***	0.089***	0.091***	0.091***	0.093***	0.096***	0.096***
	(0.020)	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
Constant	−1.038***	−1.634***	−2.174***	−0.612	−0.782*	−1.152***	−1.114**	−1.038**	−0.928**
	(0.181)	(0.199)	(0.217)	(0.319)	(0.325)	(0.341)	(0.343)	(0.346)	(0.351)
Observations	1,479	1,479	1,479	1,479	1,479	1,479	1,479	1,479	1,479
Deviance	1999.427	1858.119	1791.641	1746.891	1725.745	1707.355	1695.463	1687.544	1684.359
Log Likelihood	−999.713	−929.059	−895.821	−873.445	−862.873	−853.678	−847.731	−843.772	−842.180
Akaike Inf. Crit.	2,017.427	1,878.119	1,813.641	1,770.891	1,751.745	1,735.355	1,725.463	1,719.544	1,718.359

Note: *p<0.05; **p<0.01; ***p<0.001

Table 4.7. *Logistic regression results for the success of the navigation routes in the Least-clicks games.*

	Dependent variable:Success							
	Least-clicks Games							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Age						-0.017^{***}	-0.019^{***}	-0.020^{***}
						(0.005)	(0.005)	(0.005)
Foreign Language				0.517^{***}	0.450^{***}	0.438^{***}	0.408^{***}	0.385^{***}
				(0.102)	(0.104)	(0.104)	(0.105)	(0.106)
Liberal					0.431^{***}	0.442^{***}	0.464^{***}	0.435^{***}
					(0.106)	(0.106)	(0.107)	(0.108)
Wikipedia_1								0.098^{*}
								(0.044)
Spatial_1							0.128^{**}	0.123^{**}
							(0.044)	(0.044)
Prior (Wikigame)		0.611^{***}	0.579^{***}	0.576^{***}	0.511^{***}	0.421^{***}	0.429^{***}	0.394^{***}
		(0.096)	(0.098)	(0.099)	(0.100)	(0.102)	(0.102)	(0.103)
Prior (Target Page)			0.277^{***}	0.276^{***}	0.276^{***}	0.270^{***}	0.260^{***}	0.239^{***}
			(0.048)	(0.048)	(0.049)	(0.049)	(0.049)	(0.050)
Order	0.020	0.021	0.018	0.018	0.018	0.018	0.018	0.019
	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Constant	-0.136	-0.297	-0.760^{***}	-1.006^{***}	-1.214^{***}	-0.602^{*}	-0.511	-0.400
	(0.156)	(0.160)	(0.181)	(0.189)	(0.197)	(0.263)	(0.265)	(0.270)
Observations	1,662	1,662	1,662	1,662	1,662	1,662	1,662	1,662
Deviance	2275.619	2231.278	2197.145	2171.371	2154.814	2142.421	2133.731	2128.874
Log Likelihood	$-1,137.809$	$-1,115.639$	$-1,098.573$	$-1,085.685$	$-1,077.407$	$-1,071.211$	$-1,066.866$	$-1,064.437$
Akaike Inf. Crit.	2,293.619	2,251.278	2,219.145	2,195.371	2,180.814	2,170.421	2,163.731	2,160.874
Note:								*p<0.05; **p<0.01; ***p<0.001

Table 4.8. *We tested if participants with certain characteristics are more likely to opt for Speed-race games or Least-clicks games. For binary individual characteristics variables, a Pearson's chi-squared test was conducted; for continuous characteristics variables, a Student's t-test was conducted. Least-clicks games and Speed-race games were encoded as value 1 and 0 in the calculation.*

	chi2	t	pval
Female	34.40		0.00
Asian American	0.87		0.35
Foreign Language (Native)	2.63		0.11
Liberal	3.88		0.05
Spatial_1		4.32	0.00

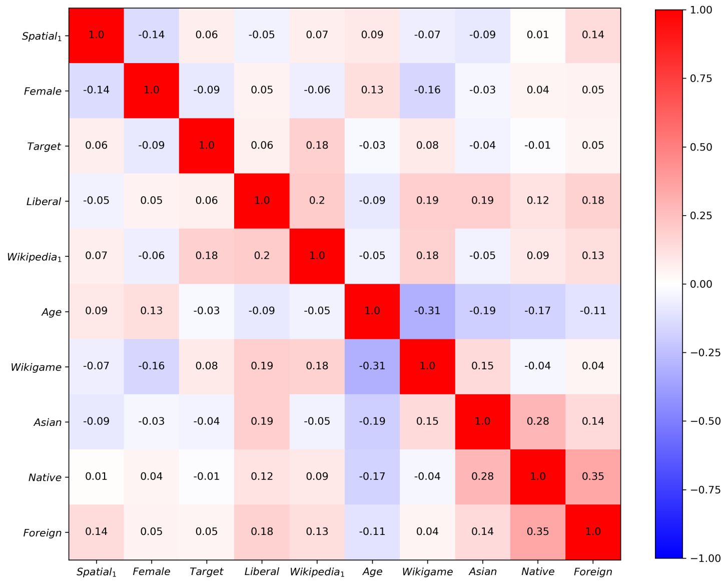

Figure 4.2. *Correlation of the significant predictors of the navigation performance.*

4.3 Results

4.3.1 Impact of individual characteristics on navigation performance

After adjusting for participants' familiarity with the Wikigame, Wikipedia, and the specific target page, fluency in a foreign language (Cohen's d = 0.57) was identified as the strongest predictor of success in both types of games (Table 4.3). Individuals skilled in a foreign language have a 40% greater likelihood of winning a game on average (Fig. 4.3b and Table 4.3), and this factor accounts for 18.7% and 13.6% of the total variance explained in Speed-race and Least-clicks games, respectively (Fig. 4.3h-i). Age also plays a significant role in influencing navigation success across both game modes (Pearson's r(395) = -0.30), with younger players significantly outperforming their older counterparts (Fig. 4.3c and Table 4.3). In particular, age contributes to 12.6% and 6.5% of the variance explained in the respective game types (Fig. 4.3h-i).

Our research indicates that past experience with the Wikipedia navigation game, adeptness in Wikipedia, and knowledge of the specific target Wikipedia page significantly benefit participants across both types of games (Table 4.3). Those who have engaged with the Wikipedia navigation game previously are 1.7 times more likely to succeed in our study than those unfamiliar with the game (Fig. 4.3a). A strong familiarity with the target page is a key predictor of overall success (Table 4.3), especially ranking as the second most influential factor in the least-clicks games, more so than in the speed-race games (Fig. 4.3h-i).

With respect to other personal traits, differences are observed across the two game variants (refer to Figure 4.4 for the distribution of game durations). In the context of games with time restrictions, male players of Asian descent who lack fluency in a foreign language at a native level often exhibit superior performance (Table 4.3, Fig. 4.3d-e). On the other hand, in games governed by distance limitations, improved performance is linked to possessing a liberal viewpoint and a higher self-reported ability in spatial navigation (Table 4.3, Fig. 4.3f-g). Additionally, our analysis shows a significant performance improvement in participants engaging in time-limited games, a trend not seen in the distance-limited games (as highlighted by the Order variable).

4.3.2 Interplay between success and uniqueness

We noted significant diversity in the outcomes and distinctiveness of participants' navigational paths: while certain participants achieved success in all games, others did not succeed in any (Fig. 4.5a). Moreover, some navigators followed popular paths, while others explored more unique routes (Fig. 4.5b). On average, unsuccessful navigation attempts were marked by higher levels of uniqueness when compared to successful ones (Fig. 4.5b). This pattern emerges possibly because becoming lost or straying from the desired route not only increases the likelihood of failure but also elevates the uniqueness score, thereby affecting the overall uniqueness distribution. Figure 4.1 displays the uniqueness scores for successful navigation paths across nine games.

Our regression analysis focusing on the uniqueness scores of navigation paths shows that, similar to success, personal traits play a role in determining

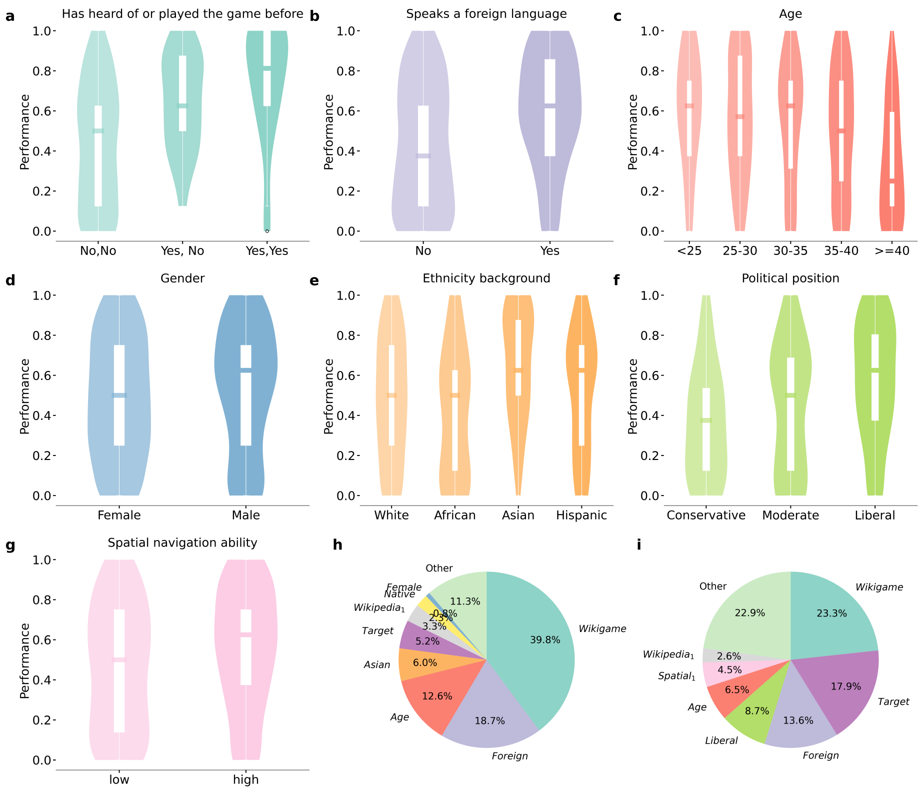

Figure 4.3. *The figure shows the navigation performance distribution for participants with different characteristics, where performance is measured as the ratio of games won by each participant. Fig. a-h shows the distribution of navigation performance for the participants' eight characteristics: a) age, b) gender, c) foreign language skills, d) ethnic background, e) political view, f) spatial navigation skills (the first principal component of the spatial navigation related questions* $Spatial_1$*), g) prior experience with the Wikipedia navigation game and h-i) the percentage of deviance explained by each variable as they were added as the covariate to the regression model (see details in Methods) in the Speed-race game and Least-clicks game respectively, normalized by the total variance explained by the multiple regression of all individual characteristics in Table 4.3.*

the uniqueness of successful routes. In particular, for participants engaging in games with time limitations, younger and left-handed individuals (represented

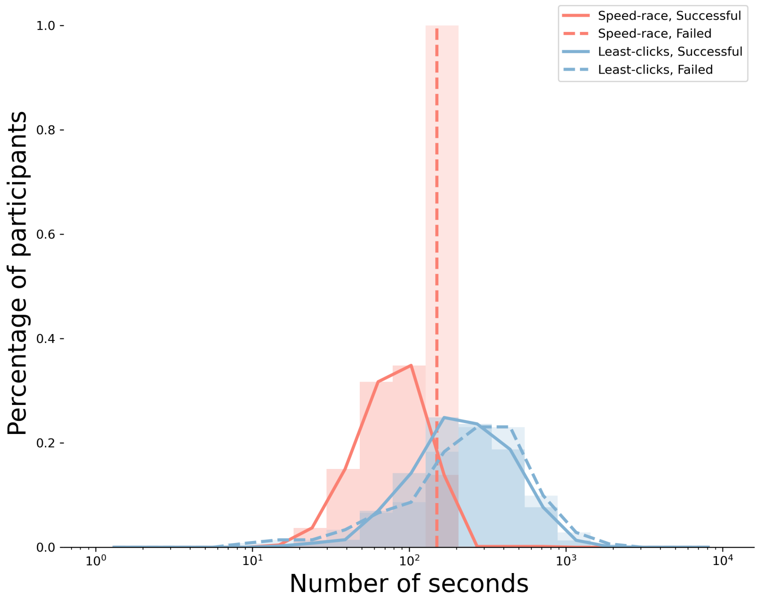

Figure 4.4. *This figure shows the distribution of time (in seconds) spent in the successful and failed Speed-race games and Least-clicks games respectively. For the Speed-race games, the median game time for the successful and failed games are 78 seconds and 150 seconds (the time limit for the game), and for Least-clicks games 214 seconds and 249 seconds.*

by the third principal component of self-reported spatial skills) are more likely to take more unique routes to the target (Table 4.3). However, for participants who opted for games with distance limitations, no specific characteristics were found to significantly correlate with the level of path uniqueness (Table 4.3).

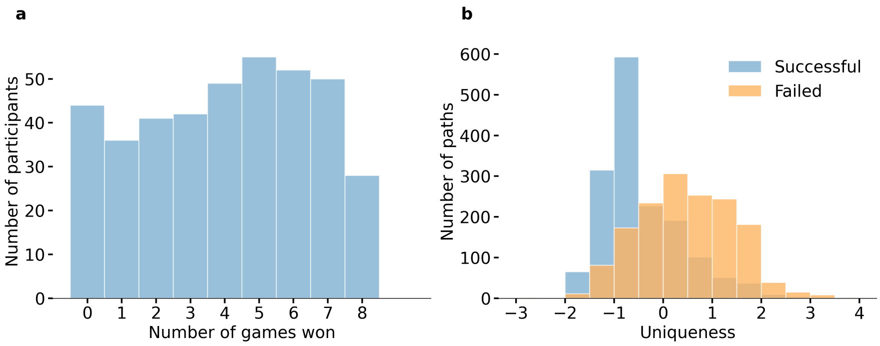

Figure 4.5. *a) shows the distribution of the total number of games the participants won in the experiment. b) shows the distribution of uniqueness scores for both successful and unsuccessful navigation paths. The uniqueness score quantifies the distinctiveness of a navigation path relative to others. For the definition and computation of the uniqueness score, refer to the Methods section.*

4.4 Discussion

Our analysis underscores the impact of personal characteristics on participants' performance in navigating knowledge spaces, with this influence being modulated by constraints like time and distance. We found that previous engagement with Wikipedia, the navigation game, and knowledge of the target page significantly enhance navigation outcomes, likely due to the game's nature. After controlling for these factors, youth and multilingualism emerge as consistent predictors of effective navigation, regardless of constraint type, highlighting the fundamental roles of age and multilingualism in navigating knowledge spaces.

Differences based on other attributes emerge when considering the types of constraints. In Speed-race games, which involve time constraints, participants that are males, or of Asian ethnicity, or with English as the only native language excel. In contrast, for Least-clicks games, which are based on distance constraints, higher performance correlates with liberal political view and superior self-assessed spatial navigation skills. Notably, participants show marked improvement in Speed-race games with additional rounds but not in Least-clicks games. Employing a measure of uniqueness we introduced, we found that individual traits also affect the distinctiveness of routes: under time constraints, younger and left-handed participants favor more unique paths, unlike in distance-constrained tasks.

Prior research has predominantly focused on linking individual differences to navigation within physical spaces. Our study extends this body of work to navigation within knowledge spaces, finding that, similar to physical space navigation [63, 34], age acts as a limiting factor, likely due to cognitive declines associated with aging, which affect fluid intelligence, perceptual speed, memory, and vocabulary [62]. Bilingualism, known for its cognitive benefits, including improved executive control and protection against cognitive decline [97], is identified as the most potent predictor of knowledge space navigation success, suggesting an additional cognitive advantage tied to multilingualism.

Moreover, our findings reveal that individual characteristics such as sex, ethnicity, language proficiency, political orientation, and self-reported spatial navigation skills significantly influence navigation performance in games with either time or distance constraints, but not both. Further investigation is needed to fully understand these effects. One potential explanation is that these factors may be associated with other cognitive processes affecting navigation performance, which were not included in our study. Our results align with research on sex differences in spatial navigation tasks [98, 99, 100], indicating that test timing may moderate these differences due to the impact of anxiety on spatial feature encoding [101, 32], with females typically reporting higher spatial anxiety and lower self-confidence than males [102, 66]. Interestingly, we did not observe a significant role for the Big Five personality traits [103] in our experiment, despite their expected impact on online navigation performance [37], possibly due to differences in navigation tasks and experimental settings. Inspired by our findings, we aim to conduct additional experiments incorporating more objective cognitive measures to derive more definitive conclusions.

Creativity, traditionally defined, encompasses both originality and effectiveness[104, 105]. Our findings go beyond merely identifying successful navigation (effectiveness) to show that the uniqueness of routes (originality) is also shaped by individual variances, in line with prior studies that have shown a tendency for frequent detour usage in navigation[106], with such detours reflecting personal attributes[107]. The traits we identified as predictors of successfully navigating to a target do not always align with those indicating an ability for creative route selection.

The implications of our study are widespread, particularly concerning governmental digital service practices. The "online only" approach has been critiqued by researchers pointing out that certain groups, especially older adults, might find it challenging to access online services, suggesting the necessity for alternative access methods[108]. Despite near-universal internet access in developed nations, the experience of being online varies significantly based on individual differences. Simply being available on the Internet doesn't guarantee accessibility for all.

Our research comes with several limitations. Firstly, for more robust outcomes, it would be beneficial to include additional variables that capture participants' engagement, working memory, levels of anxiety, and specific spatial skills, which may influence navigation performance. Secondly, the context of our findings is limited to a controlled navigation game scenario involving Wikipedia pages, which might not directly apply to broader real-world navigation tasks. Thirdly, our focus is mainly on navigating knowledge spaces, a subset of the wider online information-seeking behavior, distinct from searching for knowledge. Lastly, due to participants' ability to choose their game type, we noticed a notable self-selection bias based on gender and self-reported spatial skills (refer to Table 4.8), which must be considered in interpreting our results. To overcome these challenges, we're conducting a subsequent experiment with additional moderating variables.

Extending from previous inquiries into spatial navigation, our study delves into navigation within knowledge spaces. Future research could look into developing mathematical models that incorporate individual differences to further understand navigation behavior. Investigating ways to improve navigation experiences for people with specific characteristics through future experiments presents an interesting direction for further exploration.

CHAPTER 5

NAVIGATION STRATEGIES

5.1 Introduction

Navigating from one location to another is essential for animals, allowing them to find crucial resources like food, partners, and shelters[3, 15]. This quest for resources isn't limited to physical spaces but extends to abstract spaces, such as seeking assistance in social spaces[19, 4], or looking for answers online in knowledge spaces[5]. With the explosion of information on the internet over recent decades, we've encountered the challenge of information overload, making efficient navigation in the information space increasingly vital[6]. Understanding how we navigate this information space is the first step in addressing this challenge.

The concept of navigation in social space can be traced back to a 1929 thought experiment by Hungarian author Frigyes Karinthy[7]. Witnessing the burgeoning advancements in communication technologies that seemed to shrink the distances between people, Karinthy suggested that any two people on Earth could be connected through a short chain of acquaintances. This idea was empirically implemented decades later by Stanley Milgram[8],where he asked individuals in Omaha, Nebraska, and Wichita, Kansas, to try and send a letter to a person in Boston, Massachusetts, by handing it off through a chain of friends. Milgram found that through about six intermediaries on average the letter could reach the target person. This concept was further tested in 2003 via an email-based experiment[9] involving roughly 100,000 participants across 13 countries, which essentially replicated Milgram's results, demonstrating humans' profound ability to navigate social networks.

Research has linked our efficient social navigation capability to the structure of social networks. Watts et al.[20] noted that social connections are highly organized, with individuals possessing various identities and affiliating with groups defined by specific social attributes. This results in natural hierarchical group structures, similar to organizational structures in universities or corporations, enabling navigable social networks through simple linking rules. Utilizing a greedy decentralized search algorithm—where the next step chosen is always the closest to the target—allows for reaching any individual within a few steps. Kleinberg[21] theoretically established that networks formed from a hierarchical tree graph under certain linking conditions could be efficiently navigated using such an algorithm, a principle empirically supported by Adamic et al.[22], who demonstrated that the organizational hierarchy could facilitate finding short paths within an email network. Typically, individuals leverage either geographical or occupational hierarchies to navigate social networks[9, 23].

Extending these principles of social navigation to knowledge spaces, attention has shifted towards online information-seeking behaviors, especially on Wikipedia[24], known for its extensive range of topics and significant user interaction. Wikipedia navigation games like Wikispeedia[25] and the Wiki Game[26] have gained popularity, tasking players with navigating from one Wikipedia article to another via hyperlink chains. The digital footprints left by game participants provide a valuable dataset for analyzing human navigation behavior in knowledge spaces. West et al.[27] observed that players tend to utilize both the degree of articles and their textual similarity to the target page as navigational aids on Wikipedia, with the degree playing a more crucial role in the early stages of navigation and textual similarity becoming more important as players near their target. Compared to the greedy decentralized search model, human players display more bias and stochasticity in their navigation choices, with topic level decisions influenced by the previous steps and a tendency towards random selection in the game's initial phases[28, 29].

While these studies have illuminated aspects of how we navigate social and knowledge networks, a thorough understanding of how navigation varies based on individual preferences and its implications remains to be fully explored. Milgram's experiments showed reliance on geographical and occupational information for social navigation, yet the reasons behind these preferences and their impacts are not fully understood. Furthermore, factors such as age, gender, and origin have been shown to significantly influence navigational performance in physical spaces[30, 31, 32, 33, 34], raising questions about potential disparities in navigating knowledge spaces. However, the lack of comprehensive demographic data on participants in previous studies, both in social navigation experiments and Wikipedia navigation games, has limited the investigation into how individual traits affect navigation patterns.

In this chapter[1], using a graph embedding trained on the English Wikipedia network, I aim to identify the diverse navigation strategies employed by participants and examine how these strategies are influenced by the information landscape and individual characteristics.

5.2 Methods

5.2.1 Embedding of the Wikipedia articles

To measure the similarity between Wikipedia articles, we created a 64-dimensional node embedding for each article a_i within the English Wikipedia network G, using the DeepWalk algorithm[90]. Graph embedding is a technique that maps each node of the graph to a numerical vector $\vec{v}_i$ in a multidimensional space, placing similar nodes close together. This approach allows for the evaluation of node dissimilarity through the calculation of the distance between their vectors. In our case, we assign a 64-dimensional vector $\vec{v}_i$ to each Wikipedia article a_i, which facilitates the definition of a semantic distance measure between article pairs:

$$d(a_i, a_j) = 1 - \frac{\vec{v}_i \cdot \vec{v}_j}{\|\vec{v}_i\| \|\vec{v}_j\|},$$

where $d(a_i, a_j)$ represents the semantic distance between the articles a_i and a_j, calculated using the cosine distance between their embeddings $\vec{v}_i$ and $\vec{v}_j$. For

[1]This chapter is based on the article "Milgram's experiment in the knowledge space: Individual navigation strategies" [2].

the purpose of these calculations, all vector embeddings are normalized to unit length unless stated otherwise. To evaluate the accuracy of our embedding, we utilized the WikipediaSimilarity 353 Test[91], which is based on the WordSimilarity 353 Test[92] and assesses semantic similarity between words. Our embedding method achieved a Spearman rank correlation of 0.667 on the WikipediaSimilarity 353 test, indicating that it matches the leading current methods for measuring semantic relatedness of Wikipedia articles[93].

5.2.2 Categorization of the visited articles

The graph embedding approach allows for the classification of visited Wikipedia articles based on their semantic distances. For this classification, we applied the KMeans clustering algorithm[109], classified the visited articles into three groups in each game. The Euclidean distance was used to measure the distance between articles for clustering, as KMeans operates by calculating the centroids of data points. To account for the different visitation frequencies of articles, we assigned a weight $w_i = \log(n_i) + 0.1$ to each article a_i, where n_i is the number of times the article was visited, and the addition of 0.1 ensures that articles with only one visit still get a minimal weight. Figure 5.1 shows the clusters in a two-dimensional space, with the x-axis and y-axis representing the first and second principal components of the embedding vectors, respectively, after dimensionality reduction through Principal Components Analysis[110]. The clusters were named as follows: the cluster containing the source page was designated the Source Group, the cluster with the target page was named the Occupation Group, and the third cluster was identified as the Geography Group.

This clustering achieved an average Silhouette coefficient of 0.18 (SD = 0.020) across all games. When the distance between articles is considered within the two-dimensional space of the first two principal components, the average Silhouette coefficient increases to 0.45 (SD = 0.052). It should be noted that the primary goal is not to attain perfect clustering but to extract insights at a more coarse-grained level.

5.2.3 Wikipedia articles characterization

To explore the knowledge users depend on for navigation within the Wikipedia network, our study examined Wikipedia pages from two perspectives: their semantic connections and hierarchical positions. For semantic relationships, which pertain to how articles' meanings or contexts relate, we utilized the 64-dimensional embedding trained (detailed in 5.2.1), where each article is represented as a vector in a space where articles on similar topics are closer together, allowing us to assess the "closeness" of two articles a_i and a_j through the cosine similarity of their vectors $\vec{v}_i$ and $\vec{v}_j$:

$$c(a_i, a_j) = 1 + cos(\vec{v}_i, \vec{v}_j) \tag{5.1}$$

For the hierarchical analysis, we assessed articles based on their network connections, applying a hierarchical score[111] to determine each article's rank. This score $h(a_i)$ is computed by considering the number of links an article receives (in-degree, $k_{in}(i)$) and sends (out-degree, $k_{out}(i)$) as shown in Equation 5.2, positioning an article higher in the hierarchy if it has numerous inbound

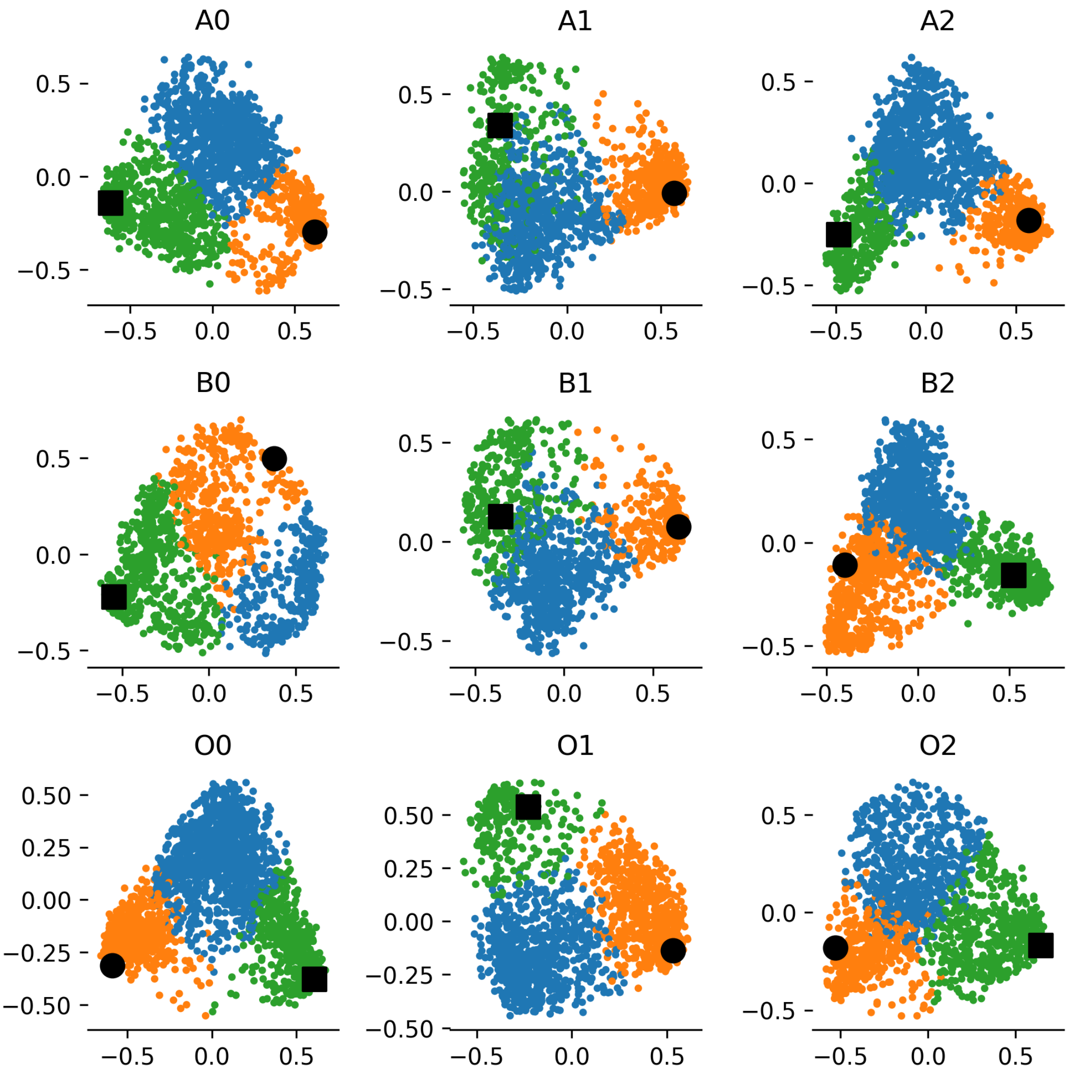

Figure 5.1. *The figure visualizes the visited Wikipedia articles in each of the nine games. The articles belonging to the Geography, Occupation and Source group (see Methods 5.2.2 for the classification method) are shown as blue, orange and green dots respectively. The source page and target page of each navigation game are shown in black square and circle. The two axis represents the first two principal components after dimension reduction using principle components analysis on the original 64 dimensional embedding of all the visited articles.*

and outbound links:

$$h(a_i) = \frac{k_{in}^{3/2}(i) + k_{out}^{3/2}(i)}{k_{in}(i) + k_{out}(i)} \tag{5.2}$$

Generalizing the closeness and hierarchical score from the article level to the path level, For each navigation path we computed the average hierarchical score (H_j) and the average closeness score (C_j). This approach considers each navigation path j, composed of a series of articles $A_j = \{a_k\}$, where N_j represents the total number of articles in that path. For each article a_k in the path, we calculated its hierarchical score ($h(a_k)$) and its closeness to the target page ($c(a_k)$).

$$H_j = \frac{1}{N_j} \sum_{a_k \in A_j} h(a_k) \tag{5.3}$$

$$C_j = \frac{1}{N_j} \sum_{a_k \in A_j} c(a_k, target) \tag{5.4}$$

5.2.4 Generation of synthetic paths

To investigate if the choice of navigation strategies by participants is influenced by the network's structure, we created 10 synthetic paths for each successful navigation path, maintaining the same length as the original. The process for generating a synthetic path for an empirical navigation path with n steps is as follows. Beginning at the source page, we identify all out-neighbors of the source that can reach the target within n steps on the Wikipedia network and randomly select one to be the next step. After moving to this selected step, we gather all out-neighbors of this new step, choosing randomly among those that can reach the target in $n-1$ steps. This selection process is continued, reducing the steps by one each time, until the target page is reached and selected as the final step.

5.3 Results

5.3.1 Navigation strategies

The vector representations of articles enabled a quantitative analysis of navigation strategies by participants. Figure 5.2 showcases the visited articles and the paths taken by participants from 'Barack Obama' to 'Vincent van Gogh' (refer to Figures 5.3 for visualizations across all nine games). Three primary navigation strategies are evident: one group followed the target's occupational aspect, heading directly from 'Barack Obama'-related pages to art-themed ones (orange paths); another utilized geographic information about the target, moving first to pages about 'Netherlands' and 'France' before heading to the target (blue paths); and a third combined both strategies, starting with European countries or cities and then transitioning through art-related pages to reach the target (red paths).

Initially, we clustered the visited articles into three categories based on their pairwise distances using the KMeans clustering algorithm[109] (refer to Methods 5.2.2 for details). Across all nine games, three distinct clusters were consistently identified, each serving unique navigational purposes:

- Occupation Group $\mathcal{O}$: This cluster includes articles semantically close to the target page (see Figure 5.4 for distribution of visited articles' proximity to the target page in the Geography, Occupation, and Source groups for each game). For example, in games where "Vincent van Gogh" is the target, art-related articles dominate this group, whereas for targets like "Albert Einstein" science-related pages would dominate this group. These articles, typically highly interlinked, serve as direct navigation objectives for participants, depicted as orange dots in Figure 5.2a.

- Geography Group $\mathcal{C}$: Comprising articles related to countries or cities associated with the target, shown as blue dots in Figure 5.2a. Though these articles are not as semantically close to the target as those in the Occupation Group (see Figure 5.4), they connect to a broader range of topics, making them intermediate goals for participants who expect links to the target page within these articles.

- Source Group $\mathcal{S}$: Consists of articles closely related to the source page (see Figure 5.5 for distribution of visited articles' proximity to the source page across the Geography, Occupation, and Source groups in each game), represented as green dots in Figure 5.2a. These articles are starting points, with users anticipating paths leading towards the Occupation or Geography Groups.

Analyzing successful navigation paths revealed that the last Wikipedia article clicked before reaching the target often belongs to either the occupation or geography groups. This observation allows for categorizing paths as 'occupational paths' (orange lines in Figure 5.3) if ending with an article from the Occupation Group, or 'geographical paths' (blue lines in Figure 5.3) if concluding with a Geography Group article. Paths involving articles from both groups are labeled "mixed paths" (red lines in Figure 5.3), showing elements of both strategies, while rare paths ending with a Source Group article are called 'other paths' (green lines in Figure 5.3). Figure 5.2a visualizes these strategies with color-coded lines, showing that this categorization effectively distinguishes the

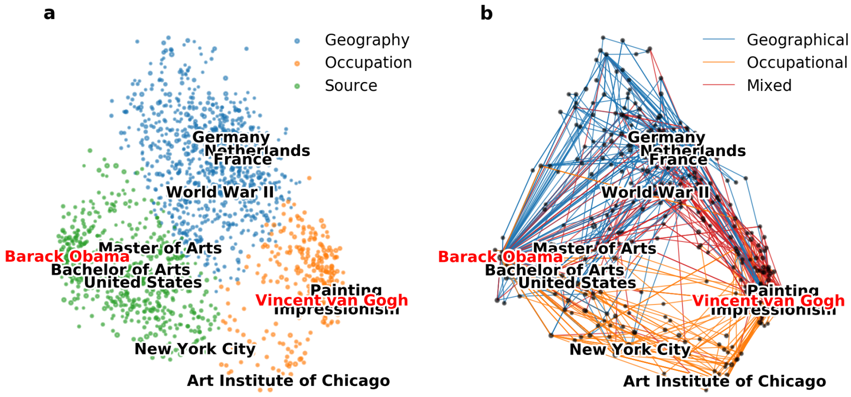

Figure 5.2. *The figures show the visited Wikipedia articles (figure a) and the successful navigation paths (figure b) in the game with the source page "Barack Obama" and the target page "Vincent van Gogh". Articles belonging to the Geography, Occupation and Source group (see Methods 5.2.2 for the classification method) are shown as blue, orange and green dots in figure a, and paths categorized as geographical, occupational and mixed are shown as blue, orange and red lines in figure b. Note that the set of mixed paths overlap with the sets of geographical paths and occupational paths. The geographical/occupational paths here refer to the paths in respective sets excluding the mixed paths.*

three main types of routes observed. For visualizations of all nine games, see Figure 5.1 and Figure 5.3.

5.3.2 Hub-driven vs proximity-driven approach

The split between occupational and geographical strategies, can be better understood by drawing parallels to the way transportation networks operate. In transportation, particularly on roads where shortcuts between distant locations are scarce, travelers tend to choose paths that are physically nearer to their ultimate destination because they must pass through nearby locations. On the other hand, in transportation systems with many shortcuts, like airlines, using major hubs that connect to various locations, albeit not always directly close to

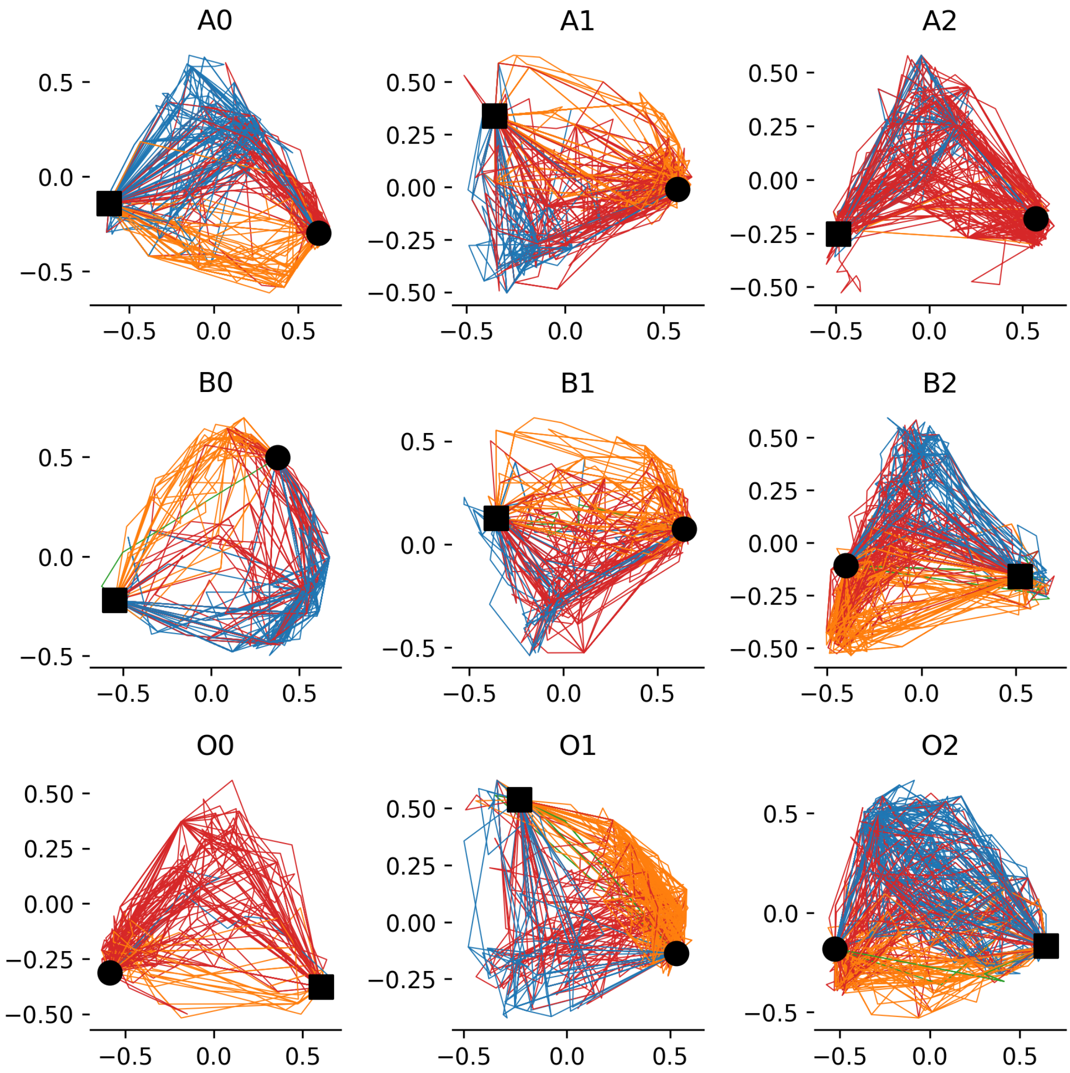

Figure 5.3. *The figure visualizes the successful navigation paths in each of the nine games. The paths categorized as geographical, occupational and mixed are shown as blue, orange and red lines respectively. The source page and target page of each navigation game are shown in black square and circle. The two axis represents the first two principal components after dimension reduction using principle components analysis on the original 64 dimensional embedding of all the visited articles.*

the final destination, becomes a more efficient strategy. This concept is mirrored in knowledge space navigation if we introduce the closeness among articles and

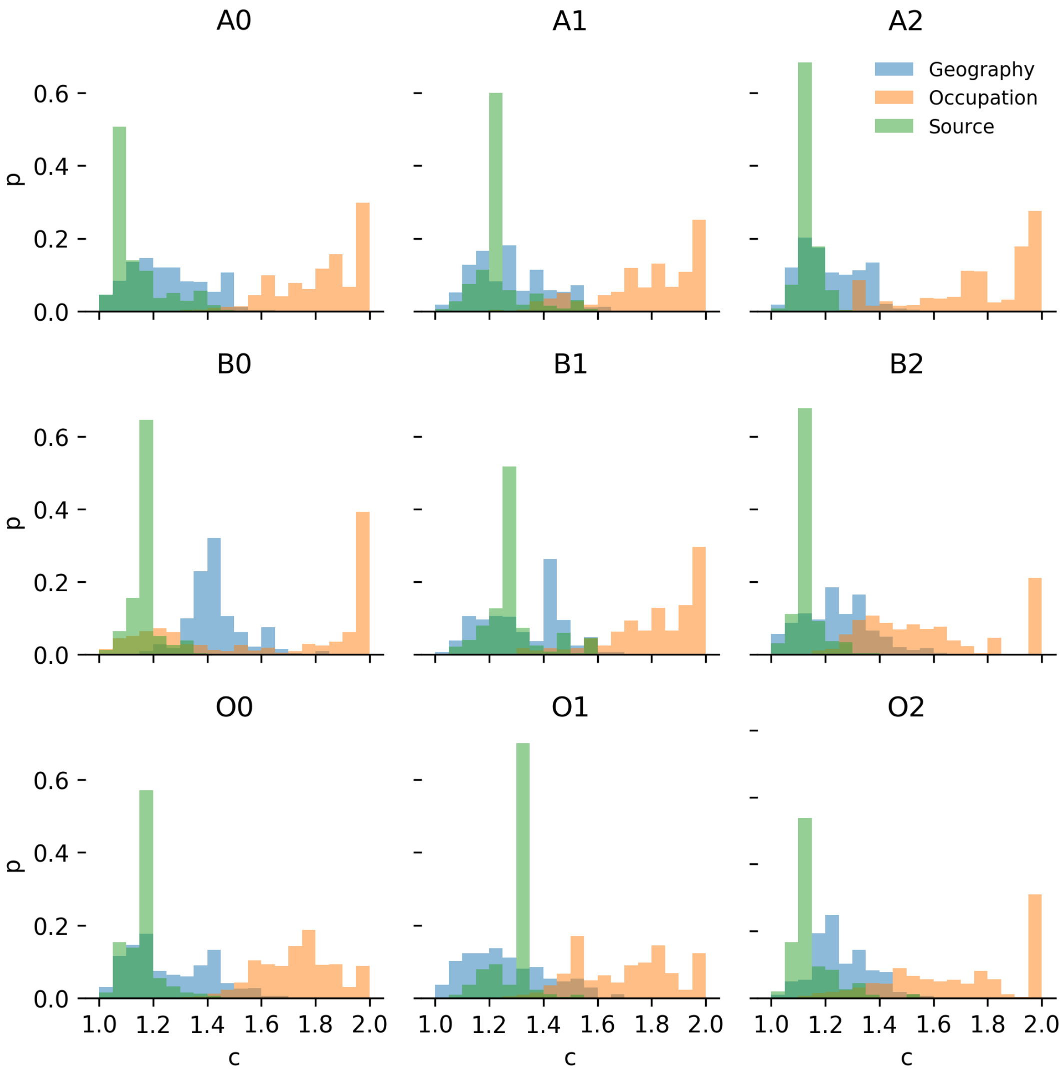

Figure 5.4. *The figure shows the distribution of the visited articles' proximity (measured as the cosine proximity defined in Eq. 5.1) to the target page in the Geography, Occupation and Source groups respectively (see classification in Method 5.2.2). The sub-figures show the distribution for each of the nine navigation games respectively.*

the articles' hierarchical scores (detailed in 5.2.3). Figure 5.6a-b displays the semantic closeness and hierarchical positioning of the last articles clicked on by users following occupational and geographical navigation paths, respectively. Occupational paths tend to end with clicks closer to the target page than ge-

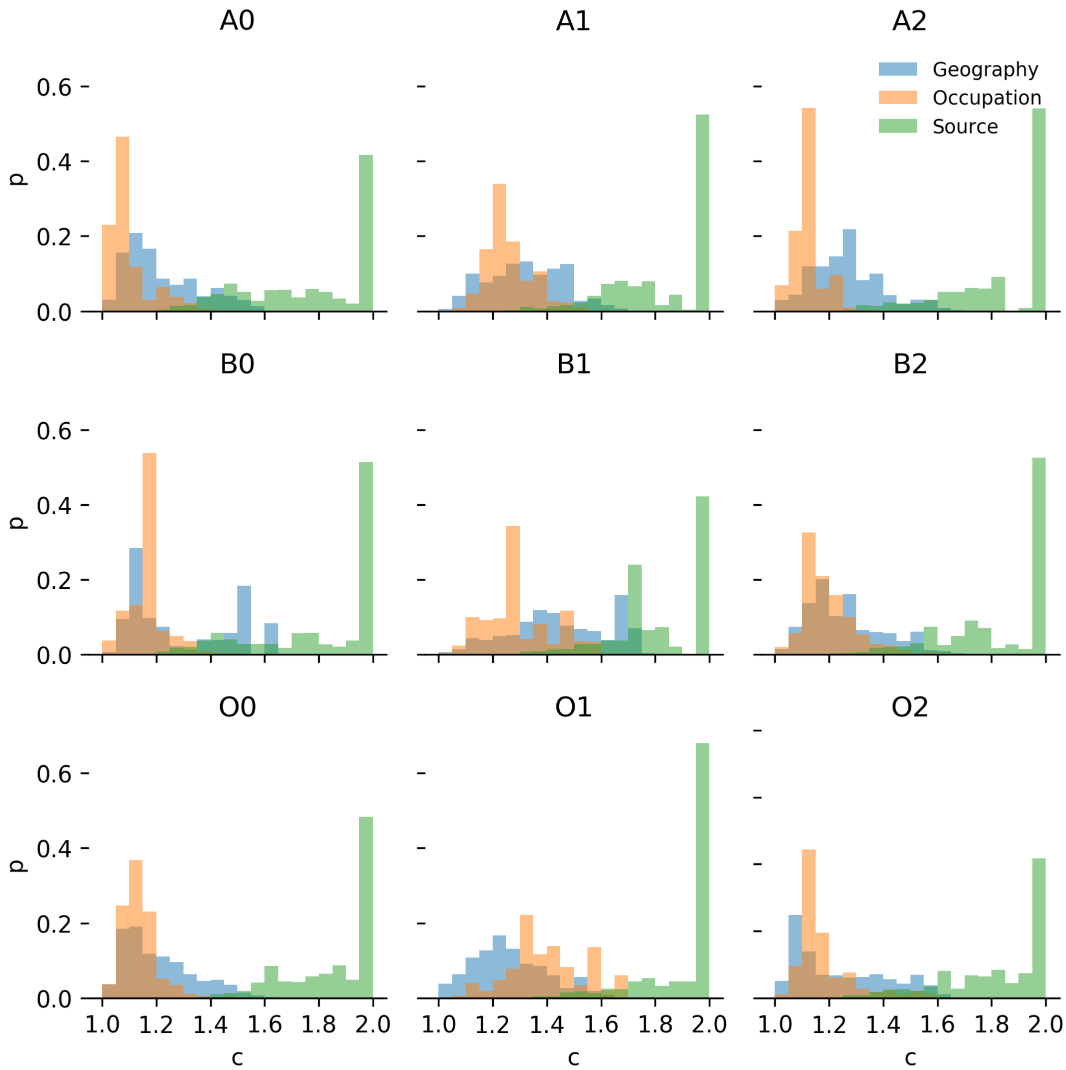

Figure 5.5. *The figure shows the distribution of the visited articles' proximity (measured as the cosine proximity defined in Eq. 5.1) to the source page in the Geography, Occupation and Source groups respectively (see classification in Method 5.2.2). The sub-figures show the distribution for each of the nine navigation games respectively.*

ographical paths, indicating a preference for proximity, whereas geographical paths often involve articles that are higher up in the knowledge hierarchy, suggesting a reliance on hubs. This distinction is further exemplified in Figure 5.6c, showcasing the semantic distance between the last clicked articles and the target page. Note that game O0 was excluded due to its anomalously low success rate and limited geographical path data.

Figure 5.7 illustrates these navigation paths with respect to the navigation paths' hierarchical score H and closeness score C defined in Equation 5.3, highlighting the difference between the proximity and hub strategies on the whole path level. Some paths exhibit a tendency to rise through the hierarchy to access hubs before converging on the target, while others aim to stay as close as possible to the target throughout. Our findings, as illustrated in the subfigures of Figure 5.7, show that occupational paths predominantly exhibit a proximity-driven strategy, with a significant statistical result (t-value M=19.04, SD=6.33), indicating these paths' articles are generally closer to the target page. On the other hand, geographical paths are more inclined towards a hub-driven strategy, as indicated by their hierarchical positioning with a distinct statistical outcome (t-value M=9.32, SD=3.22), even though some occupational paths also make strategic use of hubs to reach their destination, thereby indicating a blend of strategies.

The effectiveness of each approach is evaluated by analyzing player performance in different game settings through linear regression models, focusing on successful paths. Performance is gauged by the time and number of steps saved in Speed-race and Least-clicks games, respectively. The results, presented in Table 5.1, reveal that the navigation strategy's impact varies with the game's timing conditions. In the Least-clicks game, where time isn't a factor, both strategies enhance performance. However, in Speed-race games, where time is of the essence, the hub strategy proves advantageous while the proximity strategy can be detrimental, likely due to the additional time required to find pages closely related to the target compared to jumping to a well-connected page.

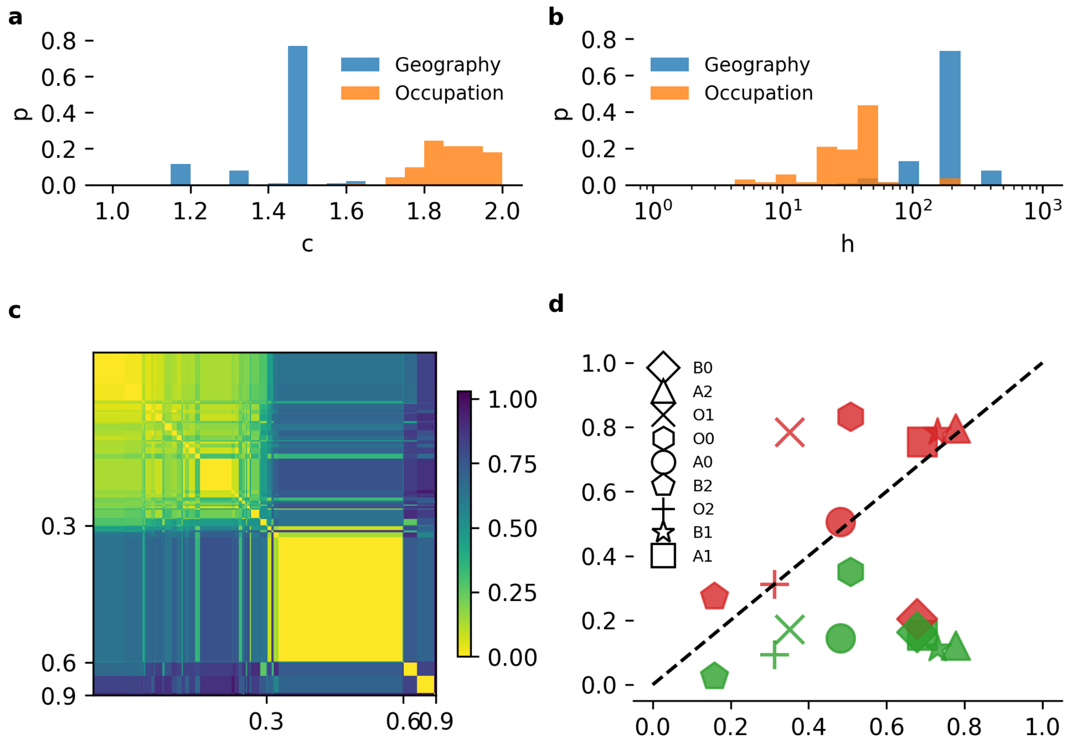

Figure 5.6. *Figure a-b show the distribution of the last clicked articles of the geographical and occupational navigation paths in terms of their closeness c (defined in Eq. 5.1) to the target and hierarchical score h (defined in Eq. 5.2) on the Wikipedia network. Figure c shows the pairwise distance among the last clicked articles ordered by its cosine distance to the target page. Figure d shows on the horizontal axis the ratio of close-by in-neighbors of the target page over all in-neighbors, and on the vertical axis the ratio of occupational paths for players' navigation paths (red) and synthetic navigation paths (green).*

5.3.3 Wisdom of the crowd

To understand the reasons why individuals favor certain navigation strategies over others—whether it stems from personal preferences or the architecture of the knowledge space—we began by dissecting the division between geograph-

Table 5.1. *The table shows the linear regression results for the fitness of hub-driven and proximity-driven navigation strategies, measured as the seconds saved in the Speed-race games or steps saved in the Least-clicks games.*

	Dependent variable:	
	Seconds saved Speed-race games	Steps saved Least-click games
Steps	-7.066^{***} (0.345)	
Seconds		-0.002^{***} (0.0001)
Hub-driven score	3.248^{**} (1.051)	0.405^{***} (0.041)
Proximity-driven score	-3.259^{**} (1.103)	0.200^{***} (0.039)
Source page knowledge	3.268^{**} (1.116)	0.030 (0.037)
Target page knowledge	1.436 (1.032)	0.034 (0.036)
Game Round	1.027^{**} (0.321)	-0.012 (0.011)
Constant	107.372^{***} (4.010)	3.209^{***} (0.110)
Observations	1,174	1,707
R^2	0.367	0.203
Adjusted R^2	0.359	0.196
Residual Std. Error	27.904 (df = 1159)	1.182 (df = 1692)
F Statistic	47.984^{***} (df = 14; 1159)	30.765^{***} (df = 14; 1692)
Note:		*p<0.05; **p<0.01; ***p<0.001

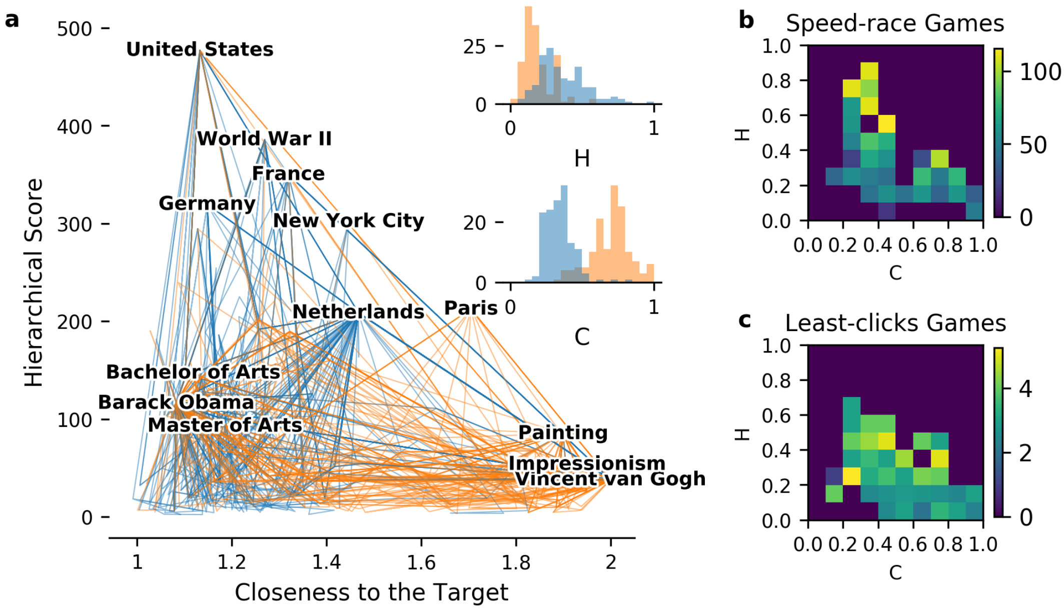

Figure 5.7. *Figure a shows the navigation paths of the participants for the game with the source article "Barack Obama" and target article "Vincent van Gogh". The horizontal axis shows the visited articles' closeness to the target, measured as the cosine similarity between the respective article and the target article in the 64 dimensional embedding space (Eq. 5.1), and the vertical axis shows the hierarchical score of the respective article calculated in Eq. 5.2 using the in-degree and out-degree of the article on the English Wikipedia network. The geographical and occupational paths are shown in blue and orange lines and the subfigures in Figure a show the distribution of H and C for paths in the respective categories. Figure b and c visualize the average seconds/steps saved in the respective types of games at given closeness score C and hierarchical score H intervals.*

ical and occupational navigation paths. This bifurcation illuminates whether participants tend toward the target through proximate articles or prefer engaging with more remote hubs. Our methodology involved analyzing all incoming neighbors of the target page within the Wikipedia network, distinguishing those that are nearby (cosine distance below 0.3) from those more distant (cosine distance above 0.3). We then contrast this analysis with player strategies: how often players achieved their target by leveraging nearby neighbors in their final step (occupational) versus opting for distant neighbors (geographical). Our findings showed that the proportion of occupational paths closely mirrors the ratio of proximate in-neighbors of the target page (adjusted R squared = 0.96) in six out of nine games, with the exceptions (O0, B0, O1) ranking among the games where players showed least familiarity with the target pages (as detailed in Table 5.2).

To discern if this phenomenon was influenced by the structure of the Wikipedia network itself, we generated 10 synthetic paths for each empirical navigation path, ensuring the path length remained consistent (as detailed in Methods 5.2.4). The analysis of synthetic paths indicated a pronounced inclination towards the geographical strategy (as depicted in Figure 5.6), a trend that persists even when adjusting the cosine distance threshold (illustrated in Figure 5.8). These insights hint at an underlying "wisdom of the crowd" effect, highlighting our collective capability to adeptly navigate the knowledge terrain surrounding a target page and devise our pathways accordingly.

To understand the impact of individual characteristics on the choice of navigation strategies during the navigation task, we devised several regression models. These models incorporated personal traits (outlined in Chapter 3) and attributes from current and prior game rounds as predictors, with the preference for geographical or occupational navigation paths as the dependent variable. We conducted separate logistic regression analyses for the initial and subsequent rounds of the experiment to identify only those relationships that exhibited enduring significance. According to the regression outcomes presented in Table 5.3, there wasn't a single personal characteristic that consistently influenced the selection of navigation strategies across the two rounds. Altering the dependent variable to be the extent of hub-driven (H) versus proximity-driven (C) tendencies in navigation did not change this outcome significantly; individual characteristics largely did not affect navigation choices in a consistent manner. An interesting exception was observed with left-handed participants, who demonstrated a pronounced preference for hub-driven strategies in both experiment rounds, as detailed in Table 5.4. The experimental conditions, particularly the timing constraints of the games, were also significant factors influencing navigation preferences. In Speed-race games, which are time-bound, a hub-driven approach was more favored. Conversely, in Least-clicks games, where the objective is to minimize the number of steps, a proximity-driven strategy was predominantly chosen. This highlights how the external conditions of a task, alongside individual differences like handedness, can shape the strategies people adopt in navigating information spaces.

Table 5.2. *The table shows the R_0 and R_{emp} defined in Results 5.3.3 and the participants' average prior knowledge of the target page in each of the nine games, ordered by increasing order.*

	Source page	Target page	R0	R_emp	Prior knowledge
B0	Donald Trump	Pyotr Ilyich Tchaikovsky	0.68	0.21	0.87
A2	Marie Curie	Chuck Berry	0.78	0.80	1.06
O1	Steve Jobs	Charlie Chaplin	0.35	0.79	1.54
O0	Alexander the Great	Tim Burton	0.51	0.83	1.58
A0	Barack Obama	Vincent van Gogh	0.48	0.51	1.76
B2	Angelina Jolie	Charles Darwin	0.16	0.27	1.97
O2	Elizabeth I of England	Albert Einstein	0.31	0.31	2.03
B1	Jeff Bezos	Kanye West	0.73	0.78	2.32
A1	Bill Gates	Eminem	0.69	0.75	2.35

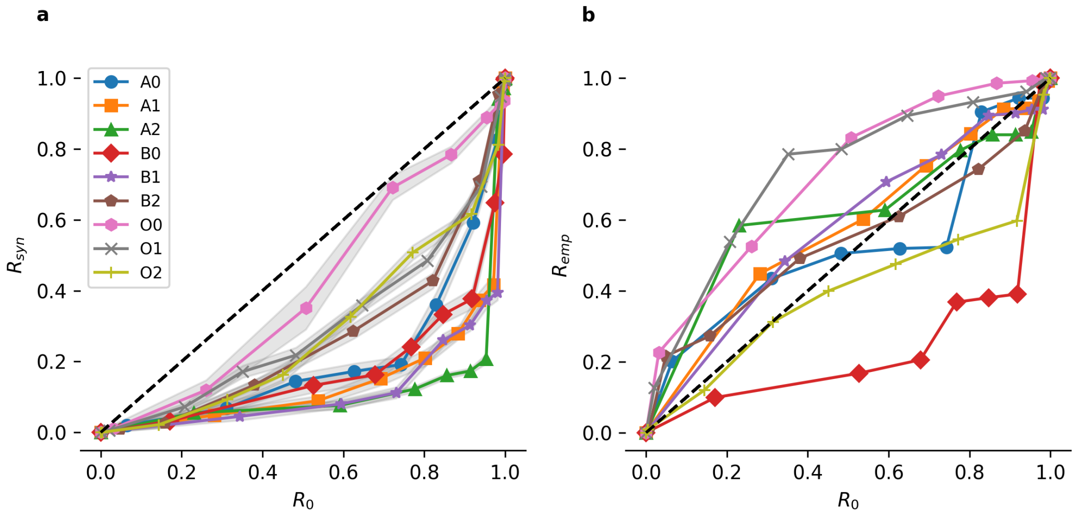

Figure 5.8. *The figure shows the ratio R_0 of the target page's in-neighbors on the Wikipedia network that are within certain distance threshold to the target page over all in-neighbors; the ratio R_{emp} of the successful navigation paths whose last clicked article is within certain distance threshold to the target page over all successful paths; and the ratio R_{syn} of the synthetic navigation paths whose last clicked article is within certain distance threshold to the target page over all synthetic paths. The distance threshold, measured as the cosine distance to the target page, varies from 0 to 1 in 0.1 increments. Figure a shows the relationship between R_0 (x axis) and R_{syn} (y axis) for all nine games, with the shade representing the standard deviation of R_{syn} among the 10 simulation trials. Figure b shows the relationship between R_0 (x axis) and R_{syn} (y axis) for all nine games.*

Table 5.3. *The table shows the logistic regression results for the geographical and occupational navigation strategies. Please note that only variables that are significant (p < 0.01) in at least one round of the experiment are shown in the table. Additionally, variables representing which game was played are also omitted from the table for better readability.*

	Dependent variable:			
	Is Geographical Path		Is Occupational Path	
	First Round	Second Round	First Round	Second Round
Play computer games frequently	0.088 (0.083)	0.387*** (0.116)	−0.088 (0.082)	−0.412*** (0.116)
Like to play computer games	0.094 (0.137)	−0.506** (0.176)	−0.077 (0.135)	0.528** (0.175)
Left-handed	0.659*** (0.192)	−0.237 (0.261)	−0.587** (0.189)	0.100 (0.257)
Job is intensive	0.197* (0.084)	0.053 (0.101)	−0.216** (0.082)	−0.067 (0.100)
Age	−0.031** (0.010)	−0.006 (0.010)	0.032** (0.010)	0.008 (0.010)
Speaks a foreign language	0.258 (0.158)	0.524** (0.193)	−0.267 (0.156)	−0.554** (0.192)
Current game round	0.032 (0.029)	0.107** (0.037)	−0.030 (0.029)	−0.104** (0.036)
Constant	1.424 (1.081)	−1.944 (1.408)	−1.098 (1.066)	1.860 (1.405)
Observations	1,495	1,095	1,495	1,095
Log Likelihood	−683.128	−474.402	−700.664	−477.243
Akaike Inf. Crit.	1,458.256	1,040.804	1,493.328	1,046.486
Note:				*p<0.05; **p<0.01; ***p<0.001

Table 5.4. *The table shows the linear regression results for the hub-driven and proximity-driven scores for the Speed-race games and Least-clicks games. Please note that only variables that are significant ($p < 0.01$) in at least one round of the experiment are shown in the table. Additionally, variables representing which game was played are also omitted from the table for better readability.*

	Dependent variable:			
	H		C	
	First Round	Second Round	First Round	Second Round
extraversion	-0.003^{**} (0.001)	0.001 (0.001)	0.004^{**} (0.001)	0.001 (0.001)
Played the game before	0.019^{**} (0.007)	−0.006 (0.008)	−0.015 (0.009)	−0.012 (0.009)
Use Wikipedia frequently	0.022^{**} (0.007)	0.021^{*} (0.008)	−0.009 (0.008)	−0.012 (0.009)
Play computer games frequently	0.001 (0.005)	0.017^{*} (0.007)	−0.009 (0.006)	-0.022^{**} (0.008)
Left-handed	0.043^{***} (0.013)	-0.033^{*} (0.016)	-0.038^{*} (0.015)	0.015 (0.018)
Employed	0.002 (0.011)	0.019 (0.013)	−0.003 (0.013)	-0.040^{**} (0.015)
Age	-0.003^{***} (0.001)	−0.001 (0.001)	0.003^{***} (0.001)	0.001 (0.001)
Has time constraint	0.038^{***} (0.010)	0.024^{*} (0.012)	-0.061^{***} (0.011)	-0.078^{***} (0.013)
Constant	0.613^{***} (0.072)	0.305^{***} (0.087)	0.241^{**} (0.083)	0.471^{***} (0.098)
Observations	1,495	1,095	1,495	1,095
R^2	0.268	0.220	0.250	0.262
Adjusted R^2	0.245	0.186	0.227	0.230
Residual Std. Error	0.170 (df = 1449)	0.166 (df = 1049)	0.198 (df = 1449)	0.186 (df = 1049)
F Statistic	11.782^{***} (df = 45; 1449)	6.566^{***} (df = 45; 1049)	10.741^{***} (df = 45; 1449)	8.281^{***} (df = 45; 1049)

Note: *p<0.05; **p<0.01; ***p<0.001

5.4 Discussion

This research delved into the navigation strategies utilized by participants within Wikipedia's vast network. By employing a graph embedding technique trained on the English Wikipedia network to measure semantic distances between articles and using a local hierarchical score to determine article hierarchy, we observed that participants generally lean on geographical and occupational insights about their target to navigate the knowledge space, which correspond to hub-driven and proximity-driven strategies respectively within the context of the English Wikipedia network. The efficiency of these strategies is significantly influenced by the timing conditions imposed on the navigation tasks. Specifically, in the Least-clicks games, where there are no time restrictions, both strategies markedly improve performance. However, in the timed Speed-race games, although the hub-driven strategy continues to offer an advantage, the proximity-driven strategy becomes counterproductive. The distinction between occupational and geographical navigation strategies unearthed a "wisdom of the crowd" phenomenon, suggesting that collective navigation strategies effectively mirror the informational terrain surrounding the target, unbiased by individual differences.

In our study, we applied the concept of social navigation in the realm of information spaces, where the pathways taken by participants indicate their cognitive processes rather than their social connections. It was found that the categorization into occupational and geographical pathways, as previously identified in the context of navigating social networks [8, 23, 9], also applies to the navigation within information spaces. Notably, such a division is reflective of the informational environment of Wikipedia related to prominent figures, implying that the representation of individuals' geographical backgrounds and professions might be integral to how we construct our mental or cognitive maps of the social world. Indeed, earlier studies have shown that the hippocampus is adept at mapping abstract concepts, including a person's social affiliations and status [19], which aids in the quest for appropriate help in securing housing or jobs.

Previous studies on navigation on the information network [27] have explored the interplay of degree and closeness of network nodes influence an individual's wayfinding. Our research builds on this and demonstrates that this interplay do not only apply to the routes taken by single individuals but are also observable across different players. As highlighted in the Results section 5.3.2, this balancing act between node proximity and connectivity is inherent to the structure of the underlying knowledge network guiding our explorations. In environments where direct routes linking far-apart points are rare, adopting a strategy that favors nearby connections becomes logical. On the other hand, when the network features central nodes that link many points together, opting for a navigation strategy that leverages these hubs can offer distinct benefits.

In our experiment, the links that people rely on to navigate to the target are Wikipedia hyperlinks instead of social ties. Created for semantic reasons, the Wikipedia linking structure contains information from more aspects than social networks. It's not guaranteed that geography and occupation are still the major hierarchies that people utilize to navigate, there might be other hierarchies, popular culture mentions for example, that people could make use of to navigate. Our results show that when it comes to social search, geography and occupation seem to remain the major hierarchies used for navigation (looking at the last clicks to the target), even though the navigation space is much larger in our experiment. This implies that geography and occupation are probably the fundamental relationships people built with the world, which also explains when meeting anybody new in social life, the first few questions always include where are you from and what do you do. Looking at the navigation paths on the information network, the hub driven and proximity driven strategies resonates with an EVN navigation strategy [112] where a general network with homophily linking structure (e.g. geography and occupation) can be efficiently navigated considering a trade-off between node degree and node homophily to the target. Our results show that such trade-off indeed exists in empirical human navigation behavior.

Our research has certain limitations. Primarily, by focusing navigation tasks solely between renowned individuals, we may not fully capture variations in navigation behavior when the start and end points involve less prominent figures or non-personal entities like objects, happenings, or abstract concepts. Secondly, our experiment did not detect the anticipated influence of participants' pre-existing knowledge of the Wikipedia pages used in the navigation tasks, potentially because the self-reported survey responses failed to accurately reflect their actual knowledge levels. Future studies should aim to include more precise measurements of variables such as spatial navigation abilities and familiarity with the content of the source and target pages. Moreover, we only considered successful navigation attempts in our analysis, overlooking the insights that could be gained from examining unsuccessful ones. Investigating methods to analyze these failed attempts could provide valuable information for customizing help to enhance users' ability to navigate online information spaces.

Expanding on previous work on how individual differences affect navigation strategies in the knowledge network, it would be valuable to explore navigation tasks that are not restricted to well-known figures. This could encompass navigation involving lesser-known people or non-human subjects like objects, events, or abstract theories. Additionally, the exploration of algorithms to improve support for online navigation offers an exciting avenue for future research.

CHAPTER 6

CONCLUSION

This thesis examines the individual differences in performance and strategies when we navigate the knowledge network. For this purpose, I designed and conducted an online experiment where I implemented a social navigation task on the Wikipedia network in the form of a navigation game. We hired 802 participants online to play the game and collected their demographic information. Given the digital traces of the participants in the game and the structure of the English Wikipedia network, I investigated the individual differences in the participants' performance with respects to their success and creativity, classified the different navigation strategies the participants adopted and examined why they chose differently.

Our experiment, detailed in Chapter 3, provides a unique dataset for studying human navigation behavior in the knowledge space. By implementing a social navigation task in the knowledge space, we recorded participants' navigation trajectories, reflecting their thought processes rather than merely the individuals in their social networks as they navigate from person to person [8, 9]. Involving a large number of participants helped us overcome the typical issue of small sample sizes in laboratory experiments. Our survey includes not only basic demographic information but also the Big Five personality test and several other factors, offering rich insights about the participants from multiple perspectives, complementing the dataset produced from the Wikispeedia [25] or Wikigame [26] where demographic information was not accessible. By providing two versions of the game, we had the opportunity to study the effects of time and distance constraints on navigation behavior. Previous studies, as well as our own, demonstrate that timing conditions indeed moderate participants' navigation behavior [32].

Our study on participants' navigation performance, detailed in Chapter 4, underscores the impact of individual characteristics on navigation within the knowledge space, with this influence moderated by constraints such as time and distance. We found that prior experience with Wikipedia, the navigation game, and familiarity with the target page significantly predict better navigation, likely due to the game's nature. Controlling for these factors, being young and multilingual are consistent predictors of superior navigation performance, regardless of the constraints applied, highlighting the fundamental role of age and multilingualism in navigating the knowledge space. In terms of other traits, differences emerge with respect to the two types of constraints. Among participants engaging in Speed-race games, which involve time constraints, superior performance is shown by male participants of Asian ethnic background who do not speak a foreign language at a native level. In contrast, for participants in Least-clicks games, which have distance constraints, better performance is linked with those identifying as liberal and reporting superior spatial navigation skills. Interestingly, as participants engage in more rounds of the game, they exhibit significant improvement in Speed-race games but not in Least-clicks games. Employing a uniqueness measure introduced in this work, we demonstrated that individual traits not only affect navigation success but also route uniqueness: under time pressure, younger and left-handed participants tend to choose more unique routes, unlike in tasks with distance constraints.

In our study on participants' navigation strategies, detailed in Chapter 5, we discovered that participants typically utilize geographical and occupational information about the target person to navigate the knowledge space, corresponding to hub-driven and proximity-driven approaches on the English Wikipedia network, respectively. The success of these strategies is moderated by the timing conditions of the navigation tasks: in Least-clicks games, which lack time constraints, both hub-driven and proximity-driven approaches significantly improve performance. Conversely, in the timed Speed-race games, although the hub-driven strategy remains beneficial, the proximity-driven approach often results in poorer outcomes. The distinction between occupational and geographical navigation strategies unveils a "wisdom of the crowd" phenomenon, where the collective strategies effectively mirror the information landscape around the target, an insight unswayed by individual traits.

Various individual traits have been shown to influence navigation in physical spaces [30]. Our study broadens this research by exploring navigation within the knowledge space. Similar to navigation in physical spaces [63, 34], age acts as a limiting factor here, likely due to the declining cognitive abilities associated with aging, which affect fluid intelligence, perceptual speed, memory, and vocabulary [62]. Bilingualism has been identified as having numerous cognitive benefits, including enhanced executive control and protection against cognitive decline [97]. Our findings highlight speaking a foreign language as a significant predictor of success in navigating the knowledge space, suggesting an additional cognitive benefit of multilingualism. Creativity, traditionally associated with originality and effectiveness [104, 105], plays a role in navigation success and the uniqueness of chosen routes. This aligns with previous studies showing a preference for frequent detour use in navigation [106], with these detours reflecting individual traits [107]. However, traits that predict successful navigation to a target do not necessarily lead to innovative route-finding.

We observed that the division between occupational and geographical navigation paths identified in earlier work on social network navigation [8, 23, 9] also exists in information space navigation. This division interestingly reflects the information landscape on Wikipedia surrounding notable individuals, indicating that the representation of geographical origin and occupation may be fundamental to our mental or cognitive map of the social world. Indeed, previous research has shown that our hippocampus can represent abstract quantities, such as a person's affiliations and power within social encounters [19], aiding in the search for suitable assistance in areas like housing or employment. Earlier studies on wayfinding in the information network [27] explored the interplay between node degree and proximity within a single player's navigation trajectory. Our findings broaden this perspective by demonstrating that this interplay not only occurs within the navigation process of individual players but also on a macro level across different players. As discussed in Chapter 5, this trade-off is naturally dictated by the structure of the knowledge network on which our navigation is based. In an environment where shortcuts connecting distant locations are rare, a proximity-driven approach is naturally favored. On the other hand, when hubs exist that connect many locations, a hub-driven navigation strategy becomes more viable.

Our study advanced the understanding of individual differences in knowledge navigation. The next logical phase of research could involve developing navigation tasks where source and target pages extend beyond well-known individuals to encompass lesser-known figures or non-human concepts like objects, events, or theoretical ideas. It also suggests constructing mathematical models that incorporate personal traits to clarify participants' navigation behavior. Furthermore, investigating whether and how navigation experiences can be improved for individuals with specific characteristics in future experiments presents a promising avenue for exploration.

CHAPTER 7

APPENDICES

The appendix contains a full list of the questions and their encoded names and values in the survey of the experiment.

Agreeableness:

0–40 calculated from the 50 question Big Five personality test

Conscientiousness:

0–40 calculated from the 50 question Big Five personality test

Extroversion:

0–40 calculated from the 50 question Big Five personality test

Neuroticism:

0–40 calculated from the 50 question Big Five personality test

Openness:

0–40 calculated from the 50 question Big Five personality test

Prior_Wikigame: Which statement best describes your previous experience with the game we asked you to play?

0 I have never heard of the game and never played it before
1 I have heard of the game but never played it before
2 I have played the game (or similar game) several times before
2 I have played the game (or similar game) many times before
nan Other. Please specify:

Discarded[note1]: Are you a Wikipedia editor?

No
Yes. Please specify in what language do you edit and approximately how many edits have you made:

W_purpose: Choose the statement/s that describes how you use Wikipedia. You can choose more than one answer.

3 I am a fan of the idea of open knowledge and uses Wikipedia a lot to explore interesting topics
2 I use wikipedia a lot for my school work or job to learn new knowledge
1 I use wikipedia only for looking up information
nan Other. Please specify:

W_frequency: How often do you use Wikipedia?

0–4 Slide bar marked evenly by Never, Occasionally, Often, Everyday, Several Times Everyday

Discarded[note2]: What topics are you most familiar with on Wikipedia? Drag and drop the topics on the left to the corresponding groups on the right

Animals & Agriculture	Not Familiar
Arts & Design	Neutral
Food, Health & Medicine	Familiar
History & Geography	
Literature & Language	
Music & Films	
Philosophy & Religion	
Cultures & Traditions	
Math & Sciences	
Society & Events	
Sports & Games	
Technology	
Politics & Economy	

C_frequency: How often do you play computer games?

0–4 Slide bar marked evenly by Never, Occasionally, Often, Everyday, Several Hours Everyday

Discarded[note2]: What computer games do you play? Drag and drop the games on the left to the corresponding groups on the right

Strategy games	Never or Almost Never
Action games	Occasionally
Simulation and sports games	Very Often
Adventure and role play games	
Puzzle and educational games	
Other (Please specify)	

C_good: How good are you at playing computer games?

5 Excellent
4 Good
3 Average
2 Poor
1 Terrible
nan Can't say. I don't play computer games very often

C_like: How much do you like playing computer games?

4 A lot
3 A moderate amount
2 A little
1 None at all
nan Can't say. I don't play computer games very often

S_good: Compared to your family and friends, how well can you navigate to places in a new environment (e.g. a city where you have never been to before)?

5 Much better
4 Slightly better
3 About the same
2 Slightly worse
1 Much worse

S_learn: Do you agree that being able to navigate well in unknown environments is a skill that can be learned by practice?

```
    5    Strongly agree
    4    Slightly agree
    3    Neither agree nor disagree
    2    Slightly disagree
    1    Strongly disagree
```

S_unknown: When you try to navigate to a place in an unknown environment by foot, which statement best describes your behavior?

```
    0    I would get a paper map. I will look at the map constantly
on my way to the place
    1    I would get a paper map. I don't look at the map
constantly. I will plan the route, memorise it and not look at the
map again except when I am very uncertain at some point
    0    I would take the route planned for me on a map app. I would
constantly look at the phone for the route or turn by turn
instructions
    1    I would take the route planned for me on a map app.  I
don't look at the phone constantly. I will memorise the route and
not look at the phone again except when I am very uncertain at some
point
    1    I would use a map app on my smart phone. Once I know where
I am and where I am heading to, I plan the route myself instead of
taking the route planned for me on the app
    nan  Other, please specify:
```

S_known: When you try to navigate to a place in an familiar environment by foot, which statement best describes your behavior?

```
    1    I would rather explore than using a map
    0    I would use a map app on my smart phone. Once I know where
I am and where I am heading to, I plan the route myself instead of
taking the route planned for me on the app
    0    I would take the route planned for me on a map app. I would
constantly look at the phone for the route or turn by turn
instructions
    0    I would take the route planned for me on a map app.  I
don't look at the phone constantly. I will memorise the route and
not look at the phone again except when I am very uncertain at some
point
    nan  Other, please specify
```

S_left: Are you left-handed or right handed?

```
    0    Right-handed
    1    Left-handed
    nan  Other, please specify:
```

Gender: What is your gender?

```
    Female    Female
```

```
    Male      Male
    Other     Other, please specify:

Age: In which year were you born?(Please enter the year number, e.g.
1990) Encoded as age of the participant

Discarded[note1]: What is your country of birth?

Discarded[note3]: What is your nationality?

Discarded[note1]: In which country do you live now?

Discarded[note4]: For how many years have you lived in the current
country?

Ethnicity: What is your ethnicity?

    White       White
    Hispanic    Hispanic or Latino
    African     Black or African American
    Other       Native American or American Indian
    Asian       Asian
    Other       Pacific Islander
    nan         Other, please specify:

Native: Is English your first/native language?

    0    Yes, and it's my only native language
    1    Yes, and I have other native language(s). Please specify:
    1    No. Please specify your native language(s):

Foreign: Do you speak foreign language(s)? You can choose more than
one answer.

    1    I speak some foreign language(s) at advanced/working level.
Please specify:
    1    I speak some foreign language(s) at intermediate level.
Please specify:
    1    I speak some foreign language(s) at beginner level. Please
specify:
    0    I don't speak any foreign languages

Political: Where would you place yourself along the political
spectrum?

    Conservative    Conservative
    Moderate        Liberal
    Other           Other
    Other           I would rather not answer

ED_years: How many years of schooling did you get (starting from
primary school)?

    0-25 Entered by the participants. If more than 25 years, enocded
```

as 25 years.

Discarded[note4]: When was the last time you were enrolled in a school?

I am still a student
1–3 years ago
4–10 years ago
More than 10 years ago

ED_highest: What is the highest level of education you have completed?

0	Less than High School
0	High School / GED
1	Some College
1	2-year College Degree
1	4-year College Degree
2	Master's Degree
2	Professional Degree (JD, MD)
2	Doctoral Degree
nan	Other, please specify:

Discarded[note2]: If you attended or are attending college, what is your major? You can choose more than one answer.

Arts and Humanities
Business and Economics
Engineering and Computer
Math and Sciences
Social and Behavioral Sciences
Health and Medicine
Other major, please specify:
Can't say. I didn't attend college

EM_status: What is your employment status?

1	Employed for wages
1	Self-employed
0	A student
0	Out of work and looking for work
0	A homemaker
0	Unable to work
0	Out of work but not currently looking for work
0	Retired
0	Other

EM_mental: Is your current job mentally challenging? (If you don't have a job now, answer for your previous job)

0–4 Slide bar marked evenly by Not at all, A bit, Neutral, A moderate amount, Very much

EM_physical: Is your current job physically challenging? (If you don't have a job now, answer for your previous job)

0-4 Slide bar marked evenly by Not at all, A bit, Neutral, A moderate amount, Very much

EM_intensive: Is your current job intensive? (If you don't have a job now, answer for your previous job)

0-4 Slide bar marked evenly by Not at all, A bit, Neutral, A moderate amount, Very much

EM_creative: Is your current job creative? (If you don't have a job now, answer for your previous job)

0-4 Slide bar marked evenly by Not at all, A bit, Neutral, A moderate amount, Very much

Prior_Source: How much did you know about the source Wikipedia page before the game?

4 A great deal
3 A lot
2 A moderate amount
1 A little
0 None at all

Prior_Target: How much did you know about the target Wikipedia page before the game?

4 A great deal
3 A lot
2 A moderate amount
1 A little
0 None at all

note1: This question is discarded in our analysis because more than 90% of the participants gave the same answer.

note2: This question is discarded in our analysis because answers to this question disperse greatly.

note3: This question is discarded in our analysis because its answers overlap greatly with the question about ethnicity.

note4: This question is discarded in our analysis because its answers has a Pearson correlation coefficient greater than 0.5 with the age of the participants.

BIBLIOGRAPHY

[1] Zhu, M., Yasseri, T., Kertész, J.: Individual differences in knowledge network navigation. Scientific Reports **14**(1), 8331 (2024)

[2] Zhu, M., Kertész, J.: Milgram's experiment in the knowledge space: Individual navigation strategies (2024). 2404.06591

[3] Epstein, R.A., Patai, E.Z., Julian, J.B., Spiers, H.J.: The cognitive map in humans: spatial navigation and beyond. Nature neuroscience **20**(11), 1504–1513 (2017)

[4] Schafer, M., Schiller, D.: Navigating social space. Neuron **100**(2), 476–489 (2018)

[5] Pirolli, P., Card, S.: Information foraging. Psychological review **106**(4), 643 (1999)

[6] Bush, V., *et al.*: As we may think. The atlantic monthly **176**(1), 101–108 (1945)

[7] Karinthy, F.: Chain-links. Everything is different, 21–26 (1929)

[8] Milgram, S.: The small world problem. Psychology today **2**(1), 60–67 (1967)

[9] Dodds, P.S., Muhamad, R., Watts, D.J.: An experimental study of search in global social networks. science **301**(5634), 827–829 (2003)

[10] O'Keefe, J., Dostrovsky, J.: The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. Brain research (1971)

[11] O'Keefe, J.: Place units in the hippocampus of the freely moving rat. Experimental neurology **51**(1), 78–109 (1976)

[12] Fyhn, M., Molden, S., Witter, M.P., Moser, E.I., Moser, M.-B.: Spatial representation in the entorhinal cortex. Science **305**(5688), 1258–1264 (2004)

[13] Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., Moser, E.I.: Microstructure of a spatial map in the entorhinal cortex. Nature **436**(7052), 801–806 (2005)

[14] Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B.L., Witter, M.P., Moser, M.-B., Moser, E.I.: Conjunctive representation of position, direction, and velocity in entorhinal cortex. Science **312**(5774), 758–762 (2006)

[15] Stachenfeld, K.L., Botvinick, M.M., Gershman, S.J.: The hippocampus as a predictive map. Nature neuroscience **20**(11), 1643–1653 (2017)

[16] Baram, A.B., Muller, T.H., Whittington, J.C., Behrens, T.E.: Intuitive planning: global navigation through cognitive maps based on grid-like codes. bioRxiv, 421461 (2018)

[17] Muller, R.U., Stead, M., Pach, J.: The hippocampus as a cognitive graph. The Journal of general physiology **107**(6), 663–694 (1996)

[18] Peer, M., Brunec, I.K., Newcombe, N.S., Epstein, R.A.: Structuring knowledge with cognitive maps and cognitive graphs. Trends in cognitive sciences **25**(1), 37–54 (2021)

[19] Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y., Schiller, D.: A map for social navigation in the human brain. Neuron **87**(1), 231–243 (2015)

[20] Watts, D.J., Dodds, P.S., Newman, M.E.: Identity and search in social networks. science **296**(5571), 1302–1305 (2002)

[21] Kleinberg, J.: Small-world phenomena and the dynamics of information. Advances in neural information processing systems **14** (2001)

[22] Adamic, L., Adar, E.: How to search a social network. Social networks **27**(3), 187–203 (2005)

[23] Killworth, P.D., Bernard, H.R.: The reversal small-world experiment. Social networks **1**(2), 159–192 (1978)

[24] Wikipedia, the free encyclopedia. (2024). `https://en.wikipedia.org/` Accessed 2024

[25] Wikispeedia (2024). `https://dlab.epfl.ch/wikispeedia/play/` Accessed 2024

[26] The Wiki Game - Wikipedia Game - Explore Wikipedia! (2024). `https://www.thewikigame.com/` Accessed 2024

[27] West, R., Leskovec, J.: Human wayfinding in information networks. In: Proceedings of the 21st International Conference on World Wide Web, pp. 619–628 (2012)

[28] Singer, P., Helic, D., Taraghi, B., Strohmaier, M.: Detecting memory and structure in human navigation patterns using markov chain models of varying order. PloS one **9**(7), 102070 (2014)

[29] Helic, D., Strohmaier, M., Granitzer, M., Scherer, R.: Models of human navigation in information networks based on decentralized search. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media, pp. 89–98 (2013)

[30] Newcombe, N.S.: Individual variation in human navigation. Current Biology **28**(17), 1004–1008 (2018)

[31] Weisberg, S.M., Schinazi, V.R., Newcombe, N.S., Shipley, T.F., Epstein, R.A.: Variations in cognitive maps: understanding individual differences in navigation. Journal of Experimental Psychology: Learning, Memory, and Cognition **40**(3), 669 (2014)

[32] Nazareth, A., Huang, X., Voyer, D., Newcombe, N.: A meta-analysis of sex differences in human navigation skills. Psychonomic bulletin & review **26**(5), 1503–1528 (2019)

[33] Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, R., Wiener, J., Hölscher, C., Hornberger, M., et al.: Entropy of city street networks linked to future spatial navigation ability. Nature, 1–7 (2022)

[34] Spiers, H.J., Coutrot, A., Hornberger, M.: Explaining world-wide variation in navigation ability from millions of people: Citizen science project sea hero quest. Topics in Cognitive Science (2021)

[35] Dutton, W.H., Blank, G.: Cultures of the internet: Five clusters of attitudes and beliefs among users in britain (2014)

[36] Van Deursen, A.J., Van Dijk, J.A.: The digital divide shifts to differences in usage. New media & society **16**(3), 507–526 (2014)

[37] Ho, S.Y.: An exploratory study of using a user remote tracker to examine web users' personality traits. In: Proceedings of the 7th International Conference on Electronic Commerce, pp. 659–665 (2005)

[38] Savolainen, R.: Cognitive barriers to information seeking: A conceptual analysis. Journal of Information Science **41**(5), 613–623 (2015)

[39] Alós-Ferrer, C., García-Segarra, J., Ritschel, A.: Generous with individuals and selfish to the masses. Nature human behaviour **6**(1), 88–96 (2022)

[40] Landaeta, V., Candia, C., Pulgar, J., Fabrega, J., Varela, J., Yaikin, T., Monge, C., Rodriguez-Sickert, C.: Game of tokens: Low cooperative relationships as a key marker for bully-victims in elementary school classroom (2023)

[41] Allen, K.R., Brändle, F., Botvinick, M., Fan, J., Gershman, S.J., Griffiths, T.L., Hartshorne, J., Hauser, T.U., Ho, M.K., de Leeuw, J., et al.: Using games to understand the mind (2023)

[42] Tolman, E.C.: Cognitive maps in rats and men. Psychological review **55**(4), 189 (1948)

[43] Kuipers, B.: The" map in the head" metaphor. Environment and behavior **14**(2), 202–220 (1982)

[44] Kuipers, B.: Modeling spatial knowledge. Cognitive science **2**(2), 129–153 (1978)

[45] Kuipers, B., Tecuci, D.G., Stankiewicz, B.J.: The skeleton in the cognitive map: A computational and empirical exploration. Environment and behavior **35**(1), 81–106 (2003)

[46] Warren, W.H.: Non-euclidean navigation. Journal of Experimental Biology **222**(Suppl_1), 187971 (2019)

[47] Eichenbaum, H., Cohen, N.J.: From conditioning to conscious recollection: Memory systems of the brain. Oxford university press (2004)

[48] Wang, C.-H., Monaco, J.D., Knierim, J.J.: Hippocampal place cells encode local surface-texture boundaries. Current Biology **30**(8), 1397–1409 (2020)

[49] Bulkin, D.A., Sinclair, D.G., Law, L.M., Smith, D.M.: Hippocampal state transitions at the boundaries between trial epochs. Hippocampus **30**(6), 582–595 (2020)

[50] Sun, C., Yang, W., Martin, J., Tonegawa, S.: Hippocampal neurons represent events as transferable units of experience. Nature neuroscience **23**(5), 651–663 (2020)

[51] Van Der Meer, M., Kurth-Nelson, Z., Redish, A.D.: Information processing in decision-making systems. The Neuroscientist **18**(4), 342–359 (2012)

[52] Hassabis, D., Kumaran, D., Vann, S.D., Maguire, E.A.: Patients with hippocampal amnesia cannot imagine new experiences. Proceedings of the National Academy of Sciences **104**(5), 1726–1731 (2007)

[53] Ólafsdóttir, H.F., Barry, C., Saleem, A.B., Hassabis, D., Spiers, H.J.: Hippocampal place cells construct reward related sequences through unexplored space. Elife **4**, 06063 (2015)

[54] Bellmund, J.L., Gärdenfors, P., Moser, E.I., Doeller, C.F.: Navigating cognition: Spatial codes for human thinking. Science **362**(6415), 6766 (2018)

[55] Behrens, T.E., Muller, T.H., Whittington, J.C., Mark, S., Baram, A.B., Stachenfeld, K.L., Kurth-Nelson, Z.: What is a cognitive map? organizing knowledge for flexible behavior. Neuron **100**(2), 490–509 (2018)

[56] West, R., Leskovec, J.: Automatic versus human navigation in information networks. Proceedings of the International AAAI Conference on Web and Social Media (2012)

[57] Scaria, A.T., Philip, R.M., West, R., Leskovec, J.: The last click: why users give up information network navigation. Proceedings of the 7th ACM international conference on Web search and data mining (2014)

[58] West, R., Paranjape, A., Leskovec, J.: Mining missing hyperlinks from human navigation traces: A case study of wikipedia. Proceedings of the 24th International Conference on World Wide Web (2015)

[59] Lamprecht, D., Lerman, K., Helic, D., Strohmaier, M.: How the structure of wikipedia articles influences user navigation. The New Review of Hypermedia and Multimedia **23**, 29–50 (2016)

[60] Trattner, C., Helic, D., Singer, P., Strohmaier, M.: Exploring differences and similarities between hierarchical decentralized search and human navigation in information networks. (2012)

[61] Lamprecht, D., Strohmaier, M., Helic, D., Nyulas, C., Tudorache, T., Noy, N., Musen, M.A.: Using ontologies to model human navigation behavior in information networks: A study based on wikipedia. Semantic web **6 4**, 403–422 (2015)

[62] Ghisletta, P., Rabbitt, P., Lunn, M., Lindenberger, U.: Two thirds of the age-based changes in fluid and crystallized intelligence, perceptual speed, and memory in adulthood are shared. Intelligence **40**(3), 260–268 (2012)

[63] Anguera, J.A., Boccanfuso, J., Rintoul, J.L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., *et al.*: Video game training enhances cognitive control in older adults. Nature **501**(7465), 97–101 (2013)

[64] Lindenberger, U.: Human cognitive aging: corriger la fortune? Science **346**(6209), 572–578 (2014)

[65] Dabbs Jr, J.M., Chang, E.-L., Strong, R.A., Milun, R.: Spatial ability, navigation strategy, and geographic knowledge among men and women. Evolution and human behavior **19**(2), 89–98 (1998)

[66] Lawton, C.A.: Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. Sex roles **30**, 765–779 (1994)

[67] Lawton, C.A., Kallai, J.: Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. Sex roles **47**, 389–401 (2002)

[68] Sharit, J., Hernández, M.A., Czaja, S.J., Pirolli, P.: Investigating the roles of knowledge and cognitive abilities in older adult information seeking on the web. ACM Transactions on Computer-Human Interaction (TOCHI) **15**(1), 1–25 (2008)

[69] Chevalier, A., Dommes, A., Marquié, J.-C.: Strategy and accuracy during information search on the web: Effects of age and complexity of the search questions. Computers in human behavior **53**, 305–315 (2015)

[70] McDonald, S., Spencer, L.: Gender differences in web navigation: Strategies, efficiency & confidence. Springer, 174–181 (2000)

[71] Liao, Y., Finn, M., Lu, J.: Information-seeking behavior of international graduate students vs. american graduate students: A user study at virginia tech 2005. College & Research Libraries **68**(1), 5–25 (2007)

[72] Zmigrod, L., Eisenberg, I.W., Bissett, P.G., Robbins, T.W., Poldrack, R.A.: The cognitive and perceptual correlates of ideological attitudes: a data-driven approach. Philosophical Transactions of the Royal Society B **376**(1822), 20200424 (2021)

[73] Marian, V., Shook, A.: The cognitive benefits of being bilingual. In: Cerebrum: the Dana Forum on Brain Science, vol. 2012 (2012). Dana Foundation

[74] Wikipedia:Policies and guidelines. (2024). https://en.wikipedia.org/w/index.php?title=Wikipedia:Policies_and_guidelines Accessed 2024

[75] Prolific — Quickly find research participants you can trust (2024). https://www.prolific.com/ Accessed 2024

[76] Douglas, B.D., Ewell, P.J., Brauer, M.: Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona. Plos one **18**(3), 0279720 (2023)

[77] Qualtrics XM - Experience Management Software (2024). https://www.qualtrics.com/uk/ Accessed 2024

[78] Wikimedia Downloads (2024). https://dumps.wikimedia.org/ Accessed 2024

[79] Official White House Photo by Pete Souza, Public domain, via Wikimedia Commons (2024). https://commons.wikimedia.org/wiki/File:President_Barack_Obama.jpg Accessed 2024

[80] President.gov.ua, CC BY 4.0, via Wikimedia Commons (2024). https://commons.wikimedia.org/wiki/File:Emmanuel_Macron_June_2022_(cropped).jpg Accessed 2024

[81] Original: Unknown, Vector: Jürgen Krause, Public domain, via Wikimedia Commons (2024). https://commons.wikimedia.org/wiki/File:Flag_of_France_official.svg Accessed 2024

[82] Vincent van Gogh, Public domain, via Wikimedia Commons (2024). https://commons.wikimedia.org/wiki/File:Vincent_van_Gogh_-_Self-Portrait_-_Google_Art_Project_(454045).jpg Accessed 2024

[83] Kkdu101, CC BY-SA 3.0, via Wikimedia Commons (2024). https://commons.wikimedia.org/wiki/File:NU_Diploma_(redacted).jpg Accessed 2024

[84] User:Husky and h3m3ls, Mischa de Muynck and Niels, CC BY-SA 3.0, via Wikimedia Commons (2024). https://commons.wikimedia.org/wiki/File:Art-portrait-collage_2.jpg Accessed 2024

[85] Goldberg, L.R.: The development of markers for the big-five factor structure. Psychological assessment **4**(1), 26 (1992)

[86] Jones, G., Macken, B.: Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. Cognition **144**, 1–13 (2015)

[87] Gill, P., Crete-Nishihata, M., Dalek, J., Goldberg, S., Senft, A., Wiseman, G.: Characterizing web censorship worldwide: Another look at the opennet initiative data. ACM Transactions on the Web (TWEB) **9**(1), 1–29 (2015)

[88] Mat-Hassan, M., Levene, M.: Associating search and navigation behavior through log analysis. Journal of the American Society for Information Science and Technology **56**(9), 913–934 (2005)

[89] Dunteman, G.H.: Principal components analysis. Sage (69) (1989)

[90] Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710 (2014)

[91] Witten, I.H., Milne, D.N.: An effective, low-cost measure of semantic relatedness obtained from wikipedia links (2008)

[92] Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., Ruppin, E.: Placing search in context: The concept revisited. In: Proceedings of the 10th International Conference on World Wide Web, pp. 406–414 (2001)

[93] Singer, P., Niebler, T., Strohmaier, M., Hotho, A.: Computing semantic relatedness from human navigational paths: A case study on wikipedia. International Journal on Semantic Web and Information Systems (IJSWIS) **9**(4), 41–70 (2013)

[94] Besse, P., Guillouet, B., Loubes, J.-M., François, R.: Review and perspective for distance based trajectory clustering. arXiv preprint arXiv:1508.04904 (2015)

[95] Rockafellar, R.T., Wets, R.J.-B.: Variational analysis. Springer science & business media **317** (2009)

[96] Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)

[97] Bialystok, E., Craik, F.I., Luk, G.: Bilingualism: consequences for mind and brain. Trends in cognitive sciences **16**(4), 240–250 (2012)

[98] Lövdén, M., Herlitz, A., Schellenbach, M., GROSSMAN-HUTTER, B., Krüger, A., Lindenberger, U.: Quantitative and qualitative sex differences in spatial navigation. Scandinavian journal of psychology **48**(5), 353–358 (2007)

[99] Padilla, L.M., Creem-Regehr, S.H., Stefanucci, J.K., Cashdan, E.A.: Sex differences in virtual navigation influenced by scale and navigation experience. Psychonomic bulletin & review **24**(2), 582–590 (2017)

[100] Grön, G., Wunderlich, A.P., Spitzer, M., Tomczak, R., Riepe, M.W.: Brain activation during human navigation: gender-different neural networks as substrate of performance. Nature neuroscience **3**(4), 404–408 (2000)

[101] Saucier, D.M., Green, S.M., Leason, J., MacFadden, A., Bell, S., Elias, L.J.: Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies? Behavioral neuroscience **116**(3), 403 (2002)

[102] Huang, X., Voyer, D.: Timing and sex effects on the "spatial orientation test": A world war ii map reading test. Spatial Cognition & Computation **17**(4), 251–272 (2017)

[103] Cobb-Clark, D.A., Schurer, S.: The stability of big-five personality traits. Economics Letters **115**(1), 11–15 (2012)

[104] Runco, M.A., Jaeger, G.J.: The standard definition of creativity. Creativity research journal **24**(1), 92–96 (2012)

[105] Campbell, C.M., Izquierdo, E.J., Goldstone, R.L.: Partial copying and the role of diversity in social learning performance. Collective Intelligence **1**(1), 26339137221081849 (2022)

[106] Spiers, H.J., Gilbert, S.J.: Solving the detour problem in navigation: a model of prefrontal and hippocampal interactions. Frontiers in human neuroscience **9**, 125 (2015)

[107] Gulyás, A., Bíró, J., Rétvári, G., Novák, M., Kőrösi, A., Slíz, M., Heszberger, Z.: The role of detours in individual human navigation patterns of complex networks. Scientific Reports **10**(1), 1098 (2020)

[108] Hunsaker, A., Hargittai, E.: A review of internet use among older adults. New media & society **20**(10), 3937–3954 (2018)

[109] MacQueen, J., *et al.*: Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281–297 (1967). Oakland, CA, USA

[110] Jolliffe, I.T.: Principal component analysis for special types of data. Springer (2002)

[111] Muchnik, L., Itzhack, R., Solomon, S., Louzoun, Y.: Self-emergence of knowledge trees: Extraction of the wikipedia hierarchies. Physical review E **76**(1), 016106 (2007)

[112] Şimşek, Ö., Jensen, D.: Navigating networks by using homophily and degree. Proceedings of the National Academy of Sciences **105**(35), 12758–12762 (2008)