Navigational Decision-Making Under Uncertainty

By

Phillip Michael Newman

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

August 11, 2023

Nashville, Tennessee

Approved:

Timothy P. McNamara, Ph.D.

Isabel Gauthier, Ph.D.

Randolph Blake, Ph.D.

Robert Bodenheimer, Ph.D.

Jennifer Trueblood, Ph.D.

TABLE OF CONTENTS

Chapter Page

CHAPTER 1

INTRODUCTION

Navigating through the environment is a ubiquitous skill for many organisms, especially humans. Every day, people around the world travel from some starting location to a desired or target location, a process referred to as wayfinding. Although this task is seemingly simple and comes naturally to most, effective wayfinding is a complex cognitive task. For example, when planning a route to a target location, a person might choose from multiple overlapping routes, considering the total distance required or the overall complexity (e.g., number of turns and decision points) of each route. In other cases, the navigator might need to wayfind to a sequence of locations (e.g., when running errands), paying special consideration to the amount of time it will take to navigate to each location and the uncertainty about how much time will be required to spend there. Consequences of wayfinding errors range widely, from arriving late to a dinner party to running out of fuel while flying over an ocean. Thus, it is not surprising that wayfinding ability and the underlying cognitive processes have received substantial attention in the spatial cognition literature (Boone, Maghen, & Hegarty, 2019; Burgess, 2008; Chrastil, 2013; Chrastil & Warren, 2012, 2014; He, McNamara, Bodenheimer, & Klippel, 2018; He, McNamara, & Brown, 2019; Hegarty, Montello, Richardson, Ishikawa, Lovelace, 2006; Ishikawa & Montello, 2006; Montello, 1998; Münzer, Fehringer, & Kühl, 2016; Shelton & Gabrieli, 2004; Siegel & White, 1975; Thorndyke & Hayes-Roth, 1982; Weisberg & Ekstrom, 2021; Weisberg, Schinazi, Newcombe, Shipley, & Epstein, 2014; Yu, Boone, He, Davis, Hegarty, Chrastil, & Jacobs, 2021, to name a few).

Despite the vast attention given to wayfinding behavior, little has been done to investigate how individuals estimate the amount of time required to visit multiple goal locations under uncertainty. However, this problem is ubiquitous to most individuals. Navigators often need to estimate how many locations can be visited within a fixed time interval. For example, if a person wishes to run errands before a job interview in two hours, they will need to consider how many errands they can run without arriving late to the interview. Consideration must be given to the uncertainty about the time to navigate to each location and the time to complete the tasks at each location. In this example, the timeframe is fixed and the number of errands to be run is contingent on the given timeframe. On the other hand, the navigator may have a fixed number of errands to run and need to estimate the timeframe required to visit them to plan the rest of their day. In this case, underestimating the required timeframe may result in arriving late to other obligations, while overestimation may result in having unnecessary free time to kill. In this dissertation, I propose a novel paradigm for investigating navigational decision-making under uncertainty. First, I will discuss the relevant literature on spatial knowledge acquisition and the contribution of individual differences. Then, I will propose a novel task in which navigators are required to estimate the amount of time required to run a set of errands under uncertainty, with estimation errors resulting in a cost. Finally, I will assess the reliability and validity of the novel errand running task.

Acquiring Spatial Knowledge from Wayfinding

As navigators are exposed to a new environment, they begin to develop knowledge of the environment that can be accessed and modified during future wayfinding (e.g., He et al., 2019; Siegel & White, 1975). Siegel and White (1975) put forth one of the earliest theoretical accounts of how spatial knowledge is acquired. According to their model, spatial knowledge comprises three primary components: landmarks, route knowledge, and survey knowledge. Landmarks are prominent places in an environment that stand out due to their distinct visual features and serve as useful reference points for orienting. Routes connect sets of locations by sequences of paths and turns. Originally, route knowledge was believed to contain little to no metric information and was composed purely of associations between navigational decisions and visual scenes (e.g., follow straight down the path, then turn left at the large oak tree). However, recent work has challenged this view, showing that when the shortest route to a target destination is blocked, navigators will often choose the next shortest route (Chrastil & Warren, 2015). Knowledge of multiple overlapping routes comprise complex networks of connections (edges) between sets of locations (nodes) and these networks constitute graph knowledge, which is conceptualized as an intermediate step between route and survey knowledge and contains metric information (Chrastil, 2013; Meilinger, 2008; Muller, 1996). Finally, survey knowledge is represented by straight-line distances and directions between locations and is conceptualized as constituting a *cognitive map* (O'Keefe & Nadel, 1978).

Although Siegel and White (1975) originally proposed that the components of spatial knowledge were acquired sequentially, a large body of literature has rendered this view outdated (see Chrastil, 2013, for an alternative framework). It is now widely accepted that route and survey knowledge are acquired simultaneously. However, the relative rates at which different types of spatial knowledge are developed depends in part on the medium through which the navigator learns the environment (Giraudo & Pailhous, 1994; Montello, Waller, Hegarty, Richardson, 2004; He et al., 2019; Lloyd, 1989; Moeser, 1988; Richardson, Montello, & Hegarty, 1999; Rossano, West, Robertson, Wayne, & Chase, 1999; Taylor & Tversky, 1996). A classic demonstration of this comes from Thorndyke and Hayes-Roth (1982), who tested two groups of participants on their spatial knowledge of a floor plan of a large corporate building. The two groups of participants comprised employees who worked at the building (navigation learners) and subjects who had never been inside the building before, but instead studied a map of the floor plan (map learners). Although both groups of participants displayed some route and survey knowledge, map learners tended to outperform navigation learners when estimating direct angles and straight-line distances to target locations, whereas navigation learners outperformed map learners when estimating route distance and orientation. Their results suggest that development of survey knowledge was supported by learning environments from survey representations (e.g., maps) while development of route knowledge was supported by navigating through the environment. However, it should be noted that navigation learners' survey knowledge improved with increased exposure, and in some cases, caught up to map learners.

A more recent example comes from He et al. (2019), who investigated how the spatial information navigators were exposed to when learning an environment affected the relative acquisition of route and survey knowledge. Two groups of participants learned a virtual shopping mall and were either exposed predominantly to route or survey information during learning. Survey exposure occurred through completing penetrable trials, wherein the walls of the buildings were penetrable by the participants so they could travel in straight lines to the target location.[1] Route exposure occurred through completing impenetrable trials, wherein the walls of the buildings were impenetrable, requiring traditional route-based wayfinding. Participants (using mouse and keyboard) who primarily received survey exposure showed gradual improvements during wayfinding on penetrable trials, but not impenetrable trials. On the other

---

[1] While inside of a building, its walls were rendered invisible such that the participant was able to view his or her surroundings from inside of the building.

hand, participants who primarily received route exposure showed the opposite pattern, suggesting that exposure to different types of spatial representations (route versus survey) differentially impacted acquisition of spatial knowledge components. This effect was attenuated when participants had access to body-based information (i.e., cues generated from the body during self-motion, such as proprioception and motor efference copies), suggesting that body-based cues to navigation might support both types of spatial knowledge development (see also Chrastil & Warren, 2012; Chrastil & Warren, 2013).

Other studies contest whether survey knowledge (i.e., a cognitive map) is ever truly developed, with some authors suggesting that knowledge of the environment is better conceptualized as a labeled graph (Chrastil & Warren, 2013; Chrastil & Warren, 2014; see also Peer et al., 2021, for a review). After learning a virtual maze, Chrastil and Warren (2013) had participants take straight-line shortcuts from a starting location to a target location. The maze was only available as an orientation cue before beginning the shortcut, at which time the maze was replaced by an infinite ground plane. Although the presence of body-based cues reduced angular error overall, mean angular error was high (roughly 70°), with half of participants near chance (90°). Only 12% of participants had absolute angular errors around 20°. However, during a route-finding task in which the shortest possible route to the target was blocked off, nearly 40% of participants were able to find the next shortest route, and 60% were at least successful at finding the target (Chrastil & Warren, 2014). Only 10% of participants were near chance performance. The ability of participants to take the next shortest route during the route-finding task suggests that many participants formed a labeled graph representation, as opposed to a purely topological representation. In a labeled graph, metric information such as path lengths and angles are encoded in the edges between nodes, whereas edges simply denote the presence (or absence) of a connection between nodes in a topological graph.

The studies reviewed thus far suggest that, although the components of spatial knowledge are acquired simultaneously, the mode by which navigators are exposed to an environment affects the relative rate of acquisition of each component. Furthermore, the type of spatial knowledge (e.g., survey/graph or route knowledge) possessed by the navigator can influence how navigation manifests (e.g., taking novel shortcuts). Thus, its plausible that the development of different types of spatial knowledge may enhance the accuracy of timeframe predictions when navigating under uncertainty, such as running errands. Importantly, when assessing the spatial knowledge of navigators, the distinction between cognitive maps (survey representations) and cognitive graphs should be considered. The former can be measured via pointing and straight-line estimation tasks (e.g., He et al., 2019), and the latter via route-finding (e.g., Chrastil & Warren, 2014, 2015). However, although many of Chrastil and Warren's (2014) participants were able to take the next shortest route to a target after the shortest was blocked, and many others were otherwise successful at locating the target, there were still many participants who failed, suggesting that there is widely ranging variability in graph knowledge development among navigators. In the next section, I consider the role of individual differences when making navigational decisions under uncertainty.

Individual differences in Spatial Knowledge Acquisition

Numerous studies have demonstrated vast individual differences in spatial knowledge and behavior, even in children as young as 20-26 months (Hazen, 1982). Perhaps the most well-known study of individual differences in the acquisition of route and survey knowledge was conducted by Ishikawa and Montello (2006). Participants were driven along two separate routes which were mostly straight (S-route) and U-shaped (U-route), respectively. While driving along each route, the experimenter pointed out four landmarks, each of which could not be seen from any other landmark's location. Following each route, participants completed distance and direction estimates between the landmarks on that route. Participants made these same estimates between landmarks from separate routes which served as a baseline for between-route estimates. During additional sessions, participants were driven along connecting routes before making between-route distance and direction estimates. Although little improvement was observed across sessions, there were marked individual differences in performance. Whereas a small group of participants showed high levels of accuracy throughout the duration of the study and another small group showed poor accuracy, a large group of participants performed intermediately. Performance, especially for landmarks along the complex U-route, also correlated with self-reported sense of direction, as measured by the Santa Barbara Sense of Direction Scale (SBSOD), which has been psychometrically validated several times and serves as a measure of large-scale environmental ability (e.g., Hegarty et al., 2006; Hegarty, Richardson, Montello, Lovelace, & Subbiah, 2002).

Less has been done to investigate individual differences in graph knowledge acquisition. Although navigators tend to develop more robust graph than survey representations, a wide range of performance has been observed on route-finding tasks probing the use of graph knowledge (Chrastil & Warren, 2014, 2015). Other studies have demonstrated large individual differences in integrating connecting routes between two previously learned routes, an ability that is crucial to developing graph knowledge (Weisberg et al., 2014; Weisberg & Newcombe, 2018). For example, Weisberg et al. (2014) had participants learn a virtual town by traveling along two main routes and two connecting routes. Participants were separated into three groups based on their pointing errors for locations within- and between-routes: good between/good within, bad between/good within, and bad between/bad within.[2] In addition to overall negative associations of mental rotation and sense of direction with within- and between-route pointing errors, participants in the good between/good within group showed significantly higher scores on both psychometric tests. However, it should be noted that environmental knowledge was tested by a pointing task, which probes survey knowledge and as noted previously, tends to yield worse performance than route-finding tasks (cf. Chrastil & Warren, 2014).

Sex differences in spatial knowledge acquisition and wayfinding strategy use have also been well documented (e.g., Boone, Gong, & Hegarty, 2018; Chrastil & Warren, 2015; Coluccia & Louse, 2004; Coutrot et al., 2022; Yu et al., 2021; Mofat, Hampson, & Hatzipantelis, 1998). For example, on average, men have been shown to develop more accurate survey representations, rely more on cardinal directions during wayfinding, and tend to take more shortcuts than women, while women tend to follow known routes and rely on associations between landmarks and wayfinding decisions (e.g., turn left at the library) (Boone et al., 2018). A recent demonstration of this comes from Yu et al. (2021), who examined sex and age-related differences in spatial knowledge and wayfinding strategy. Young men were more successful at finding target locations within the time limit and used more shortcuts than young women. However, this difference disappeared for older men and women (cf. Coutrot et al., 2018).

The link between spatial reasoning ability and spatial knowledge acquisition has been demonstrated using structural equation modeling techniques (for a review of individual differences in psychometric tests of spatial reasoning, see Hegarty & Waller, 2005). Hegarty et al. (2006) showed that small-scale (e.g., mental rotation) and large-scale (e.g., sense of direction)

---

[2] Only a single participant demonstrated good between- but bad within-route pointing errors.

spatial reasoning abilities predicted learning of an environmental layout from both direct experience and media. However, they differed in their relative ability to predict learning from each method of exposure. Specifically, small-scale spatial reasoning ability was a stronger predictor of learning from media, while large-scale spatial reasoning ability was a stronger predictor of learning from direct experience. A person's ability to make navigational decisions under uncertainty is inextricably linked to their knowledge of the environment they are making decisions about. If individuals' ability to develop knowledge (cognitive maps or graphs) of the environment covaries with their small- and large-scale spatial abilities, it is plausible that such abilities are also predictive of the decisions navigators make when navigating under uncertainty (e.g., running errands).

Currently, the role of individual differences in navigational decision making under uncertainty has not been investigated to my knowledge. Therefore, as part of this dissertation, I will test whether navigators' decision-making and performance under uncertainty while running errands can be predicted by individual differences in small- and large-scale spatial reasoning abilities, wayfinding anxiety, and spatial knowledge of the environment. Given the substantial evidence that individuals vary widely in their ability to acquire spatial knowledge, it is reasonable to speculate that navigators with greater spatial ability (and hence more comprehensive cognitive maps/graphs after learning) will be more successful at estimating the amount of time required to run a fixed number of errands.

Decision-Making Under Uncertainty

Navigation requires complex cognitive processing and involves making decisions under uncertainty. For example, when running errands, navigators must consider factors such as the time required to complete the tasks at each errand destination, in addition to how long it will take to travel there. Some of this uncertainty is, in principle, knowable in advance. For example, the time it will take to travel from one location to another is influenced by the amount of knowledge a navigator has about the environment, such as knowledge of various routes and their path lengths (e.g., labeled graph knowledge). Other aspects of uncertainty are unknowable in advance, stemming from random or stochastic processes. For example, navigators might be slowed down by traffic, or there could be long lines at some of the errand destinations (e.g., when checking out at the grocery store). These two dimensions of uncertainty are referred to as epistemic and aleatory uncertainty and have recently gained a fair amount of attention in the decision-making literature (Fox & Ülkümen, 2011; Krijnen, Ülkümen, Bogard, & Fox, 2022; Tannenbaum, Fox, & Ülkümen, 2017; Ülkümen, Fox, & Malle, 2016; Walters, Ülkümen, & Tannenbaum, 2022).

Epistemic uncertainty stems from an absence of knowledge about a fact that is either true or false. For example, a navigator may be uncertain about whether a store opens at 7 am, but this uncertainty could be eliminated by looking up the store hours. Aleatory uncertainty is uncertainty about the results of probabilistic or stochastic processes. For example, when rolling a fair die, it is impossible to know what the result will be. Individuals vary in their perceptions about whether events are epistemic or aleatory in nature, and these perceptions are biased by factors such as linguistic framing (Ülkümen, Fox, & Malle, 2016), probabilistic extremity (Tannenbaum, Fox, & Ülkümen, 2017), and representation (single case vs. class of possible outcomes; Fox & Ülkümen, 2011). Fox and colleagues developed the Epistemic-Aleatory Rating Scale (EARS), which has been used to predict financial decision-making and classification of linguistic framings of probability judgments (confidence vs. likelihood statements). As mentioned, navigational decision-making likely involves both types of uncertainty. For example, estimating the amount of time required to complete a set of errands could be influenced by the navigator's knowledge and familiarity with the environment (epistemic) and consideration of random variables such as traffic and long lines at errand destinations (aleatory).

In addition to uncertainty, decision-making researchers are also concerned with measuring individuals' predisposition towards risky decision-making and behavior (Lauriola, Levin, & Hart, 2007). For example, during the Balloon Analogue Risk Task (BART) (Lejuez et al., 2002), participants are presented with a balloon and two choices: pump or cash out. Every time the balloon is pumped, 5 cents is added to the potential reward for that trial. If the participant chooses to cash out, the potential reward for that trial is added to their total reward. However, each time the balloon is pumped, the probability that it will pop on subsequent pumps increases. If the balloon pops, the potential reward for that trial is forfeited. The trial ends only if the participant cashes out or the balloon pops. The average number of pumps on trials in which the participant cashes out before popping the balloon is often used as an indication of that participant's predisposition toward risky decision-making. The BART has been shown to correlate with self-reported risk behaviors such as smoking, alcohol use, gambling, and risky sexual behavior (Lejuez et al., 2002).

However, decision-making researchers are unclear about what kind of risky decision-making the BART is measuring. Using exploratory and confirmatory factor analysis, Buelow and Blaine (2015) recently found evidence that the BART and other prominent risky decision-making tasks, such as the Iowa Gambling Task and Columbia Card Task, measure different constructs. One possibility is that propensities toward risky behaviors differ across contexts and may need to be measured separately. Weber and colleagues (2002, 2006) developed the

Domain-Specific Risk-Taking Scale (DOSPERT), a self-report survey which measures risk behaviors and perceptions across five domains: financial, health/safety, recreational, ethical, and social. Although it is possible to compute a comprehensive score, the domain subscales of the DOSPERT have shown high discriminant validity from one another, loading onto distinct factors, further evidencing that risk propensity may be contextual.

Errand running involves elements of uncertainty and risk. Lack of spatial knowledge about the environment and common impediments (e.g., traffic, long lines) create uncertainty about the time needed to run a set of errands. Under and overestimating the amount of time required to run a set of errands impose costs (i.e., late arrival and extra time to fill, respectively). As part of the current project, both the BART and DOSPERT will be administered to explore the correlates of timeframe estimation when running errands. Furthermore, the role of perceptions about uncertainty will also be explored by having navigators rate the epistemic and aleatory nature of the errand running task. For example, navigators who perceive errand running to involve more aleatory uncertainty might overestimate the amount of time required to complete a set of errands. Furthermore, individuals with greater aversion to risk might also overestimate the required time.

Current Project

Wayfinding tasks used in research on spatial navigation are often very similar. Typically, participants first learn an environment via a predetermined method of exposure (e.g., following a predefined route, free exploration, viewing a map, etc.). Then, participants are taken to a starting location and are required to find their way to a target location, usually within a timeframe specified by the researcher. However, in a naturalistic setting it is often the case that navigators must decide for themselves how much time is required to visit one or multiple locations or how many locations they can visit within a given timeframe. Consider the following scenario: You are making plans to meet a friend and need to determine when you will arrive at your agreed upon meeting location. However, you also need to run a few errands before your meeting and would like to arrange a meeting time that allows you to complete your errands. You will need to consider the uncertainty involved in the errands to minimize costs of misestimating the required amount of time to complete your errands. Despite being a quotidian problem, surprisingly little has been done to investigate how individuals make navigational decisions under uncertainty.

Errand running tasks are not entirely novel, however. Researchers have used the Multiple Errands Task (MET) to study clinical populations with frontal lobe damage as an ecologically valid and alternative test of executive functioning (Burgess, Alderman, Evans, & Wilson, 1998; Burgess et al., 2006; Dawson et al., 2009; Knight, Alderman, & Burgess, 2002; Shallice & Burgess, 1991). Among several other tasks, participants are required to visit several locations (usually in a hospital setting where participants are currently receiving treatment) to gather various items, and then meet the experimenter at a predesignated location within a given timeframe. However, these tasks are rarely, if ever, used to study spatial navigation per se, but rather impairments in performance (e.g., failure to complete various tasks, number of rule violations, task efficiency, etc.) relative to healthy controls. Additionally, the paradigm almost always involves a fixed set of errands to be run and a fixed timeframe in which to complete them, and thus decision-making is seldom examined closely.

Several concerns have been raised regarding the MET’s validity and reliability. For example, in a systematic review, Rotenberg et al. (2020) found that the internal consistency was indeterminant and test-retest reliability to be poor. A likely issue is that the MET must be adapted to the physical setting in which it is being implemented (e.g., Burns, Dawson, Perea, &

Vas, 2019), introducing task variability across studies (see McGeorge et al., 2001, and Rand, Weiss, & Katz, 2009, for virtual implementations). Additionally, the MET is made up of several distinct subtasks, none of which has been validated in isolation.

Important to task design is demonstrating its internal consistency, defined as the degree to which a set of items in an instrument all measure the same trait, and its ability to reliably measure the same trait across time (i.e., test-retest) (Magasi, Gohil, Burghart, & Wallisch, 2017). Therefore, the first goal of the proposed project is to establish the internal consistency and reliability of a virtual errand running task in which participants are required to make decisions about the time required to perform the task. Secondly, I investigated the degree to which various psychometric measures of spatial ability and decision-making are predictive of errand running performance, including, for example, decisions about the required timeframe to complete the errands, success rates, and time to complete the errands. An overview of experiments designed to meet the above goals is presented below.

First, the internal consistency of the errand running task was examined via two pilot experiments. To this aim, participants completed a learning phase followed by the errand running task in a virtual shopping mall (Figure 1) across two within-subjects conditions that varied the amount of uncertainty. Participants were informed of the storefronts that they would need to visit at the beginning of the trial, at which time they selected from seven options how long they thought it would take to complete the errands. During narrow and wide trials, the target storefronts on the errand list were associated with narrowly and widely ranging waiting times, respectively, to manipulate the amount of uncertainty when selecting a timeframe. Participants were told that each timeframe was associated with a monetary payoff to be added to their total earnings for the study if they successfully completed all the errands and returned to the starting location within the selected timeframe for that trial. However, the payment schedule was negatively related to the amount of time selected (i.e., selecting a longer timeframe to complete the errands yielded less reward). If the participant was unsuccessful at completing the errands and returning to the starting location within the selected timeframe, the bonus for that trial was forfeited. Cronbach's alpha was computed to assess the internal consistency of both errand running times and timeframe selections across all trials, as well as wide and narrow trials separately.

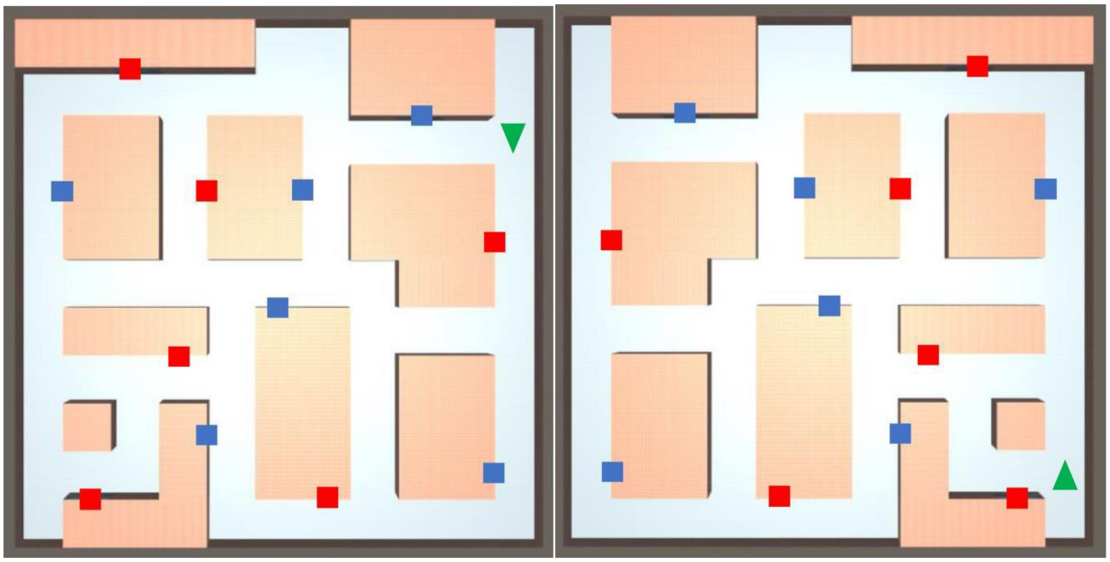


Figure 1. Bird's-eye view of the original (top left) and mirror (top right) virtual shopping malls, and a ground-level view of the original environment (bottom). The green arrow represents the starting location. Blue and red squares represent storefronts with narrow (8-12 seconds) and wide (2-18 seconds) ranging waiting time distributions, respectively.

Second, test-retest reliability was assessed by having participants perform the errand running task across two sessions and environments. The second environment was a mirrored version of the virtual shopping mall used in the pilot experiments, with an alternate starting position and alternate storefronts similar to the procedure used by Boone et al. (2019). The order in which the environments were presented across sessions was counterbalanced. Correlations were conducted to assess the relationship between performance across the two environments.

Third, the construct validity of the errand running task was examined by correlating errand running performance with various measures of spatial knowledge of the environment, ability, and decision-making. In the first session, participants completed the errand running task, judgments of relative direction for storefronts in the virtual shopping mall, and a route-finding task in which participants were told to go from one storefront to another as quickly as possible. Errand running trials varied in the amount of aleatory uncertainty involved by manipulating the uncertainty around waiting times at each destination on the errand list. In the second session, participants completed several psychometric measures of spatial ability (e.g., spatial working memory, sense of direction, and mental rotation) and decision-making (e.g., BART and DOSPERT). The psychometric measures from the second session (and spatial knowledge measures from the first session) were tested as predictors of errand running performance.

CHAPTER 2

PILOT EXPERIMENTS

Two pilot experiments were conducted to assess the internal consistency of the errand running task, whether participants understood the directions and task, and whether the timeframe options presented to participants before each trial were appropriate based on how long participants took to complete their errands.

Pilot Experiment 1

Methods

Participants

Participants were United States citizens (*n* = 16; 53% male; *M* age = 26.38, *SD* = 4.53) recruited through Prolific (https://www.prolific.co/) and compensated monetarily for participation. Regarding race, 75% identified as white/Caucasian, 12.5% identified as black/African American, 6.25% identified as Asian, and 6.25% identified as biracial. Participants were categorized as having completed a high school education or GED equivalent (18.75%), completed or attended college (75%), or completed or attended graduate school (6.25%). Participants completed the experiment on a laptop (62.5%) or desktop (37.5%) computer.

Materials

The shopping mall environment's (Figure 1) layout was identical to Marchette et al.'s (2011) maze used in the dual-solution paradigm, except that the landmarks were replaced by storefronts from Newman et al. (2007). The environment was 33 x 33 m in virtual space. Each storefront had an associated waiting time range such that six storefronts were associated with a narrow range (8-12 sec) and the remaining six were associated with a wide range (2-18 sec). Storefronts had waiting time ranges assigned to them such that both wide and narrow ranges were (approximately) evenly distributed throughout the maze. To reduce erroneous between-subject variability, waiting times were pre-generated from each storefront's associated waiting time range such that all participants waited the same amount of time at each storefront for each trial (Goodhew & Edwards, 2019) (see Appendix A for the list of waiting times pre-generated for each storefront for each trial). Participants were not told about waiting times being pre-generated. The textures on the ground and walls were gray-colored stone and red brick, respectively. No distal landmarks were present.

Procedure

Participants were first prompted to answer a few demographic questions about their age, gender, race, highest level of education achieved, device (desktop or laptop), and screen/monitor size. Participants then moved to the learning phase (wayfinding task) and were instructed that they would perform a navigation task in which they would navigate to several storefronts in a virtual shopping mall. They were informed that each storefront was associated with either a wide or narrow waiting time range, and that these ranges were 8-12 seconds and 2-18 seconds, respectively. Participants navigated through the environment using the WASD keys (W: translate forward; S: translate backward; D: rotate right; A: rotate left). Once participants arrived at a target storefront, they pressed the SPACEBAR key to confirm their arrival and initiate the waiting time countdown. During the countdown, participants were unable to translate or rotate their perspective while an on-screen countdown timer, beginning at the pre-generated waiting time, counted down to zero, at which time the next target storefront was presented, and participants regained translational and rotational movement. This process repeated until participants had navigated to all 12 storefronts in the virtual shopping mall. All participants visited each storefront in the same order (see Appendix A).

Once all 12 storefronts had been visited, participants were instructed to navigate back to where they started the learning phase (Figure 1), which was marked by a blue post. If the participant visited all 12 storefronts and returned to the starting location within 15 minutes, they moved onto the errand running phase; otherwise, they repeated the learning phase up to 3 times. Participants who were unsuccessful at completing the learning phase within 15 minutes were debriefed and the experiment concluded. Participants were not informed of the 15-minute timer.

During the errand running phase, participants were first informed of the nature of the task. Participants were told that an errand list of 3 storefronts (and their associated waiting time ranges) would be displayed before each trial, and their task was to navigate to each storefront in the same order as the storefronts were presented in the errand list. However, while viewing the errand list, participants were required to choose from a list of timeframe options the total amount of time predicted to visit each storefront on the errand list and return to the starting location of each trial, which was the same starting location from the learning phase and marked by a blue post. If participants completed the errands and returned to the starting location within the selected timeframe, a monetary bonus was added to their compensation for participation. However, there was a catch: the monetary bonus associated with each timeframe option was negatively related to the timeframe selected (i.e., giving oneself more time would yield less reward). If the participant failed to complete the errands and return to the starting location within the selected timeframe, the bonus for that trial was forfeited.

Participants were then shown an example of an errand list and the timeframe options. While the errand list varied across trials, the timeframe options remained constant. The 7 timeframe options started at 1 minute and 30 seconds and increased by increments of 15 seconds to a maximum timeframe of 3 minutes. The payment schedule for successful trials began at 35 cents for the shortest timeframe and decreased by 5 cents as the timeframe increased. When participants finished viewing the example, they began the experimental trials. There were 32 trials in total, and participants were told how many trials remained at the beginning of each. Waiting time uncertainty was manipulated within participants: 16 wide trials presented errand lists containing 3 storefronts with wide ranging waiting times and 16 narrow trials presented errand lists containing 3 storefronts with narrow ranging waiting time. Wide and narrow trials were interleaved, beginning with a wide trial. Participants were presented with an errand list containing the 3 storefronts with their associated waiting time ranges on the right side of the screen and the timeframe options with their associated potential bonuses on the left side of the screen. Participants then selected a timeframe by pressing the a, b, c, d, e, f, or g keys, with timeframe length increasing alphabetically. An instructional screen reiterating the selected timeframe and rules for receiving a bonus for that trial was presented before participants began the navigation phase of the trial.

Participants began the navigation phase of each errand running trial at the same location as during learning (Figure 1), which was marked by a blue post. The WASD keys were used to translate and rotate the perspective. The errand list was always visible at the top left corner of the screen (Figure 2), with the current target storefront in green font (past and future targets in black font). Like during learning, participants pressed the SPACEBAR key upon arrival at each target and were unable to move their perspective until the waiting time countdown expired. Another countdown timer was presented at the bottom of the screen for the duration of each trial and continuously counted down from the selected timeframe until it reached zero or the trial terminated. Trials only terminated upon arrival at the starting location after all errands had been run, regardless of whether the countdown timer had reached zero. Upon returning to the starting location, participants were given feedback on whether they successfully completed the trial within the selected timeframe. Participants repeated this process for all 32 trials. Following the final trial, participants were debriefed and thanked for their participation.

1. Coffee Store (wait time is 8-12 sec)
2. Toy Store (wait time is 8-12 sec)
3. Butcher Shop (wait time is 8-12 sec)

0:47.88

Move and rotate with the arrow keys.

Figure 2. An example of a participant's perspective during an errand running trial. The errand list is presented in the top left corner of the screen and the current target is presented in green

font. At the bottom of the screen are the remaining time to complete the trial (beginning of the countdown depended on the selected timeframe), and instructions to move and rotate the view.

Analyses

Internal consistency was examined by computing Cronbach's alpha for errand running time, defined as the total amount of time (seconds) participants took to visit all three errand destinations in a trial and return to the starting location, and timeframe selection for all 32 trials, as well as wide and narrow trials separately (see Table 1).

To examine the extent to which participants were able to accurately estimate the timeframe required to complete the errands for each trial, the average number of successful trials, and average excess time on successful trials (henceforth referred to as average excess time for brevity) were computed. The Shapiro-Wilk test revealed that several pairs of variables violated the assumption of bivariate normality, and so Spearman's rank correlation coefficient was computed instead of Pearson's correlation coefficient. Spearman correlations were computed for several variables (Table 4). Of particular interest here were the relationships between number of successful trials, excess time on successful trials, average errand running time, and average timeframe selection. It should be noted that the pilot sample was relatively small, thus significance tests for these relationships were underpowered.

To examine the relationship between wayfinding performance and errand running performance, the average wayfinding time across all 12 wayfinding trials of the first wayfinding attempt was computed (henceforth referred to as average wayfinding time for brevity). The relationship between average wayfinding time and average errand running time, average timeframe selection, average excess time, and number of successful trials was examined.

To examine differences in errand running performance across wide and narrow trials, average errand running time, average timeframe selection, average excess time, and average number of successful trials were computed separately for each trial type. Because the sample size was small, null hypothesis testing was not conducted. Instead, means are reported in Table 2. Finally, males and females were compared across the same dependent measures above and average wayfinding time. As with trial type, null hypothesis testing was not conducted on gender differences. One participant did not report their gender and was not included in the marginal means. Means are reported in Table 3.

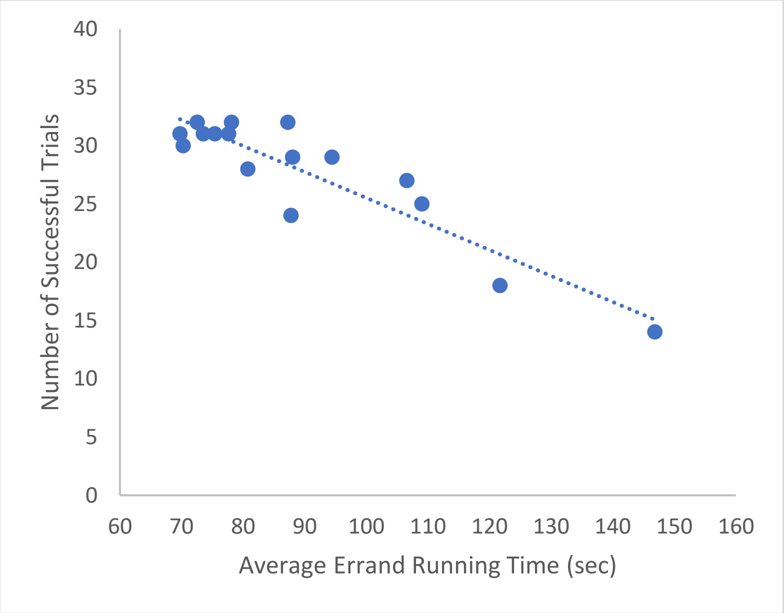
Results

Cronbach's alpha was high for errand running time and timeframe selection for all trials, as well as wide and narrow trials separately (Table 1).

	Cronbach's Alpha
Errand Running Time All Trials	.95
Errand Running Time Wide Trials	.92
Errand Running Time Narrow Trials	.88
Timeframe Selection All Trials	.98
Timeframe Selection Wide Trials	.95
Timeframe Selection Narrow Trials	.97

Table 1. Cronbach's alpha for errand running time and timeframe selection for all 32 trials, as well as wide and narrow trials separately.

On average, participants completed every errand on the errand list and returned to the starting location within the selected timeframe on most trials (*M* = 27.75; *min* = 14, *max* = 32). There was a strong relationship between average errand running time and number of successful trials, such that participants who took longer to complete errands were less successful at completing the errands within the chosen timeframe on average (Figure 3). However, the relationship between average timeframe selection and number of successful trials was unreliable. On average, participants who chose shorter timeframes were successful at completing the errands within the selected timeframe as often as those who chose longer timeframes. There was also a strong relationship between average errand running time and average timeframe selection, suggesting that participants were able to predict the time required to run the errands (Figure 3). Despite this, participants tended to err on the side of caution, overestimating the required timeframe by roughly 32 seconds on average (Table 4).



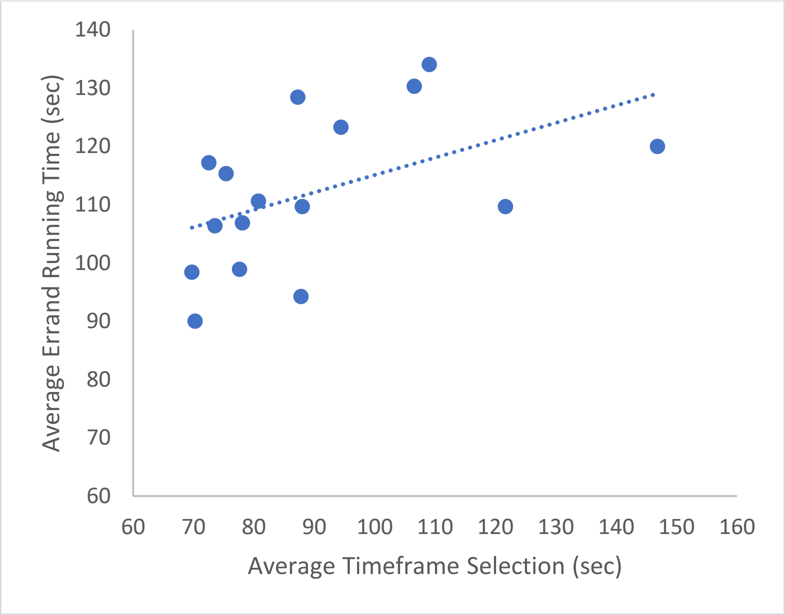


Figure 3. Top: The relationship between average errand running time and number of successful trials. Bottom: The relationship between average errand running time and average timeframe selection.

Nearly all participants completed the wayfinding task within 15 minutes during the first attempt (15 out of 16), with one participant requiring two attempts. There was a strong relationship between average wayfinding time and average errand running time, such that participants who were faster during learning were also faster errand runners on average. However, the relationship between average wayfinding time and average timeframe selection was unreliable. There was also a strong relationship between average wayfinding time and number of successful trials during the errand running task, such that participants who learned the environment faster also completed more errand running trials within the selected timeframe on average.

	Wide	Narrow
Avg. Errand Running Time	91.69 (21.44)	88.26 (22.33)
Avg. Timeframe Selection	113.44 (12.55)	110.74 (13.90)
Avg. Excess Time	32.57 (8.79)	30.73 (9.24)
Avg. # of Successful Trials	13.63 (2.80)	14.13 (2.83)

Table 2. Displays means (and standard deviations) for average errand running time, average timeframe selection, average excess time, and average number of successful trials during wide and narrow trials. Time is reported in seconds.

	Male	Female
Avg. Errand Running Time	76.29 (6.08)	107.69 (21.26)
Avg. Timeframe Selection	102.83 (9.02)	122.21 (9.70)
Avg. Excess Time	29.48 (8.81)	32.82 (8.75)
Avg. # of Successful Trials	29.88 (2.70)	24.86 (6.52)
Avg. Wayfinding Time	26.22 (6.65)	43.54 (8.37)

Table 3. Displays means (and standard deviations) for average errand running time, average timeframe selection, average excess time, average total successful trials, and average wayfinding time for males and females. Time is reported in seconds.

	M (SD)	Skewness	Kurtosis	Min, Max	1.	2.	3.	4.	5.
1. Avg. Errand Running Time	89.97 (21.41)	1.36	4.15	[69.74, 146.83]	-				
2. Avg. Timeframe Selection	112.09 (13.03)	0.44	2.08	[90.00, 134.06]	.58*	-			
3. Avg. Excess Time	31.68 (8.73)	-0.13	1.77	[18.13, 44.66]	-.07	.63**	-		
4. Number of Successful Trials	27.75 (5.23)	-1.54	4.40	[14, 32]	-.75***	-.15	.34	-	
5. Avg. Wayfinding Time	33.32 (11.78)	0.47	2.04	[18.59, 56.56]	.80***	.39	-.24	-.64**	-

Table 4. Spearman correlation matrix for all dependent variables. * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

The primary goal of this pilot experiment was to test the errand running task's internal consistency, a measure of the degree to which each item (or trial) measures the same construct. To this aim, the pilot demonstrated a high degree of internal consistency as assessed by Cronbach's alpha for both errand running time and timeframe selection. While internal consistency was slightly lower for errand running time when analyzed separately for wide and narrow trials, it was still acceptable. Furthermore, a large majority of participants successfully completed the learning phase (wayfinding task) within the first attempt (only one participant required two attempts) and participants were mostly successful at completing their errands within the selected timeframe (nearly 28 successes out of 32 trials on average). The latter of these two results suggests that participants were able to understand the instructions and that the task was not too difficult.

Several correlations between dependent measures were examined as well. While the correlation between average errand running time and average timeframe selection suggested that there was a strong relationship between errand running times and selected timeframes, the grand mean of average errand running times indicated that timeframe options were not well calibrated to match the time requirements of the errand lists. The shortest timeframe option was 1 minute and 30 seconds, and participants completed their errands in that amount of time almost exactly on average (i.e., 89.97 sec). In other words, the average errand running time was less than the shortest timeframe, and thus timeframe options could have been better calibrated if they were reduced by a constant.

There was a strong negative relationship between errand running time and number of successful trials, which may suggest that participants who took longer to complete their errands had a weaker representation of the environment and therefore underestimated the required timeframe. This result is seemingly at odds with the lack of a relationship between average timeframe selection and number of successful trials. However, the absence of a relationship between average timeframe selection and number of successful trials was possibly driven by participants who greatly overestimated the required time to complete their errands. In fact, average excess time and average timeframe selection were highly correlated, suggesting that participants who selected longer timeframes were greatly overestimating the required timeframes. It is possible that this result is attributable to the timeframe options having been incommensurate with the average errand running time.

Not surprisingly, participants who learned the environment faster during the wayfinding task were also faster at running errands as evidenced by the large correlation between average wayfinding time and average errand running time. Furthermore, participants who completed the wayfinding task faster selected shorter timeframes on average and were also more likely to be successful at completing their errands during the errand running task. While the relationship between average wayfinding time and number of successful trials was strong, it was still much weaker than the relationship between average errand running time and number of successful trials.

When wide and narrow trials were compared, few differences emerged. It is possible that navigators are more concerned about differences in the average waiting time at each destination than they are about differences in the range of waiting times. Finally, males and females were compared. Four of the five dependent measures in Table 4 were in favor of males. On average, males completed their errands faster, selected shorter timeframes, were successful on more trials, and completed the wayfinding task faster. While males had less excess time on successful trials on average, there was only a 3 second difference. Generally, the results of the errand running task were consistent with previous work showing that males were faster at wayfinding tasks (Boone et al., 2018; Munion et al., 2019; Yu et al., 2021). However, these results were inconsistent with previous work showing that differences between males and females during wayfinding were reduced when participants were instructed to navigate to a target as quickly as possible. In the current task, efficient navigation is incentivized by increased pay for faster errand running times. However, participants were free to choose their timeframe during each trial, thus there were no explicit instructions to navigate as quickly as possible.

While the current pilot was a success insofar as the errand running task was demonstrated to have high internal consistency, it also revealed that the timeframe options were not calibrated to the actual time to complete the errands. Thus, a second pilot study was conducted in which timeframe options were reduced by a constant of 30 seconds to better calibrate the timeframe options with the actual time requirements of the errand lists.

Pilot Experiment 2

The goal of the second experiment was to calibrate the timeframe options with the actual time requirements of the errand lists by reducing the timeframe options by a constant of 30 seconds.

Methods

Participants

Participants were United States citizens (*n* = 24; 54% male; *M* age = 27.46, *SD* = 5.11) recruited through Prolific and compensated monetarily for participation. Regarding race, 70.83% identified as white/Caucasian, 8.33% identified as black/African American, 8.33% identified as Asian, 4.17% identified as biracial, and 8.33% identified as Hispanic. Participants were categorized as having completed less than a high school education (8.33%), completed high school education or GED equivalent (20.83%), completed or attended college (62.5%), or completed or attended graduate school (4.17%). Participants completed the experiment on a laptop (62.5%) or desktop (37.5%) computer.

Materials, Procedure, and Analyses

The materials and procedure were identical to Pilot Experiment 1, except that the timeframes were reduced by 30 seconds, with the shortest timeframe of 1 minute and longest timeframe of 2 minutes and 30 seconds. The payment schedule was identical to Pilot Experiment 1. The analyses were also identical to Pilot Experiment 1. Two participants did not report their gender and were not included in the marginal means when comparing males and females. Means by trial type and gender are reported in Tables 6 and 7, respectively.

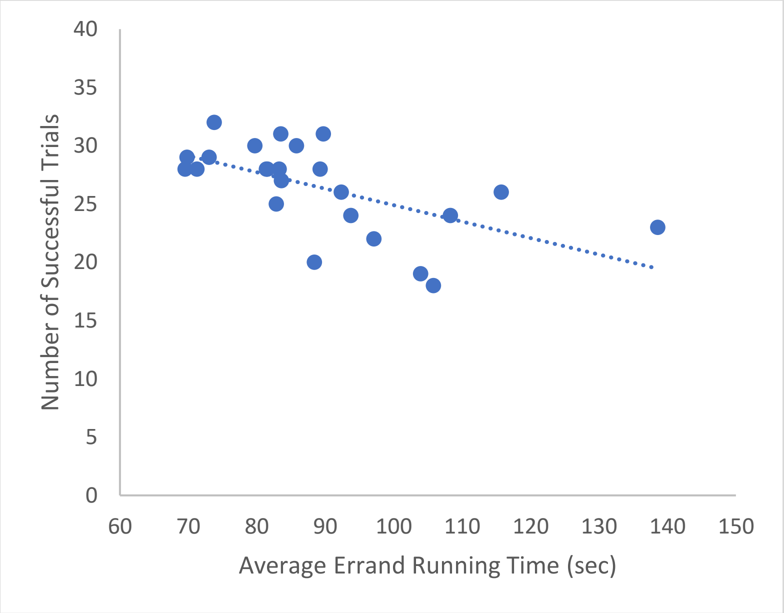
Results

Cronbach's alpha was high for errand running time and timeframe selection for all 32 trials, as well as wide and narrow trials separately (Table 5).

	Cronbach's Alpha
Errand Running Time All Trials	.95
Errand Running Time Wide Trials	.93
Errand Running Time Narrow Trials	.87
Timeframe Selection All Trials	.97
Timeframe Selection Wide Trials	.96
Timeframe Selection Narrow Trials	.95

Table 5. Cronbach's alpha for errand running time and timeframe selection for all 32 trials, as well as wide and narrow trials separately.

On average, participants completed every errand on the errand list and returned to the starting location within the selected timeframe on most trials (*M* = 26.42; *min* = 18, *max* = 32). There was a strong relationship between average errand running time and number of successful trials, such that participants who took longer to complete errands were less successful at completing errands within the selected timeframe on average (Figure 4). However, the relationship between average timeframe selection and number of successful trials was unreliable. There was a strong relationship between average errand running time and average timeframe selection, suggesting that participants were able to predict the time required to run the errands (Figure 4). Despite this, participants tended to err on the side of caution, overestimating the required timeframe by roughly 28 seconds on average (Table 8).



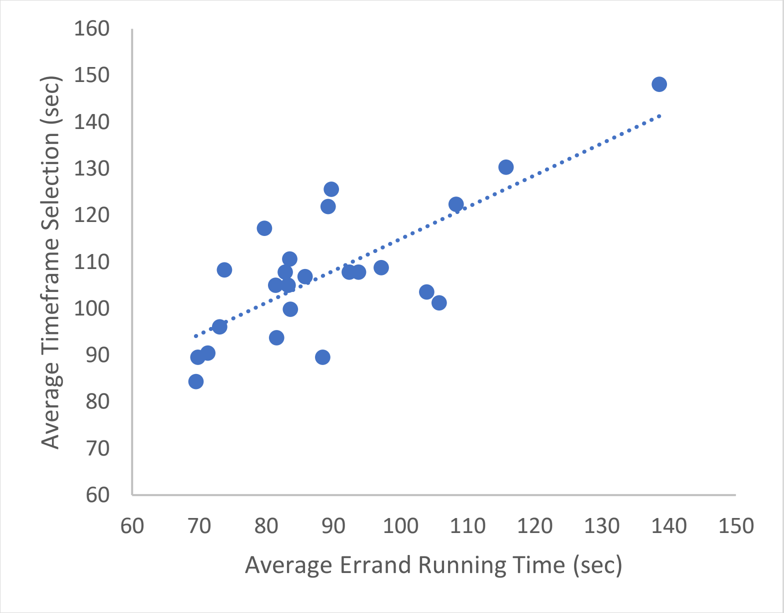


Figure 4. Top: The relationship between average errand running time and number of successful trials. Bottom: The relationship between average errand running time and average timeframe selection.

All participants completed the wayfinding task within 15 minutes during the first attempt. Unlike Pilot Experiment 1, average wayfinding time was not reliably associated with any of the dependent variables from the errand running task, suggesting that performance during learning was not predictive of performance during errand running.

	Wide	Narrow
Avg. Errand Running Time	91.14 (18.93)	87.34 (14.25)
Avg. Timeframe Selection	108.28 (15.81)	106.88 (14.86)
Avg. Excess Time	27.32 (8.81)	29.34 (6.38)
Avg. # of Successful Trials	12.79 (2.41)	13.63 (1.92)

Table 6. Displays means (and standard deviations) for average errand running time, average timeframe selection, average excess time, and average number of successful trials during wide and narrow trials. Time is reported in seconds.

	Male	Female
Avg. Errand Running Time	84.79 (14.10)	96.38 (19.02)
Avg. Timeframe Selection	101.47 (12.46)	114.79 (15.52)
Avg. Excess Time	25.12 (4.45)	30.97 (6.93)
Avg. # of Successful Trials	26.92 (3.93)	25.67 (4.21)
Avg. Wayfinding Time	31.38 (8.74)	40.37 (8.35)

Table 7. Displays means (and standard deviations) for average errand running time, average timeframe selection, average excess time, average total successful trials, and average wayfinding time for males and females. Time is reported in seconds.

	M (SD)	Skewness	Kurtosis	Min, Max	1.	2.	3.	4.	5.
1. Avg. Errand Running Time	89.24 (16.20)	1.28	4.75	[69.51, 138.58]	-				
2. Avg. Timeframe Selection	107.58 (14.67)	0.80	3.75	[84.58, 148.13]	.60**	-			
3. Avg. Excess Time	28.33 (6.57)	0.46	2.60	[17.56, 42.15]	.18	.68***	-		
4. Number of Successful Trials	26.42 (3.84)	-0.71	2.63	[18, 32]	-.65***	.05	.16	-	
5. Avg. Wayfinding Time	34.24 (9.53)	0.12	1.96	[19.45, 50.23]	.26	.14	.00	-.24	-

Table 8. Spearman correlation matrix for all dependent variables. * $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

As in Pilot Experiment 1, the internal consistency (Cronbach’s alpha) of the errand running task was high for both errand running time and timeframe selection. This was true for all 32 trials, as well as wide and narrow trials separately. Additionally, most of the dependent variables showed similar means to those observed in Pilot Experiment 1, suggesting that the task is robust across samples, albeit relatively small samples. Importantly, the relationship between average errand running time and average timeframe selection was relatively unaffected by the change in the timeframe options between the pilot experiments.

While internal consistency is a requirement of valid psychometric assessments, it is insufficient to adequately demonstrate reliability. A psychometric assessment must also be reliable across time, such that there is *consistency* in the rank-ordering of participants and *agreement* between absolute scores across testing sessions (Koo & Li, 2016; Liu et al., 2016; Shrout & Fleiss, 1979; Vaz et al., 2013). A common approach to demonstrating reliability is to use the test-retest procedure in which the same participants complete the same test across two different testing sessions. In Experiment 1, I tested the test-retest reliability of the errand running task by having participants perform the task across two testing sessions, which were separated by at least 24 hours.

CHAPTER 3

EXPERIMENT 1

The purpose of Experiment 1 is to assess the test-retest reliability of the errand running task by having participants perform the task across two testing sessions, which were separated by at least 24 hours. While it is often recommended to wait a long enough period between testing sessions as to avoid learning effects from the first testing session, another way to circumvent possible learning effects is to use the alternate-forms procedure (Rotenberg et al., 2020). In the alternate-forms procedure, two different versions of the test are provided across testing sessions; in this case, participants completed the same errand running task across two different environments (see Methods). The first environment was identical to the pilot experiments. The second environment was a mirrored version of the first environment, and the storefronts were replaced by different storefronts. However, to equate task difficulty across environments, the trials were matched across environments such that the relative locations of the storefronts were identical, and the same trials were administered in the same order across environments (Appendix A). Pre-generated waiting times were also matched across environments.

Methods

Participants

Participants were United States citizens (*n* = 63; 50% male; *M* age = 28.16, *SD* = 5.25) recruited through Prolific and compensated monetarily for participation. Power analyses showed that a sample size of 59 would be sufficient to detect a moderate test-retest correlation (Correlation: H1 $\rho = .35$, $\alpha = .05$, Power = .80) (Faul, Erdfelder, Buchner, & Lang, 2009). Regarding race, 60.66% identified as white/Caucasian, 21.31% identified as black/African American, 6.56% identified as Asian, 1.64% identified as biracial, 6.56% identified as Hispanic, and 3.28% identified as Native American. Participants were categorized as having completed high school education or GED equivalent (26.67%), completed or attended college (65%), or completed or attended graduate school (8.33%). Participants completed the experiment on a laptop (67.8%) or desktop (32.2%) computer.

Materials

The errand running task was implemented in two virtual shopping mall environments. The *original environment* was identical to the environment used in the pilot experiments. The *mirror environment* differed from the original environment in three ways (see Figure 1): first, the configuration of the environment was mirrored; second, all storefronts were replaced by different storefronts from Newman et al. (2007); third, the starting position was moved to a different corner of the environment relative to the original environment's configuration. However, trials were matched across environments such that the relative locations of the storefronts were identical, and the same trials were administered in the same order across environments. Pre-generated waiting times were also matched across environments. Full details of the wayfinding and errand running trials for both environments are reported in Appendix A.

Procedure

All participants received both environments, one during each session, with the order of environment counterbalanced across participants. Participants logged into the first session using their unique Prolific identification number (PID), which was used to keep a record of which environment they received during the first session. Then, they answered demographic questions and completed the wayfinding and errand running tasks following the same procedures as the pilot experiments. Participants logged into the second session by entering their PID, and then proceeded with the wayfinding and errand running tasks in the environment opposite to the first session. Participants were then debriefed and thanked for their participation. The second session was not made available to participants until at least 24 hours following the first session.

Analyses

As in the pilot experiments, Cronbach's alpha was computed for errand running time and timeframe selection for all 32 trials, as well as wide and narrow trials separately. These analyses were conducted on data from both the original and mirror environments separately (Table 9). Due to a server-side issue, data from 1 ($n = 5$), 2 ($n = 8$), or 3 trials ($n = 5$) were lost for some participants during the first session, and 3 trials were lost for one participant during the second session. Cronbach's alpha was only computed for participants with all 32 trials (original: $n = 51$; mirror: $n = 54$).

Test-retest reliability was examined by computing both consistency and agreement intra-class correlations (ICC) between testing sessions (Shrout & Fleiss, 1979). Consistency ICC (denoted ICC_C below) is like Pearson's correlation coefficient in that a higher ICC_C is indicative of a higher consistency of relative rankings of participants across testing sessions but does not penalize shifts in overall means across testing sessions. On the other hand, agreement ICC (denoted ICC_A below) treats testing session as a random effect, and like ICC_C, measures the consistency of the relative rankings of participants across testing sessions, but also penalizes for shifts in overall means. In other words, ICC_C places an upper bound on ICC_A. Finally, for each type of ICC, the Spearman-Brown formula was applied (denoted ICC_{CS} and ICC_{AS} below), yielding the proportion of variance in composite scores averaged across testing sessions that is due to individual differences. Each ICC type (ICC_C, ICC_A, ICC_{CS}, and ICC_{AS}) was computed for average errand running time, average timeframe selection, average excess time, average excess time, and number of successful trials. For brevity, only the Spearman-Brown adjusted ICCs (ICC_{CS} and ICC_{AS}) are reported below, but an exhaustive list is provided in Table 10.

Differences in average errand running time, average timeframe selection, average excess time, and number of successful trials between males and females, environmental orders, wide and narrow trials, and sessions were examined by separate four-way mixed ANOVAs. Specifically, each ANOVA followed a 2 (gender) x 2 (environmental order) x 2 (trial type) x 2 (environment) design, with gender and environmental order as between-subjects variables and trial type and session as within-subjects variables (see Tables 11 and 12). Two participants did not report their gender and another participant, due to experimenter error, completed the errand running task in the mirror environment during both sessions. These three participants were not included in the analyses.

The same correlations computed in the pilot experiments were computed for each environment separately (Tables 13 and 14). The participant who received the mirror environment in both sessions was not included in these analyses.

Results

Cronbach's alpha was high for errand running time and timeframe selection for all 32 trials, as well as wide and narrow trials separately (Table 9). This was true for both original and mirror environments.

	Cronbach's Alpha

	Original Environment	Mirror Environment
Errand Running Time All Trials	.96	.96
Errand Running Time Wide Trials	.93	.93
Errand Running Time Narrow Trials	.91	.92
Timeframe Selection All Trials	.98	.99
Timeframe Selection Wide Trials	.97	.97
Timeframe Selection Narrow Trials	.98	.98

Table 9. Cronbach's alpha for errand running time and timeframe selection for all 32 trials, as well as wide and narrow trials separately. Columns are separated by environment.

The errand running task showed high test-retest reliability across all four dependent measures: average errand running time, average timeframe selection, average excess time, and number of successful trials (Table 10). This was true for both consistency and agreement ICCs, suggesting that participants' ranks and mean scores were similar across testing sessions.

	ICC_C	ICC_A	ICC_{CS}	ICC_{AS}
Avg. Errand Running Time	.86	.86	.93	.92
Avg. Timeframe Selection	.95	.94	.97	.97
Avg. Excess Time	.84	.84	.91	.91

Number of Successful Trials	.83	.81	.91	.90

Table 10. Table containing between-session consistency and agreement ICCs, including Spearman-Brown formula adjusted ICCs (ICC_{C}, ICC_{A}, ICC_{CS}, and ICC_{AS}, respectively), for average errand running time, average timeframe selection, average excess time, and number of successful trials.

Average errand running time was analyzed by a four-way Mixed ANOVA (see Analyses). The main effect of gender was significant, $F(1,55) = 8.81$, $p = .004$, $\eta_p^2 = .14$. On average, males ($M = 80.49$, $SD = 13.11$) completed errands faster than females ($M = 93.12$, $SD = 18.14$). The interaction between gender and environmental order was significant, $F(1,55) = 4.01$, $p = .050$, $\eta_p^2 = .07$. However, males completed errands faster than females regardless of which environment was received first. The interaction between trial type and environment was significant, $F(1,55) = 44.20$, $p < .001$, $\eta_p^2 = .45$. Participants were slower to complete wide trials than narrow trials in the original environment but faster to complete wide trials than narrow trials in the mirror environment (Table 11).

Average timeframe selection was analyzed by a four-way Mixed ANOVA. The main effect of gender was significant, $F(1,55) = 13.43$, $p < .001$, $\eta_p^2 = .20$. On average, males ($M = 99.48$, $SD = 12.02$) selected shorter timeframes than females ($M = 113.51$, $SD = 16.24$). The interaction between trial type and environment was significant, $F(1,55) = 26.30$, $p < .001$, $\eta_p^2 = .32$. Participants selected shorter timeframes during narrow trials than wide trials in the original environment but selected shorter timeframes during wide trials than narrow trials in the mirror environment (Table 11).

Average excess time was analyzed by a four-way Mixed ANOVA. The main effect of gender was significant, *F*(1,55) = 6.26, *p* = .015, η_p^2 = .10. On average, males had less excess time (*M* = 24.84, *SD* = 7.30) than females (*M* = 29.51, *SD* = 8.14). The interaction between trial type and environment was significant, *F*(1,55) = 17.64, *p* < .001, η_p^2 = .24. Participants had more excess time during narrow trials than wide trials in the original environment but had more excess time during wide trials than narrow trials in the mirror environment (Table 11).

Number of successful trials was analyzed by a four-way Mixed ANOVA. The main effect of trial type was significant, *F*(1,55) = 9.20, *p* = .004, η_p^2 = .14. On average, participants successfully completed more wide trials (*M* = 13.31, *SD* = 2.46) than narrow trials (*M* = 13.92, *SD* = 2.41). The interaction between environment and order of environment was significant, *F*(1,55) = 10.35, *p* = .002, η_p^2 = .16. The interaction between gender and environmental order was significant, *F*(1,55) = 4.95, *p* = .030, η_p^2 = .08. However, none of the pairwise comparisons were significant. All other *p*s > .050.

	Original Env.		Mirror Env.	
	Wide	Narrow	Wide	Narrow
Avg. Errand Running Time	89.03 (16.39)	85.18 (19.41)	82.97 (17.31)	89.92 (17.06)
Avg. Timeframe Selection	106.98 (16.63)	105.93 (15.72)	104.28 (17.43)	108.31 (15.31)
Avg. Excess Time	27.52 (8.97)	28.57 (8.64)	28.64 (9.66)	25.25 (8.25)
Avg. # of Successful Trials	13.02 (2.84)	13.92 (2.53)	13.59 (2.66)	13.93 (2.74)

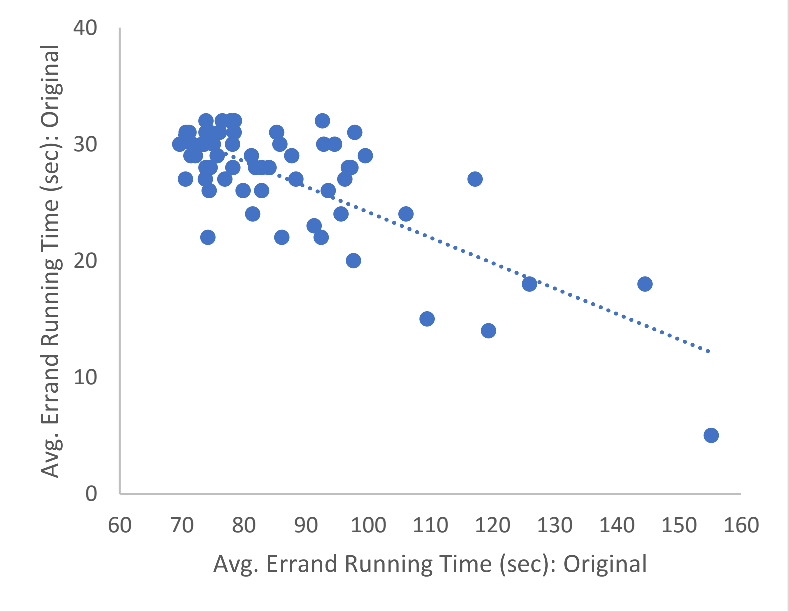
Table 11. Displays means (and standard deviations) for average errand running time, average timeframe selection, average excess time, and average number of successful trials during wide and narrow trials for each environment. Time is reported in seconds.

	Original Env.		Mirror Env.	
	Male	Female	Male	Female
Avg. Errand Running Time	81.69 (16.05)	92.71 (17.50)	79.29 (9.44)	93.53 (19.07)
Avg. Timeframe Selection	99.72 (12.39)	113.43 (16.26)	99.25 (11.85)	113.59 (16.51)
Avg. Excess Time	25.85 (7.48)	30.37 (8.59)	24.56 (7.03)	29.56 (8.58)
Avg. # of Successful Trials	27..30 (5.23)	26.55 (4.90)	28.37 (4.21)	26.66 (5.74)
Avg. Wayfinding Time	34.59 (23.55)	40.81 (24.30)	33.61 (21.05)	36.91 (17.73)

Table 12. Displays means (and standard deviations) for average errand running time, average timeframe selection, average excess time, average total successful trials, and average wayfinding time for males and females for each environment. Time is reported in seconds.

Original Environment Correlations

On average, participants completed every errand on the errand list and returned to the starting location within the selected timeframe on most trials in the original environment (*M* = 27.07; *min* = 5, *max* = 32). There was a strong relationship between average errand running time and number of successful trials, such that participants who took longer to complete errands were less successful at completing errands within the selected timeframe on average (Figure 5). There was also a weaker relationship between average excess time and number of successful trials, such that participants who played it safe by overestimating the required timeframe to complete the errands were more successful across trials on average. However, the relationship between average timeframe selection and number of successful trials was unreliable. There was a strong relationship between average errand running time and timeframe selection (Figure 5). As in the pilot experiments, participants were able to predict the time required to complete errands, but tended to err on the side of caution, overestimating the required timeframe by roughly 28 seconds on average (Table 13).



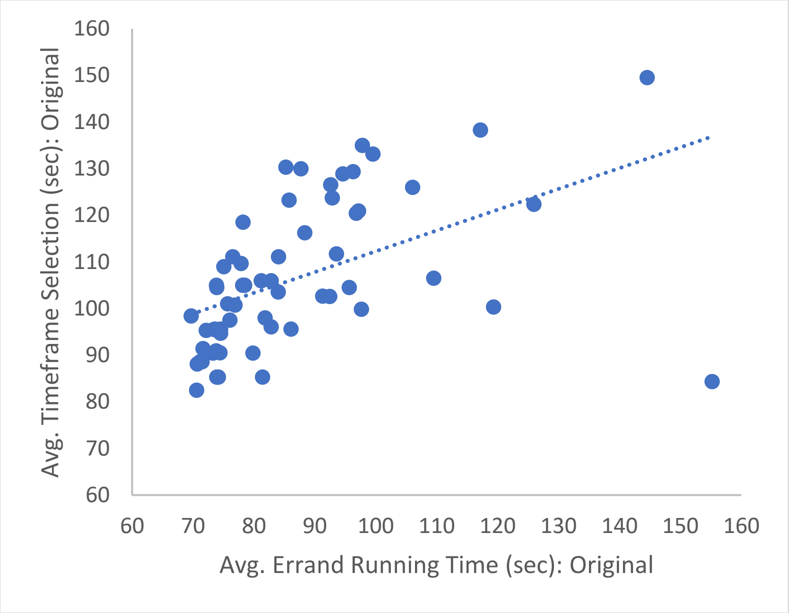


Figure 5. The relationship between average errand running time and number of successful trials in the original environment. Bottom: The relationship between average errand running time and average timeframe selection in the original environment.

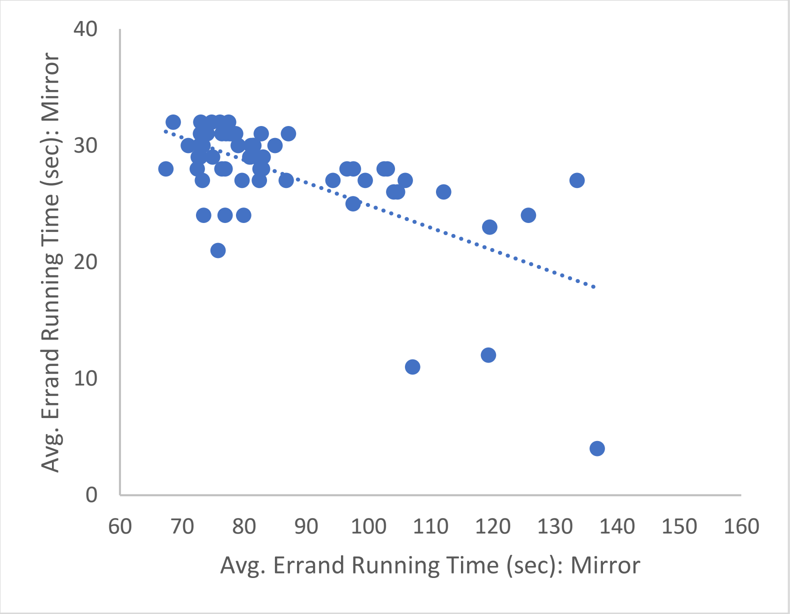
Nearly all participants completed the wayfinding task within 15 minutes during the first attempt (57 out of 61), with two participants requiring two attempts and two participants requiring three. There was a strong relationship between average wayfinding time and average errand running time, such that participants who were faster during learning were also faster errand runners on average. There was a weaker relationship between average wayfinding time and average timeframe selection, suggesting that participants who were faster during learning also selected shorter timeframes on average. There was also a relationship between average wayfinding time and number of successful trials, such that participants that were faster during learning were successful more often across errand running trials on average.

	Original Environment								
	M (SD)	Skewness	Kurtosis	Min, Max	1.	2.	3.	4.	5.
1. Avg. Errand Running Time	86.71 (17.38)	1.95	7.23	[69.67, 155.23]	-				
2. Avg. Timeframe Selection	106.36 (15.69)	0.63	2.64	[82.50, 149.53]	.65***	-			
3. Avg. Excess Time	28.15 (8.20)	0.44	2.41	[14.61, 47.09]	.42***	.89***	-		
4. Number of Successful Trials	27.07 (5.01)	-2.10	8.38	[5, 32]	-.49***	.12	.27*	-	
5. Avg. Wayfinding Time	37.38 (23.60)	3.67	18.03	[17.74, 154. 80]	.51***	.26*	.13	-.45***	-

Table 13. Spearman correlation matrix for all dependent variables from the original environment. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Mirror Environment Correlations

On average, participants completed every errand on the errand list and returned to the starting location within the selected timeframe on most trials in the mirror environment (*M* = 27.52; *min* = 4, *max* = 32). There was a strong relationship between average errand running time and number of successful trials, such that participants who took longer to complete errands were less successful at completing errands within the selected timeframe on average (Figure 6). However, the relationship between average errand running time and average timeframe selection was unreliable. There was a strong relationship between average errand running time and average timeframe selection, such that participants were able to predict the time required to run errands on average (Figure 6). Despite this, participants tended to err on the side of caution, overestimating the required timeframe by roughly 27 seconds on average (Table 14).



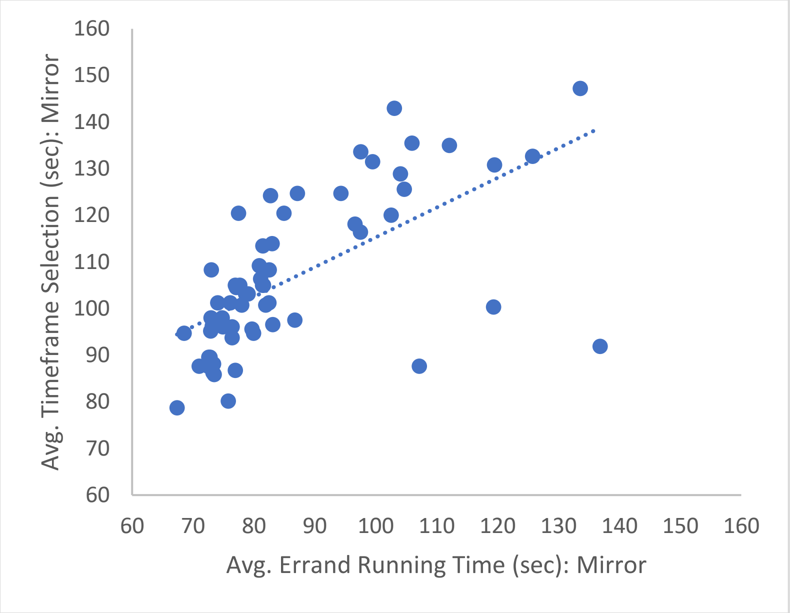


Figure 6. The relationship between average errand running time and number of successful trials in the mirror environment. Bottom: The relationship between average errand running time and average timeframe selection in the mirror environment.

Nearly all participants completed the wayfinding task within 15 minutes during the first attempt (56 out of 61), with four participants requiring two attempts and one participant requiring three attempts. There was a strong relationship between average wayfinding time and average errand running time, such that participants who were faster during learning also completed errands faster on average. There was also a relationship between average wayfinding time and average timeframe selection, such that participants who were faster during learning also selected shorter timeframes during errand running on average. Lastly, there was a weaker relationship between average wayfinding time and number of successful trials, such that participants who were faster during learning were successful during errand running more often on average.

Mirror Environment									
	M (SD)	Skewness	Kurtosis	Min, Max	1.	2.	3.	4.	5.
1. Avg. Errand Running Time	86.35 (16.43)	1.41	4.24	[67.37, 136.84]	-				
2. Avg. Timeframe Selection	106.57 (16.57)	0.57	2.44	[78.75, 147.19]	$.70^{***}$	-			
3. Avg. Excess Time	27.23 (8.71)	0.57	2.97	[10.77, 51.05]	$.44^{***}$	$.90^{***}$	-		
4. Number of Successful Trials	27.52 (4.97)	-2.86	12.39	[4, 32]	$-.53^{***}$	-.01	.17	-	
5. Avg. Wayfinding Time	35.62 (19.75)	2.39	10.20	[15.99, 128.72]	$.49^{***}$	$.32^{*}$	.23	-.20	-

Table 14. Spearman correlation matrix for all dependent variables from the mirror environment. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Discussion

The purpose of Experiment 1 was to assess the test-retest reliability of the errand running task. To this aim, participants completed the errand running task across two sessions and environments, with environments differing in the names of storefronts and being mirrors of each other. Test-retest was analyzed using consistency and agreement ICCs, which demonstrated that participants were consistent in their rank ordering across sessions (consistency) and mean scores shifted very little across sessions (agreement). This pattern held for all four primary dependent variables: average errand running time, average timeframe selection, average excess time, and number of successful trials. Thus, the errand running task has high test-retest reliability. Furthermore, consistent with the pilot experiments, internal consistency was again found to be high for both environments, this time with a much larger sample size. There was also consistency in the correlations between environments. Tables 13 and 14 showed that many of the pairwise correlations between the dependent variables of the errand running ask were highly similar between environments.

Concerning average errand running time, males, regardless of which environment was received first, outperformed females who received the original environment first. However, while not significant, average errand running time was in the same relative direction for females who received the mirror environment first (*M* = 86.57, *SD* = 12.33). There was also faster performance during wide trials in the original and the slowest during narrow trials in the mirror environment. However, average errand running times were within 80 to 90 seconds for each trial type by environment pair, suggesting that errand running time was fairly consistent across trial types and environments. The repeated-measures components of the ANOVAs were highly powered given the sample size, thus it is not surprising that mean differences between trial type by environment were observed.

Interestingly, correlations between average wayfinding time and errand running performance suggested that the errand running task is not simply another wayfinding task. While the correlation between average wayfinding time and average errand running time was still high, there was still a substantial amount of variance in average errand running time left unaccounted for. Similarly, the correlations between average wayfinding time and average timeframe selection, average excess time, and number of successful trials were much weaker.

While evidence of high test-retest reliability is promising, the construct validity of the errand running task needs to be assessed. In Experiment 2, participants completed the errand running task in the original environment only, as well as tests of their survey and route knowledge, in the first session. During the second session, participants completed several tests of spatial ability, general intelligence, and risk-taking behavior, and questionnaires about their navigation ability and anxiety, risk behaviors and perceptions, and perceptions of uncertainty about errand running. The relationships between these variables and errand running performance were examined.

CHAPTER 4

EXPERIMENT 2

The purpose of Experiment 2 was to assess the construct validity of the errand running task. To this aim, participants completed the same errand running task as in the previous experiments in the original environment. Several individual difference measures were administered to assess route and survey knowledge of the environment, self-reported sense of direction, mental rotation ability, spatial working memory, general intelligence, self-reported perceptions about the epistemic and aleatory nature of the errand running task, and risk-taking behaviors and perceptions.

Methods

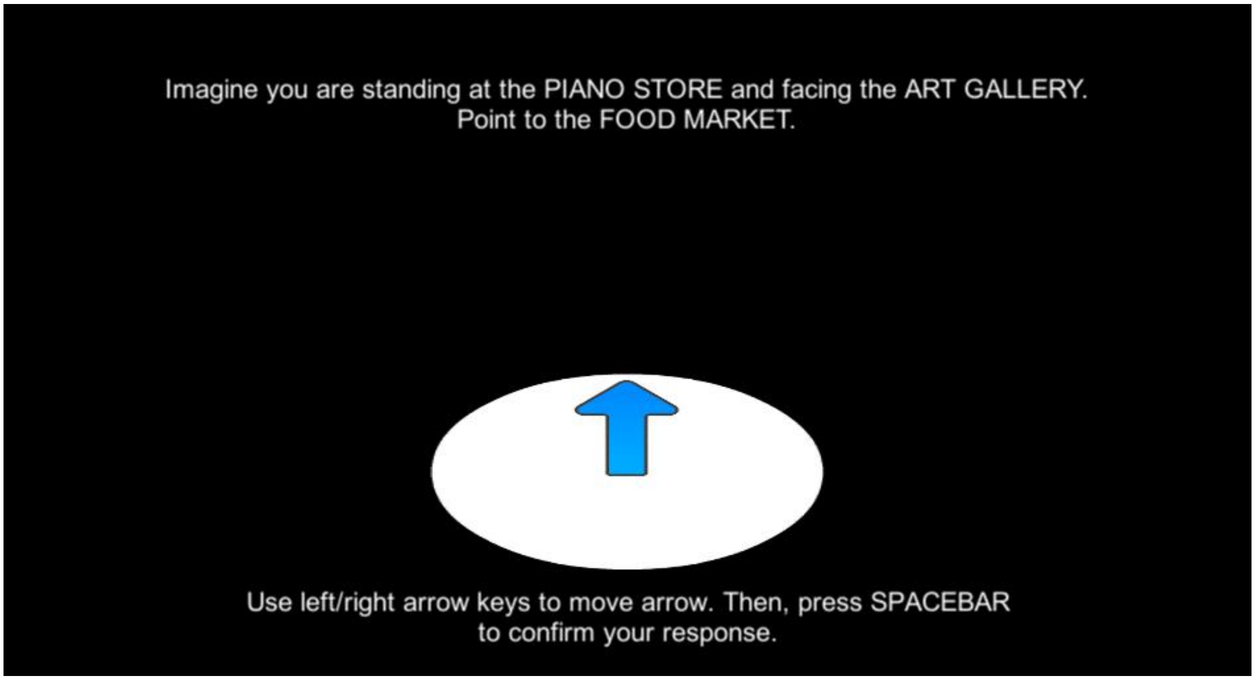
Participants

Participants were United States citizens (*n* = 127; 52.03% male; *M* age = 27.48, SD = 4.57) recruited through Prolific and compensated monetarily for participation. Power analyses showed that a sample size of 120 would be sufficient to detect relatively weak to moderate correlations (Correlation: H1 $\rho = .25$, $\alpha = .05$, Power = .80) (Faul et al., 2009). Regarding race, 64.57% identified as white/Caucasian, 17.32% identified as black/African American, 3.15% identified as Asian, 11.02% identified as biracial, 3.15% identified as Hispanic, and 0.79% identified as Native American. Participants were categorized as having completed less than a high school education (1.57%), completed high school or GED equivalent (27.56%), completed or attended college (61.42%), or completed or attended graduate school (9.45%). Participants completed the experiment on a laptop (63.11%) or desktop (36.89%) computer.

Materials

Only the original environment was used in Experiment 2. In addition to the wayfinding and errand running tasks, several other measures of individual differences were used:

The Judgments of Relative Direction Task (JRD) is a measure of survey knowledge that requires participants to imagine standing at one location and facing another, then point to a third location (McNamara, 2003). Each trial presented participants with a prompt (e.g., “Imagine standing at the COFFEE SHOP and facing the TOY STORE. Point to the BUTCHER SHOP”), a dial with an arrow that could be rotated around the center of the dial by pressing the left and right arrow keys, and instructions for how to rotate the arrow and submit a response (Figure 7). Responses were submitted by pressing the SPACEBAR key. Text prompts were in white font against a black background. The dial and arrow were white and blue, respectively, and the dial was rotated around the x axis by 30 degrees to provide depth. The JRD task comprised 20 trials, 12 of which featured unique storefronts as the imagined standing position such that all storefronts served as the imagined standing position at least once and were identical for all participants. Instructions and an example were presented, and participants were given an opportunity to practice rotating the arrow around the dial prior to beginning the task. Participants were also informed how many trials remained before each trial. Angular error and response time were the primary dependent measures. The JRD task has shown high construct validity, correlating highly with other measures of survey knowledge, (e.g., map drawing), improves with training, is correlated with confidence ratings, and has demonstrated high test-retest reliability ($r = .83$) (Huffman & Ekstrom, 2018).



Imagine you are standing at the PIANO STORE and facing the ART GALLERY.
Point to the FOOD MARKET.

Use left/right arrow keys to move arrow. Then, press SPACEBAR
to confirm your response.

Figure 7. Example trial from JRD task. Participants were presented with a storefront to imagine standing at, another to imagine facing, and a third to point to. The left and right arrow keys were used to rotate the arrow around the dial. Participants pressed the SPACEBAR key to confirm their response.

The Route-Finding Task asked participants to travel from one storefront to another as quickly as possible. At the beginning of each trial (*n* = 12), participants were teleported in front of and facing the starting storefront. Just like during the errand running task, the name of the target storefront was presented on the screen throughout the duration of the trial, but how much time had elapsed was not shown. The participant was instructed to find the target storefront using the shortest possible route. However, participants had as much time as needed to find the target and were not required to wait at targets upon arrival. Trials were terminated once the participant arrived at the target and pressed the SPACEBAR key while standing in front.

Participants were presented with instructions at the beginning of the task and were informed how many trials remained prior to each trial. All storefronts served as a target once.

The Three-Figure Mental Rotation Test (3F-MRT; adapted from Jost & Jahnsen, 2020; see also Peters & Battista, 2008) was used to measure mental rotation ability (Figure 8). Each trial consisted of two images on the left and right sides of the screen constructed from individual cubes and a third image in the bottom-middle of the screen that was a mirror of either the left or right image and a rotation of the other. Participants were instructed to choose which of the two images, left or right, was a rotation of the bottom-middle image. Participants were given one minute to answer as many items as possible and were awarded one point for correct answers and rescinded one point for incorrect answers. Text and images were presented against a black screen. Participants selected their answer by clicking the toggle box associated with the desired response and pressing the submit button. The trial did not terminate unless an answer choice was made. Trials were presented to all participants in the same order. Accuracy and response time were the primary dependent measures. Participants were given instructions and completed a practice trial before beginning the scored trials. To my knowledge, reliability measures have not been reported for the 3F-MRT. However, a similar and widely used mental rotation test by Vandenberg & Kuse (1978) has demonstrated high test-retest reliability (*r* = .83). Because the number of completed trials varied by participant, Cronbach's alpha was not computed here.

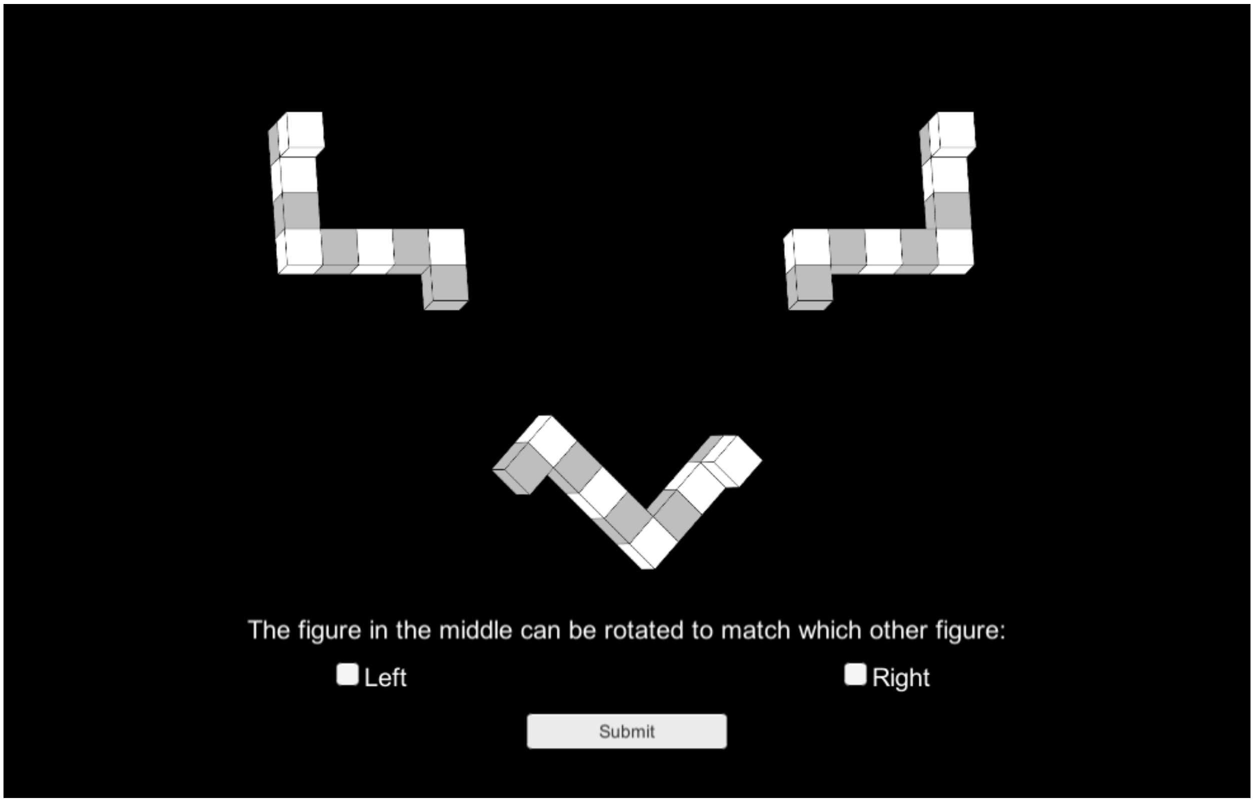

The figure in the middle can be rotated to match which other figure:

Figure 8. Example trial from the 3F-MRT. Participants indicated which of the left or right figures the middle figure was a rotation of. In this example, the right figure is the correct answer. Participants submitted their responses by pressing the submit button.

The Corsi Block-Tapping Task (CBT; adapted from Corsi, 1972) was used to measure spatial working memory (Figure 9). During each of 16 trials, participants were presented with a set of blue blocks (squares). During the learning phase of each trial, several blocks would sequentially light up by turning yellow and then return to blue just prior to the next block lighting up. No block repeatedly lit up within a trial. The number of blocks that lit up per trial began at two, and then increased by one block every two trials until all nine blocks lit up in the final two trials. During the testing phase of each trial, all blocks had returned to their initial blue color and the participant was instructed to "Go", at which time their task was to click on the blocks in the same order in which they lit up. The trial terminated once the participant clicked the same number of blocks that had lit up during the learning phase of that trial. Trials were identical for all participants. Participants were presented with instructions and completed two practice trials before beginning the scored trials. Text and blocks were presented against a black background. Trials were matched for all participants. The primary dependent variable was the product score, which is the product of the number of correct trials and the maximum span correctly repeated. The CBT has demonstrated high discriminate validity against other measures of working memory such as the Digit-Span Task (Kessels, van den Berg Carla Ruis, & Brands, 2008).

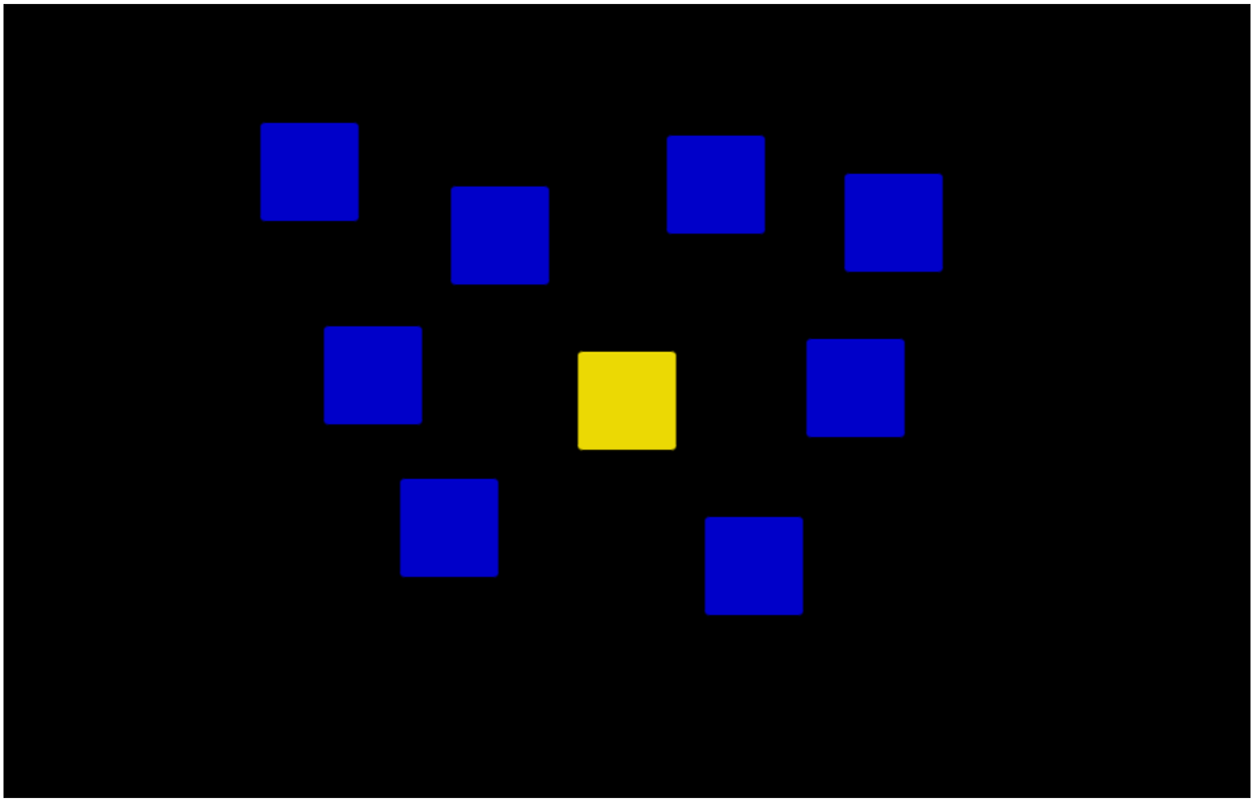


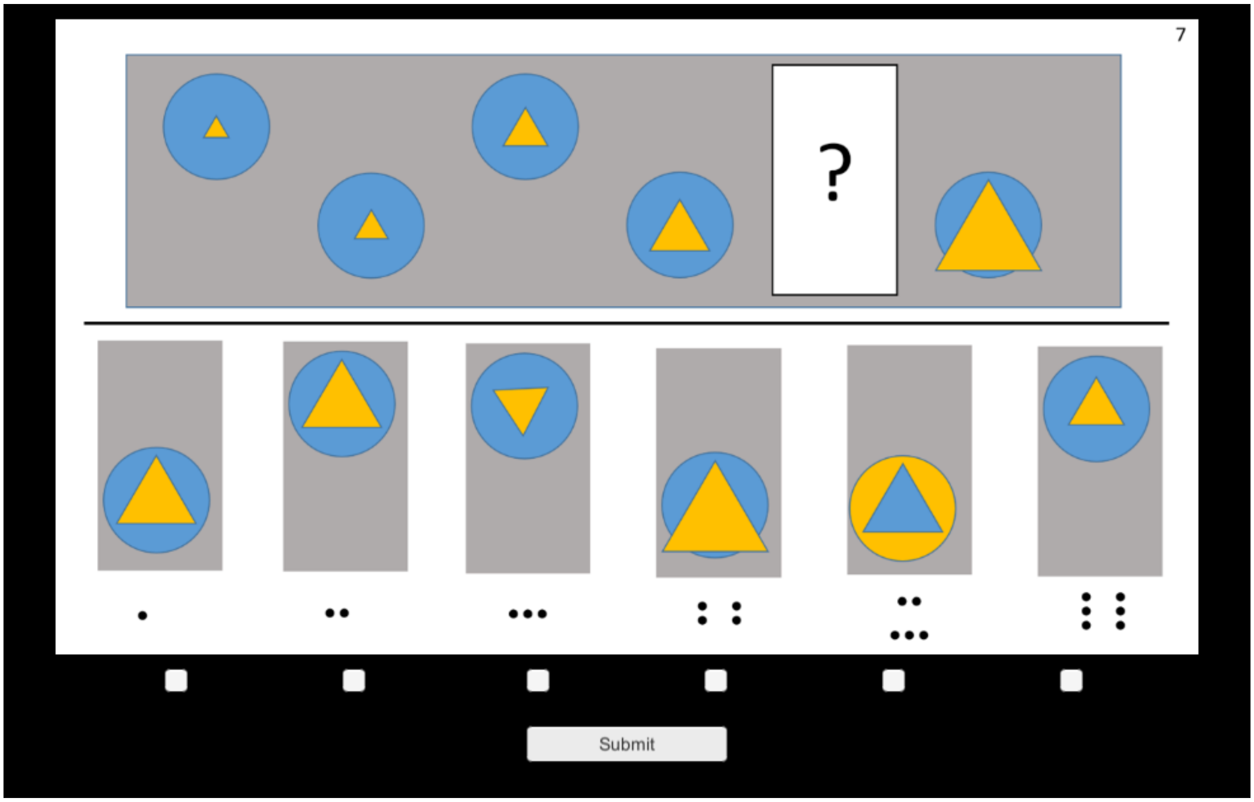
Figure 9. An image from an example trial from the CBT. Participants watched a sequence of blue squares lighting up. Following the sequence, a message appeared at the bottom of the

screen instructing participants to “Go”. Participants were tasked with clicking the squares in the same sequence as they originally lit up. The trial terminated after the same number of squares that lit up was clicked.

The Santa Barbara Sense of Direction Scale (SBSOD; Hegarty et al., 2020) consisted of 15 self-report items (e.g., “I very easily get lost in a new city”) on a 7-point Likert scale, ranging from 1 (Strongly Agree) to 7 (Strongly Disagree), which measured one’s self-reported navigation ability (Appendix B). Higher scores reflect greater navigation ability (positively worded items were reverse coded). The SBSOD was administered on a gray GUI against a black background. Participants were allowed to leave items blank and submitted their responses all at once by clicking a submit button at the bottom of the GUI. The SBSOD has shown high test-retest reliability (r = .91) (Hegarty et al., 2020).

The Wayfinding Anxiety Scale (WAS; Lawton & Kallai, 2002) consisted of 8 self-report items (e.g., “Finding my way to an appointment in an unfamiliar area of a city or town”) on a 5-point Likert scale, ranging from 1 (Not at all anxious) to 5 (Very anxious), in which participants rated their anxiety for various wayfinding scenarios (Appendix D). The WAS was administered on a gray GUI against a black background. Participants were allowed to leave items blank and submitted their responses all at once by clicking a submit button at the bottom of the GUI.

The Matrix Matching Test (MMT; Pluck, 2019) is a brief 24-item measure of intelligence (Figure 10). Each of visuospatial and semantic reasoning were measured by two 12-item subsets (visuospatial first), with items in each subset progressing in difficulty. For both subsets, a series of images was shown with one image (or two in the final half of the semantic reasoning subset) left blank and the participant’s task was to choose from several options which image most logically completed the series. Participants were required to make the appropriate number of responses before they could submit their answer. Instructions and a practice trial were presented before each subset and before the two-answer items in the verbal reasoning subset. The MMT correlates highly with the WAIS-IV full-scale IQ score (*r* = .89) and has demonstrated high test-retest reliability (*r* = .93) (Pluck, 2019).



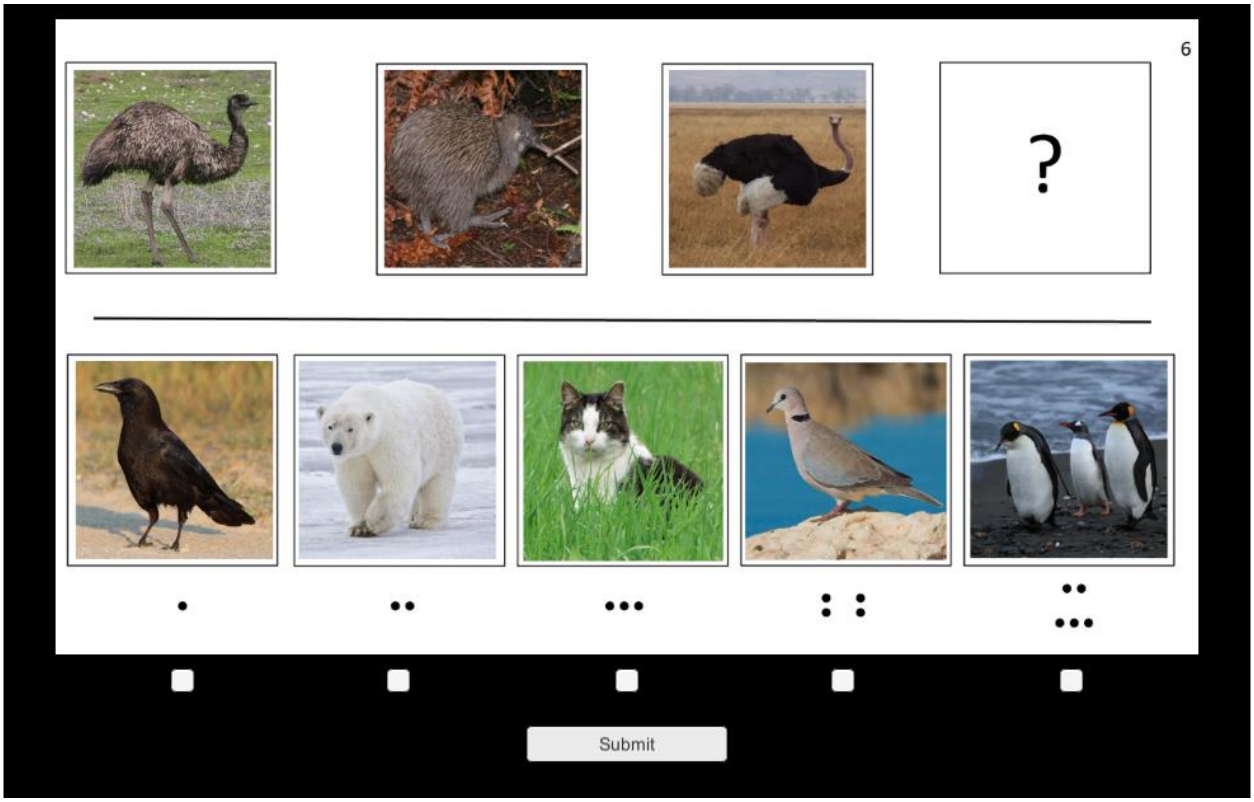


Figure 10. Example items from the MMT visuospatial (top) and semantic reasoning (bottom) subscales. Participants completed each trial by selecting the image that best completed the sequence and pressing the Submit button. The correct answers are images 2 and 5, respectively.

The Balloon Analogue Risk Task (BART; Lejuez et al., 2002) is a behavioral measure of risk taking (Figure 11). Participants completed 10 trials in which they were presented with a balloon and two options: pump or cash out. Each time the balloon was pumped, 5 cents was added to the potential reward for a given trial (although this value was not displayed to the participant until the trial ended), and a random integer from an array of integers 1-128 was sampled without replacement. Only when a value of 1 was sampled did the balloon pop, in which case the potential reward for that trial was voided. Thus, the probability of the balloon popping increased with each pump. The average number of pumps before the balloon popped is equal to the maximum value of the array (i.e., 128) divided by two; in this case, 64. If the participant decided to cash out, the potential reward was added to the sum across trials, which was always visible to the participant but was not updated until a trial was completed. The number of pumps before the balloon popped was the same for all participants across trials. The average number of pumps on trials in which the participant cashes out before popping the balloon was the primary dependent variable. This measure has been demonstrated to show high construct validity, correlating with self-report measures of risk-taking behaviors, as well as high discriminant validity against self-report measures of anxious and depressive symptoms (Lejuez et al., 2002). While the original BART comprised 30 trials, participants completed an abbreviated 10-trial version here. Results from Lejuez et al. (2002) suggest that participants' number of pumps were similar when trials were blocked in groups of 10. Because the pumping threshold to popping the balloon varied across trials, Cronbach's alpha was not computed here.

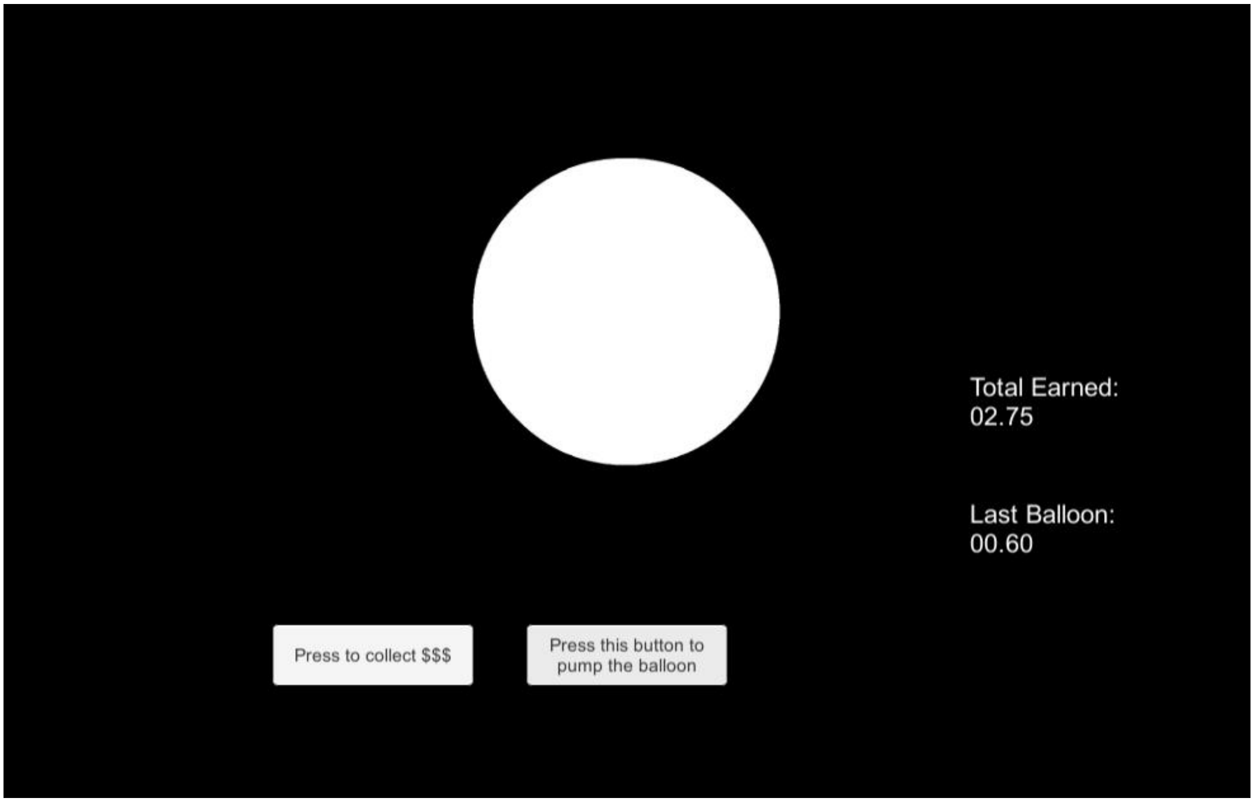


Figure 11. An example trial from the BART. Participants either pumped the balloon or collected the potential reward for that trial by pressing the respective button. If the balloon popped, the potential reward was lost for that trial and the next trial began. If they collected the potential reward, the trial also ended and the next trial began. The balloon increased in diameter with each pump.

The Revised Domain-Specific Risk-Taking (DOSPERT; Blais & Weber, 2006) consisted of 60 self-report items measuring both risk-taking behaviors (DOSPERT-RB) and risk perceptions (DOSPERT-RP) across five domains (i.e., financial, health/safety, recreational, ethical, and social) on a 7-point Likert scale, allowing for analysis of domain-specific scores for each of the risk-taking behaviors and risk perceptions subscales, as well as global scores within each subscale. Both the risk-taking behaviors and risk perceptions subscales had identical items.

In the risk-taking behaviors subscale, participants rated how likely it was that they would engage in the prompted behavior from 1 (Extremely Unlikely) to 7 (Extremely Likely). In the risk perceptions subscale, participants rated how risky they perceived the prompted behavior to be from 1 (Not at all Risky) to 7 (Extremely Risky). Both subscales were administered one at a time on a gray GUI against a black background. Participants were allowed to leave items blank and submitted their responses all at once by clicking a submit button at the bottom of the GUI. The test-retest reliability of the original full-length DOSPERT ranged from low to high depending on the domain and subscale (*min* r = .42, *max* r = .80) (Weber et al., 2002).

The Epistemic-Aleatory Rating Scale (EARS; Walters et al., 2022) consisted of 6 items on a 7-point Likert scale, ranging from 1 (Not at all) to 7 (Very much). Three items assessed perceptions of the epistemicness of the errand running task while the remaining 3 items assessed perceptions about aleatoriness (see Appendix E for the prompt). The EARS was administered on a gray GUI against a black background. Participants were allowed to leave items blank and submitted their responses all at once by clicking a submit button at the bottom of the GUI. The epistemic and aleatory subscales have been shown to consistently load onto separate factors (Walters et al., 2002).

Procedure

Experiment 2 took place across two sessions. Participants logged into the first session using their PID and completed the demographics questionnaires. Then, they completed the errand running task in the original environment as in the pilot experiments, as well as the JRD and Route-Finding tasks, in that order. After at least 24 hours had passed, participants logged into the second session using the PID they entered during the first session and completed the

SBSOD, DOSPERT, WAS, EARS, CBT, 3F-MRT, MMT, and BART, in that order. Participants were then debriefed and thanked for their participation.

Analyses

As in the previous experiments, Cronbach's alpha was computed for all 32 trials, as well as wide and narrow trials separately (Table 15). Due to server issues, one participant was missing a single wide trial and was removed from internal consistency analyses on all 32 trials as well as wide trial. Additionally, Cronbach's alpha was computed for several of the individual difference tasks (Table 15). Similarly, the same correlations between the dependent variables from the errand running task were computed.

Differences in average errand running time, average timeframe selection, average excess time, and number of successful trials between males and females and wide and narrow trials were analyzed by two-way Mixed ANOVAs, with gender and trial type as between- and within-subjects variables, respectively. Four participants did not report their gender and were removed from analyses. Means by trial type and gender are reported in Tables 17 and 18, respectively.

To examine the relationships between the dependent variables from the errand running task and the individual difference measures, Spearman's rank correlation coefficients were computed (see Table 16). Due to experimenter error, two participants were missing data from the WAS and MMT. These participants were not included in correlations involving these variables.

Hierarchical regression analyses were computed to assess the degree to which measures of spatial ability and environmental knowledge could predict errand running performance above and beyond general intelligence. Because the MMT comprises visuospatial and semantic reasoning subtests, measures of spatial ability (including the MMT visuospatial score, Corsi product score, and MRT score) were tested against MMT semantic reasoning score in separate regression models. Each model followed two steps: in the first step, MMT semantic reasoning score was input as the predictor; in the second step, the measure of spatial ability was input. Measures of environmental knowledge (average JRD angular error and average route-finding time) were tested against MMT total score in separate regression models. These models followed the same two steps, except that MMT total score was input as the predictor in the first step. Because the SBSOD is considered a measure of large-scale navigation ability rather than small-scale spatial ability (Hegarty et al., 2006), SBSOD score was tested against the MMT total score as well. Sandwich estimators were used to estimate the variability of regression coefficients and are robust to commonly violated assumptions of regression analyses, namely violations of homoscedasticity (Zeileis, 2004, 2006).

Finally, gender differences in survey (average JRD angular error) and route (average route-finding time) knowledge, spatial ability (MMT visuospatial score, Corsi product score, MRT score, and SBSOD), and risk-taking behaviors (DOSPERT-RB global score and BART score) and perceptions (DOSPERT-RP global score) were analyzed by separate one-way ANOVAs.

Results

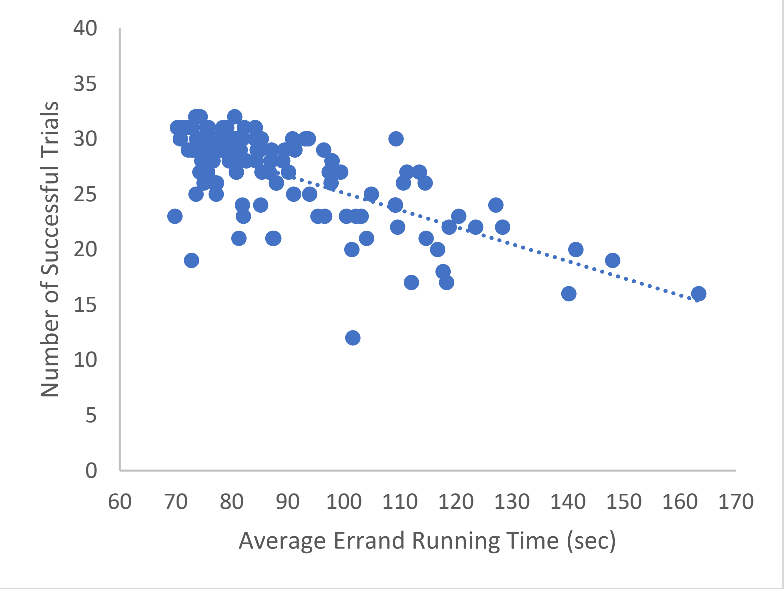
Cronbach’s alpha was high for errand running time and timeframe selection for all 32 trials, as well as wide and narrow trials separately (Table 15). Cronbach’s alpha for the individual difference tasks varied considerably.

	Cronbach’s Alpha
Errand Running Time All Trials	.93

Errand Running Time Wide Trials	.89
Errand Running Time Narrow Trials	.85
Timeframe Selection All Trials	.98
Timeframe Selection Wide Trials	.96
Timeframe Selection Narrow Trials	.97
JRD	.57
Route-Finding Task	.84
CBT	.63
SBSOD	.94
WAS	.91
MMT	.66
MMT Visuospatial	.57
MMT Semantic Reasoning	.52
DOSPERT-RB	.84
DOSPERT-RP	.85
EARS Epistemic	.71
EARS Aleatory	.75

Table 15. Cronbach's alpha for errand running time, timeframe selection, and individual difference tasks.

On average, participants completed every errand on the errand list and returned to the starting location within the selected timeframe on most trials (*M* = 26.70; *min* = 12, *max* = 32). There was a strong relationship between average errand running time and number of successful trials, such that participants who completed errands faster successfully completed errands within the selected timeframe more often on average (Figure 12). However, the relationship between average timeframe selection and number of successful trials was unreliable. There was also a strong relationship between average errand running time and average timeframe selection, suggesting that participants were able to predict the time required to complete errands (Figure 12). Participants tended to err on the side of caution, overestimating the timeframe required to complete errands by roughly 28 seconds on average (Table 16).



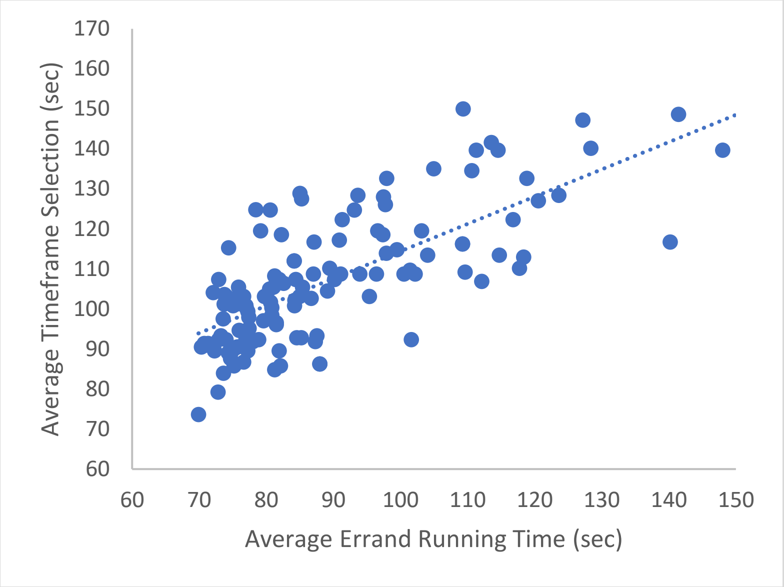


Figure 12. Top: The relationship between average errand running time and number of successful trials. Bottom: The relationship between average errand running time and average timeframe selection.

Nearly all participants completed the wayfinding task within 15 minutes during the first attempt (116 out of 127), with eight participants requiring two attempts and three participants requiring three attempts. There were relationships between average wayfinding time and all four dependent variables of the errand running task, such that, on average, participants who were faster during learning completed errands faster, selected shorter timeframes, had less excess time, and successfully completed errands within the selected timeframe more often.

	M (SD)	Skewness	Kurtosis	Min, Max	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.
1. Avg. Errand Running Time	89.70 (17.90)	1.52	5.37	[69.88, 163.40]	-																												
2. Avg. Timeframe Selection	107.41 (16.41)	0.64	2.85	[73.59, 150.00]	$.73^{***}$	-																											
3. Avg. Excess Time	27.92 (9.09)	0.64	2.58	[9.35, 49.31]	$.38^{***}$	$.84^{***}$	-																										
4. Number of Successful Trials	26.70 (4.04)	-1.12	3.86	[12, 32]	$-.60^{***}$	-.09	$.19^{*}$	-																									
5. Avg. Wayfinding Time	37.30 (16.68)	3.23	18.88	[17.49, 148.59]	$.63^{***}$	$.61^{***}$	$.38^{***}$	$-.28^{**}$	-																								
6. Avg. JRD Angular Error	87.08 (8.28)	0.47	3.40	[71.35, 113.51]	$.35^{***}$	.17	.01	$-.29^{***}$	$.22^{*}$	-																							
7. Avg. Route-Finding Time	20.77 (5.43)	1.44	4.87	[14.42, 39.88]	$.82^{***}$	$.59^{***}$	$.28^{**}$	$-.53^{***}$	$.60^{***}$	$.39^{***}$	-																						
8. SBSOD Score	3.95 (1.44)	-0.02	2.12	[1.00, 6.87]	$-.18^{*}$	$-.21^{*}$	-.09	.07	-.10	.03	$-.23^{*}$	-																					
9. DOSPERT-RB Global Score	2.91 (0.70)	0.78	3.93	[1.40, 5.57]	-.02	-.15	$-.20^{*}$	-.14	.00	.10	.08	.16	-																				
10. DOSPERT-RB Ethical Score	2.07 (0.95)	1.32	4.81	[1.00, 5.83]	.06	-.05	-.16	-.10	.03	.09	.14	-.16	$.60^{***}$	-																			

11. DOSPERT-RB Financial Score	2.74 (1.13)	0.81	3.28	[1.00, 6.33]	.08	-.05	-.09	-.15	.04	.18*	.17	.16	.66***	.33***	-									
12.DOSPERT-RB Health Score	2.44 (0.97)	0.75	3.49	[1.00, 5.50]	.05	-.12	-.22**	-.07	.05	.01	.11	-.01	.65***	.45***	.23**	-								
13. DOSPERT-RB Recreational Score	2.50 (1.24)	0.98	3.47	[1.00, 6.67]	-.14	-.24**	-.21*	-.07	.00	.05	-.08	.21*	.68***	.21*	.41***	.32***	-							
14. DOSPERT-RB Social Score	4.79 (0.98)	-0.18	2.83	[2.00, 7.00]	-.10	-.06	.02	-.02	-.05	-.05	-.07	.12	.55***	.15	.12	.22*	.25**	-						
15. DOSPERT-RP Global Score	4.40 (0.63)	-0.08	2.81	[2.70, 5.97]	.15	.22*	.26**	-.01	.18*	-.03	.10	-.18*	-.33***	-.35***	-.22*	-.23**	-.15	-.12	-					
16. DOSPERT-RP Ethical Score	4.58 (0.96)	0.02	2.91	[2.17, 7.00]	.09	.12	.15	-.02	.10	-.06	.02	-.05	-.29***	-.46***	-.12	-.23**	-.06	-.07	.78***	-				
17. DOSPERT-RP Financial Score	4.78 (1.00)	-0.43	3.19	[2.00, 7.00]	.10	.15	.20*	-.01	.19*	-.04	.04	-.06	-.29***	-.29***	-.36***	-.17	-.11	-.02	.70***	.46***	-			
18. DOSPERT-RP Health Score	5.13 (0.85)	-0.41	2.79	[2.83, 6.83]	.10	.19*	.15	-.02	.14	-.08	.00	-.23**	-.22*	-.27**	-.16	-.28**	-.04	.02	.69***	.54***	.39***	-		
19. DOSPERT-RP Recreational Score	4.58 (0.98)	-0.13	2.62	[2.00, 6.83]	.15	.20*	.21*	.03	.13	.02	.17	-.20*	-.32***	-.25**	-.11	-.19*	-.35***	-.16	.74***	.47***	.36***	.42***	-	
20. DOSPERT-RP Social Score	2.94 (0.85)	0.57	3.47	[5.83, 1.17]	.03	.04	.14	.03	.03	.09	.04	.01	.01	.11	.10	.05	.08	-.28**	.44***	.16	.15	.11	.23*	-

21. WAS Score	2.58 (0.98)	0.17	2.38	[1.00, 5.00]	.25**	.26**	.18*	-.02	.14	.03	.16	-.51***	-.25**	-.06	-.21*	-.10	-.21*	-.11	.44***	.27**	.23**	.35***	.37***	.33***	-								
22. EARS Epistemic Score	5.33 (1.12)	-0.67	3.57	[1.33, 7.00]	-.13	.00	.10	.24**	.01	-.10	-.13	-.05	-.17	-.22*	-.05	-.08	-.07	-.05	-.03	.00	.03	.02	-.10	-.07	-.09	-							
23. EARS Aleatory Score	4.82 (1.21)	-0.09	2.67	[1.67, 7.00]	-.14	-.21*	-.21*	-.01	-.17	.00	-.02	-.13	.06	.12	-.11	.05	-.01	.08	.05	-.03	.09	.09	.06	-.06	.01	-.13	-						
24. Corsi Product Score	75.55 (26.68)	0.33	2.39	[16, 135]	-.34***	-.34***	-.24**	.22*	-.26**	-.01	-.27**	.03	.08	-.08	-.06	.08	.17	.13	-.11	-.03	-.17	-.07	-.03	-.04	-.13	.22*	-.02	-					
25. MRT Score	2.60 (3.06)	0.12	3.74	[-7, 12]	-.18	.01	.14	.19*	-.22*	-.05	-.18*	-.03	-.7	-.01	-.02	-.06	-.05	-.01	-.14	-.11	-.18*	-.11	.06	-.13	-.04	-.01	.01	.08	-				
26. MMT Total Score	17.20 (2.95)	-0.06	3.02	[9, 24]	-.42***	-.33***	-.15	.34***	-.40***	-.11	-.42***	.10	-.03	-.10	-.10	-.02	-.03	.11	-.02	-.02	.04	.03	-.01	-.03	-.11	.05	.11	.29**	.12	-			
27. MMT Visuospatial Score	9.18 (1.72)	-0.45	3.04	[4, 12]	-.37***	-.35***	-.20*	.27**	-.37***	-.04	-.40***	.03	-.03	-.01	-.06	-.06	.04	.03	.04	.04	.05	.01	-.02	.03	-.08	.07	.18*	.29**	.03	.79***	-		
28. MMT Semantic Reasoning Score	8.02 (1.88)	-0.02	2.73	[3, 12]	-.29***	-.17	-.06	.27**	-.28**	-.09	-.30***	.12	-.03	-.13	-.09	.03	-.11	.11	-.09	-.07	-.03	.03	-.04	-.08	-.11	.01	.02	.18*	.15	.82***	.32***	-	
29. BART Score	155.46 (95.89)	1.20	4.26	[21, 455]	-.03	-.08	-.10	-.04	-.03	.00	.01	-.15	-.14	-.14	-.09	-.03	.01	-.12	.00	-.02	-.04	-.05	.05	-.01	.00	.15	-.15	.13	.11	.11	.06	.11	-

Table 16. Spearman correlation matrix for all dependent variables. * $p < .05$, ** $p < .01$, *** $p < .001$

Average errand running time was analyzed by a two-way Mixed ANOVA (see Analyses). The main effect of gender was significant (Table 18), $F(1,121) = 5.19$, $p = .024$, $\eta_p^2 = .04$. On average, males completed errands faster than females. The main effect of trial type was also significant (Table 17), $F(1,121) = 14.67$, $p < .001$, η_p^2 = .1l. On average, participants were faster during narrow trials than wide trials. The gender by trial type interaction was not significant, $p > .050$.

Average timeframe selection was analyzed by a two-way Mixed ANOVA. The main effect of gender was significant, $F(1,121) = 12.78$, $p < .001$, $\eta_p^2 = .10$. On average, males selected shorter timeframes than women. The main effect of trial type was also significant (Table 17), $F(1,121) = 20.19$, $p < .001$, $\eta_p^2 = .14$. On average, participants selected shorter timeframes during narrow trials than wide trials. The gender by trial type interaction was not significant, $p > .050$.

Average excess time was analyzed by a two-way Mixed ANOVA. The main effect of gender was significant, $F(1,121) = 8.07$, $p = .005$, $\eta_p^2 = .06$. On average, males had less excess time than females. Both the main effect of trial type and the gender by trial type interaction were not significant, *p*s > .050.

Number of successful trials was analyzed by a two-way Mixed ANOVA. The main effect of trial type was significant (Table 17), $F(1,121) = 18.23$, $p < .001$, $\eta_p^2 = .13$. On average, participants successfully completed more narrow trials than wide trials. Both the main effect of gender and the gender by trial type interaction were not significant, *p*s > .050.

	Wide	Narrow

Avg. Errand Running Time	91.86 (19.81)	88.27 (17.74)
Avg. Timeframe Selection	108.88 (16.49)	106.19 (17.40)
Avg. Excess Time	27.89 (9.14)	28.45 (9.16)
Avg. # of Successful Trials	12.85 (2.55)	13.73 (2.06)

Table 17. Displays means (and standard deviations) for dependent variables of the wayfinding and errand running tasks, as well as individual difference measures, for wide and narrow trials. Time is reported in seconds.

	Male	Female
Avg. Errand Running Time	86.56 (17.07)	93.87 (18.47)
Avg. Timeframe Selection	102.63 (15.82)	112.86 (15.92)
Avg. Excess Time	25.70 (8.49)	30.27 (9.29)
Avg. # of Successful Trials	26.50 (4.37)	26.68 (3.70)
Avg. Wayfinding Time	35.70 (19.87)	39.42 (12.77)
Avg. JRD Angular Error	87.99 (9.46)	86.19 (6.66)
Avg. Route-Finding Time	20.39 (5.31)	21.44 (6.27)
SBSOD Score	4.35 (1.51)	3.57 (1.26)
DOSPERT-RB Score	3.19 (0.72)	2.64 (0.56)
DOSPERT-RP Score	4.24 (0.64)	4.57 (0.57)
WAS Score	2.19 (0.95)	2.93 (0.93)
EARS Epistemic Score	5.31 (1.16)	5.31 (1.10)
EARS Aleatory Score	4.88 (1.14)	4.78 (1.26)
Corsi Product Score	77.25 (27.06)	73.73 (26.50)

MRT Score	2.67 (3.32)	2.41 (2.79)
MMT Total Score	17.14 (3.87)	17.29 (3.48)
MMT Visuospatial Score	9.27 (2.05)	9.14 (2.10)
MMT Semantic Reasoning Score	7.87 (2.24)	8.16 (2.05)
BART Score	155.53 (99.41)	160.46 (93.71)

Table 18. Displays means (and standard deviations) for dependent variables of the wayfinding and errand running tasks, as well as individual difference measures, for males and females. Time is reported in seconds.

Route and survey knowledge (average route-finding time and average JRD angular error, respectively) were both related to errand running performance (Table 16). Participants who made greater angular errors during the JRD task tended to have greater errand running times and successfully completed errands within the chosen timeframe less often. Route knowledge was much more strongly related to errand running performance. On average, participants with greater route-finding times also showed greater errand running times, greater timeframe selections, reduced excess times, and were less successful at completing errands within the selected timeframe.

Large-scale spatial ability (SBSOD score) was related to average errand running time and average timeframe selection (Table 16). On average participants with greater SBSOD scores showed reduced errand running times and selected shorter timeframes.

In general, both risk behaviors (DOSPERT-RB) and risk perceptions (DOSPER-RP) were related to decision-making during errand running (Table 16). On average, greater

DOSPERT-RB scores were associated with selecting shorter timeframes and reduced excess time. Additionally, on average, greater DOSPER-RP scores were associated with selecting longer timeframes and greater excess time. However, the relationship between BART scores and errand running performance was unreliable.

Wayfinding anxiety (WAS) was related to errand running performance (Table 16). On average, greater WAS scores were associated with greater errand running times, selection of longer timeframes, and greater excess times.

Perceptions of the types of uncertainty involved in errand running were related to errand running performance (Table 16). On average, participants who rated errand running as involving greater epistemic uncertainty (EARS epistemic score) were successful at completing errands within the selected timeframe less often. Furthermore, on average, participants who rated errand running as involving greater aleatory uncertainty (EARS aleatory score) selected shorter timeframes and had less excess time.

Spatial working memory (Corsi product score) was related to all four dependent variables of the errand running task (Table 16). On average, participants with greater Corsi product scores showed reduced errand running times, shorter timeframe selections, reduced excess time, and greater success at completing errands within the selected timeframe.

Mental rotation ability was weakly related to errand running performance, such that greater MRT scores were associated with reduced errand running time and greater success at completing errands within the selected timeframe.

General intelligence (MMT total score) was related to errand running performance (Table 16). Greater total scores on the MMT were associated with reduced errand running times, shorter timeframe selections, and greater success at completing errands within the selected timeframe. Additionally, greater scores on the visuospatial subscale of the MMT were associated with reduced errand running times, shorter timeframe selections, reduced excess times, and greater success at completing errands within the selected timeframe. Finally, greater scores on the semantic reasoning subscale of the MMT were associated with reduced errand running time and greater success at completing errands within the selected timeframe.

Hierarchical regression analyses were conducted to assess the contributions of the MMT semantic reasoning score, MMT visuospatial score, Corsi product score, and MRT score to predicting errand running performance (Table 19). In each analysis, the MMT semantic reasoning score was entered as a predictor in the first step. Alone, the MMT semantic reasoning score was a significant predictor of average errand running time, average timeframe selection, and number of successful trials. However, when the MMT visuospatial score was entered in step two, the MMT semantic reasoning score was no longer a significant predictor of any of the dependent variables of the errand running task. Furthermore, the MMT visuospatial score was a significant predictor of all four dependent variables of the errand running task. Similarly, when the Corsi product score was entered in step two, the MMT semantic reasoning score was only a significant predictor of average errand running time. The Corsi product score was a significant predictor of average errand running time, average timeframe selection, and average excess time. On the other hand, when the MRT score was entered in step two, the MMT semantic reasoning score remained a significant predictor of average errand running time and average timeframe selection. The MRT score was not a significant predictor of any dependent variable of the errand running task.

Hierarchical regression analyses were conducted to assess the contributions of the MMT total score, SBSOD score, average JRD angular error, and average timeframe selection to predicting errand running performance (Table 19). In each analysis, the MMT total score was entered as a predictor in the first step. Alone, the MMT total score was a significant predictor of all four dependent variables of the errand running task. When the SBSOD score was entered in step two, the MMT total score remained a significant predictor of all four dependent variables. The SBSOD score was not a significant predictor of any of the dependent variables. Similarly, the MMT total score remained a significant predictor of all four dependent variables when average JRD angular error was entered in step two. However, average JRD angular error was a significant predictor of average errand running time and number of successful trials. When average route-finding time was entered in step two, the MMT total score was no longer a significant predictor of any of the dependent variables. However, average route-finding time was a significant predictor of all four dependent variables of the errand running task.

			Avg. Errand Running Time				Avg. Timeframe Selection				Avg. Excess Time				Number of Successful Trials			
	Model	Predictor	b	β	R^2	ΔR^2	b	β	R^2	ΔR^2	b	β	R^2	ΔR^2	b	β	R^2	ΔR^2
MMT Visuospatial Score	1	(Intercept)	109.55				121.20				32.41				23.41			
		MMT Semantic Reasoning Score	-2.47	-0.26^{**}			-1.73	-0.20^{*}			-0.56	-0.12			0.41	0.19^{*}		
					$.07^{**}$				$.04^{*}$				.01				$.04^{*}$	
	2	(Intercept)	141.39				143.53				39.62				18.47			
		MMT Semantic Reasoning Score	-0.93	-0.09			-0.64	-0.07			-0.21	-0.04			0.17	0.08		
		MMT Visuospatial Score	-4.81	-0.46^{***}			-3.38	-0.35^{***}			-1.09	-0.21^{*}			0.75	0.32^{***}		
					$.25^{***}$	$.18^{***}$			$.15^{***}$	$.11^{***}$			$.05^{*}$	$.04^{*}$			$.12^{***}$	$.08^{***}$
Corsi Product Score	1	(Intercept)	109.55				121.20				32.41				23.41			
		MMT Semantic Reasoning Score	-2.47	-0.26^{**}			-1.73	-0.20^{*}			-0.56	-0.12			0.41	0.19^{*}		
					$.07^{**}$				$.04^{*}$				.01				$.04^{*}$	
	2	(Intercept)	120.64				132.11				36.93				22.14			
		MMT Semantic Reasoning Score	-1.94	-0.20^{*}			-1.21	-0.14			-0.35	-0.07			0.34	0.16		
		Corsi Product Score	-0.20	-0.30^{**}			-0.20	-0.33^{***}			-0.08	-0.24^{**}			0.02	0.15		
					$.15^{***}$	$.08^{***}$			$.14^{***}$	$.10^{***}$			0.07^{**}	0.06^{**}			$.06^{*}$	.02
MRT Score	1	(Intercept)	109.55				121.20				32.41				23.41			
		MMT Semantic Reasoning Score	-2.47	-0.26^{**}			-1.73	-0.20^{*}			-0.56	-0.12			0.41	0.19^{*}		
					$.07^{**}$				$.04^{*}$				.01				$.04^{*}$	
	2	(Intercept)	109.82				120.96				32.09				23.27			
		MMT Semantic Reasoning Score	-2.39	-0.25^{**}			-1.80	-0.21^{*}			-0.56	-0.12			0.37	0.17		
		MRT Score	-0.36	-0.06			0.31	0.06			0.42	0.14			0.18	0.14		
					$.07^{**}$	.00			$.04^{*}$	.00			.03	.02			$.05^{*}$	.01
SBSOD	1	(Intercept)	136.98				140.44				38.63				19.13			
		MMT Total Score	-2.75	-0.45^{***}			-1.92	-0.34^{***}			-0.62	-0.20^{*}			0.44	0.32^{***}		
					$.20^{***}$				$.12^{***}$				$.04^{*}$				$.10^{***}$	

	2	(Intercept)	140.28				145.93				39.82				19.41			
		MMT Total Score	-2.70	-.44***			-1.84	-0.33**			-0.61	-0.20*			0.44	0.32***		
		SBSOD Score	-1.05	-0.08			-1.75	-0.15			-0.38	-0.06			-0.09	-0.03		
					.21***	.01			.14***	.02			.04*	.00			.10***	.00
Avg. JRD Angular Error	1	(Intercept)	136.98				140.44				38.63				19.13			
		MMT Total Score	-2.75	-0.45***			-1.92	-0.34***			-0.62	-0.20*			0.44	0.32***		
					.20***				.12***				.04*				.10***	
	2	(Intercept)	83.10				118.04				42.75				30.60			
		MMT Total Score	-2.59	-0.42***			-1.86	-0.34**			-0.64	-0.21*			0.41	0.30***		
		Avg. JRD Angular Error	0.59	0.27***			0.24	0.12			-0.04	-0.04			-0.13	-0.26***		
					.28***	.08***			.14***	.02			.04*	.00			.17***	.07**
Avg. Route-Finding Time	1	(Intercept)	136.98				140.44				38.63				19.13			
		MMT Total Score	-2.75	-0.45***			-1.92	-0.34***			-0.62	-0.20*			0.44	0.32***		
					.20***				.12***				.04*				.10***	
	2	(Intercept)	48.67				85.97				26.78				33.44			
		MMT Total Score	-0.65	-0.11			-0.70	-0.12			-0.40	-0.13			0.09	0.07		
		Avg. Route-Finding Time	2.50	0.78***			1.61	0.53***			0.39	0.23**			-0.40	-0.55***		
					.69***	.49***			.35***	.23***			.10**	.06**			.34***	.24***

Table 19. Hierarchical linear models predicting average errand running time, average timeframe selection, average excess time, and number of successful trials. The far-left column identifies each model by the variable added in the second step of each model. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

Gender differences in survey (average JRD angular error) and route (average route-finding time) knowledge were analyzed by separate one-way ANOVAs. The mean difference in average JRD angular error between males and females was not significant, nor was the mean difference in average route-finding time, *p*s > .050.

Gender differences in spatial abilities were analyzed by separate one-way ANOVAs. On average, males rated themselves higher in sense of direction (SBSOD score) than females, $F(1,121) = 9.66$, $p = .002$, $\eta_p^2 = .07$. Similarly, females rated themselves higher in wayfinding anxiety (WAS score) than males, $F(1,121) = 20.93$, $p < .001$, $\eta_p^2 = .15$. However, the males and females did not differ significantly in spatial working memory (Corsi product score), mental rotation ability (MRT score), or visuospatial ability (MMT visuospatial score), *p*s > .050.

Gender differences in risk-taking behaviors and perceptions were analyzed by separate one-way ANOVAs. Males rated themselves as more likely to engage in risk-taking behaviors (DOSPERT-RB score) than females, $F(1,121) = 22.372$, $p < .001$, $\eta_p^2 = .16$. Furthermore, females perceived the scenarios presented in the DOSPERT-RP as being riskier than did males, $F(1,121) = 9.13$, $p = .003$, $\eta_p^2 = .07$. However, BART scores were similar for males and females, $p > .050$.

Discussion

Once again, the errand running task showed high internal consistency. Furthermore, many of the correlations between the dependent variables of the errand running task were consistent with the previous experiments. While average wayfinding time was highly correlated with average errand running time and average timeframe selection, there was still plenty of unexplained variance left. As in Experiment 1, males required less errand running time, selected shorter timeframes, and had less excess time than females, but there was no difference in number of successful trials. There was also a trend for better performance on narrow trials (i.e., less errand running time and shorter timeframe selections), however, these findings are likely attributed to high power as the mean differences were only by a few seconds. Overall, these results mirror those of the previous experiments. However, the purpose of Experiment 2 was to test for relationships between errand running performance and environmental knowledge, spatial ability, risk behaviors and perceptions, and perceptions of uncertainty.

Concerning environmental knowledge, route knowledge was highly correlated with errand running performance, particularly average errand running time, timeframe selection, and number of successful trials. This is not surprising given that the route-finding task asked participants to go from one storefront to another using the fastest possible route. Participants with a strong understanding of the configural network of storefronts are surely going to plan a more optimal route to visiting multiple storefronts on an errand list and have an easier time estimating the time requirements of completing the errands. On the other hand, survey knowledge, as measured by average JRD angular error, was much less correlated with errand running performance. One explanation is that the errand running task involves learning the environment and completing errands by traveling routes. Survey knowledge can be enhanced by having participants learn an environment through survey representations or navigating in straight lines (He et al., 2018). However, participants had little exposure to such learning here, which was consistent with near-chance JRD performance. Another explanation is that the JRD task showed poor internal consistency, and thus the task, at least as implemented here, may have been unreliable.

Concerning spatial abilities, large-scale spatial ability (SBSOD score) was only weakly correlated with average errand running time average timeframe selection. However, it was not a significant predictor of either when both SBSOD score and MMT total score were included as predictors in a hierarchical regression model, suggesting that the relationship between SBSOD score and errand running performance may have been due to individual differences in general intelligence. Similarly, small-scale spatial ability (MRT score) was only weakly correlated with average errand running time and number of successful trials but was not a significant predictor in a hierarchical regression with semantic reasoning score. On the other hand, spatial working memory and visuospatial ability were correlated with all four dependent variables of the errand running task and continued to be significant predictors above and beyond MMT semantic reasoning score, suggesting that the spatial ability components of the CBT and MMT visuospatial subscale were particularly relevant to errand running performance. However, it should be noted that the internal consistencies of Corsi product score and MMT visuospatial score were relatively weak.

Concerning risk behaviors and perceptions, BART score was not correlated with errand running performance, but DOSPERT scores (global and some subscales) were correlated with average timeframe selection and excess time. While correlations between DOSPERT-RB scores and errand running performance were relatively weak, they did suggest that individuals who were more likely to engage in risky behaviors selected shorter timeframes on average and had less excess time on successful trials on average. In other words, individuals who were more likely to engage in risky behaviors (particularly in recreational domains) made riskier timeframe estimates to complete errands. Similarly, correlations between DOSPERT-RP scores and errand running performance suggested that individuals who perceived various scenarios presented in the DOSPERT-RP subscale as riskier selected longer timeframes on average and had more excess time on successful trials on average. In other words, individuals who perceived higher risk on the DOSPERT-RP made less risky choices during errand running. Furthermore, participants who had greater anxiety about wayfinding (WAS score) also selected longer timeframes on average and had more excess time on successful trials on average, in addition to greater errand running time on average.

Interestingly, while females took longer to run errands than males on average, there were no gender differences in average route-finding time or average JRD angular error, suggesting that males and females had similar levels of environmental knowledge, both in terms of route and survey knowledge. Completing errands in a faster time yields greater rewards given that a shorter timeframe was also selected. However, because females selected longer timeframes on average, they may have felt less pressure to use the shortest possible routes when running errands, which would explain why there was a gender difference in average errand running time but not average route-finding time. It is possible that males and females would show similar errand running times if a short timeframe was imposed upon them, as opposed to giving participants agency over timeframe selection.

While males and females did not differ in environmental knowledge or spatial ability on average, males rated themselves higher in sense of direction (SBSOD score) and lower in wayfinding anxiety (WAS score) than females, which is in line with previous findings (Lawton & Kallai, 2002). Additionally, females rated themselves as less likely to engage in risky behaviors (DOSPERT-RB score) and perceived scenarios as riskier (DOSPERT-RP scores) than males on average. Thus, to the previous point, it is possible that females gave themselves longer timeframes to run errands, not because they are slower at running errands per se, but because they attributed greater risk to shorter timeframes and potentially arriving late to the destination than males. Consequently, there is less pressure to run errands using the shortest possible routes because the selected timeframe permits suboptimal navigation. However, when arbitrary timeframes are imposed on navigators, as in the route-finding task, by telling participants to use the shortest possible routes, males and females were both equally capable of navigating more efficiently.

Concerning perceptions of uncertainty, participants who rated the errand running task as involving greater epistemic uncertainty, that is, the time required to run a set of errands as knowable in advance, were successful on more trials. On the other hand, participants who rated the errand running task as involving more aleatory uncertainty, that is, the time required to run a set of errands as described by a random process, selected shorter timeframes and had less excess time on successful trials on average. This finding is seemingly contradictory as one might expect navigators would give themselves more time to account for randomness. However, in the current task participants were awarded more money for completing trials within shorter timeframes. Thus, participants may have thought that, if the time to run the errands involved random factors (e.g., waiting times), they might as well hedge their bets on shorter timeframes to earn more reward.

Overall, the errand running task showed convergent validity with environmental knowledge, particularly route knowledge, spatial ability, particularly spatial working memory and visuospatial ability, and risk behaviors and perceptions. Furthermore, spatial abilities went above and beyond semantic reasoning to predict errand running performance, with semantic reasoning often no longer being a significant predictor after spatial ability measures were added as predictors. However, a concern about the present study is that several of the individual difference measures showed low internal consistency (see General Discussion). Males and females showed mean differences in errand running performance, risk behaviors and perceptions, as well as anxiety about wayfinding. However, differences errand running ability may be attributable to differences in these latter variables as both males and females showed comparable route-finding times on average.

CHAPTER 5

GENERAL DISCUSSION

The purpose of this dissertation was to study navigational decision-making in ubiquitous complex navigational tasks such as errand running, as well as to establish the reliability and validity of an experimental paradigm for studying errand running behavior. To this end, two pilot experiments were conducted to assess the internal consistency of the errand running task and to ensure that participants were able to comprehend the task and complete it remotely. Second, the test-retest reliability of the errand running task was assessed by having participants complete the errand running task across two sessions on different days, across two different environments. Finally, construct validity was assessed by examining the relationship between errand running performance and environmental knowledge, spatial ability, and risk-taking behaviors and perceptions. Gender differences and perceptions of uncertainty surrounding errand running were also examined.

Participants understood the instructions of the errand running task and had little difficulty completing the experiment remotely. Across all experiments, a large majority of participants completed the wayfinding task in a single attempt and correlations between average errand running time and average timeframe selection were high. Additionally, Cronbach's alpha suggested that the errand running task had very high internal consistency, and this was true for both wide and narrow trials alike. One concern following Pilot Experiment 1 was the high correlation between average errand running time and average wayfinding time. However, the later experiments with larger sample sizes confirmed that there was still much variance in errand running performance that was unaccounted for by average wayfinding time. Thus, the errand running task is separable from standard wayfinding paradigms used in past research and reliably measures a common construct across trials.

Experiment 1 demonstrated high test-retest reliability of consistency and agreement of the errand running task. First, ICC analyses on the relationship between errand running performances across sessions were greater than .80 for all four of the primary dependent measures of the errand running task. Secondly, while some significant differences between environments emerged, mean errand running performance was relatively similar across the original and mirror environments. One possible explanation for differences in errand running performance across environments is that the relative starting position was also different for each environment. While the relative locations of the storefronts across trials were identical in each storefront, trials did not conclude until the participant had successfully returned to the starting location. Thus, it is possible that a change in the starting location may have had a systematic impact on the time to complete errands on some trials. This would explain why there was an interaction between trial type and environment for average errand running time. Finally, the correlation matrices of the relationships between the dependent variables of the errand running task were also similar across environments. Overall, the results of Experiment 1 painted a promising picture for the test-retest reliability of the errand running task.

Experiment 2 assessed the construct validity of the errand running task by examining the relationship between errand running performance and environmental knowledge, spatial ability, and risk-taking behaviors and perceptions, which I will discuss in turn. Beginning with environmental knowledge, errand running performance and route knowledge of the environment were strongly linked. Average route-finding time was strongly associated with average errand running time, average timeframe selection, and number of successful trials, and to a weaker extent, average excess time. This relationship is not surprising as the route-finding task asked participants to navigate from one storefront to another using the fastest possible route. During the errand running task, participants are indirectly encouraged to complete the errands within the shortest timeframe possible to maximize their rewards.

On the other hand, errand running performance and survey knowledge were not as strongly linked, showing correlations between average JRD angular error and average errand running time and number of successful trials. One reason why survey knowledge was not as strongly associated with errand running performance is because participants interacted with the environment during learning (wayfinding) and testing (errand running) by traveling along routes. While these routes were not predetermined, that is, participants determined their own routes to navigation during both learning and testing, survey knowledge is more tightly linked to learning and testing knowledge of environments via maps and straight-line distance estimates (He et al., 2002; Thorndyke & Hayes-Roth, 1982). In other words, performance on the route-finding task had an advantage over performance on the JRD task at predicting errand running performance because of an alignment in task demands. Another possibility, which is not mutually exclusive, may be due to the poor internal consistency of the JRD task. Of all the measures analyzed in this study, the JRD task had the lowest Cronbach’s alpha. Additionally, performance on the JRD task was near chance, which may have limited its ability to predict errand running performance. Despite these concerns, however, average JRD angular error was correlated with errand running performance, suggesting that survey knowledge may play a role in errand running behavior.

Spatial ability was also tightly linked to errand running performance, which was correlated with performance on the CBT and MMT visuospatial subtest. Sense of direction and mental rotation ability were also correlated with errand running performance, but to a weaker extent. Furthermore, neither sense of direction nor mental rotation ability were able to predict errand running performance after accounting for general intelligence and semantic reasoning ability, respectively. As for sense of direction, the SBSOD has been shown to be a poor predictor of environmental learning of virtual environments (Hegarty et al., 2006), and thus may only be a predictor of errand running performance in naturalistic settings. It is not clear why mental rotation ability was not correlated with errand running performance. Past research has shown significant correlations between mental rotation ability and wayfinding performance; however, these correlations were relatively weak (Malinowski, 2001). On the other hand, performance on the CBT and MMT visuospatial subtests predicted errand running performance above and beyond semantic reasoning ability, which was no longer a predictor of errand running performance after accounting for performance on the CBT and MMT visuospatial subtests. These are promising results as they demonstrate that the errand running task has a spatial component that cannot be accounted for by other measures of general intelligence (semantic reasoning).

Risk-taking behaviors and perceptions were also linked to errand running performance; however, perceptions of risk were more strongly linked than risk behaviors. DOSPERT-RP global scores, as well as several subscale scores, were associated with average timeframe selection and average excess time, suggesting that participants who perceived greater risk in the various scenarios of the DOSPERT-RP tended to pick safer timeframe options at the expense of greater monetary reward. In other words, these participants were willing to forgo a greater potential reward to avoid losing the reward entirely by arriving late to the final destination (starting location). While the DOSPERT-RB was not as strongly linked to errand running performance, there were some negative associations with average timeframe selection and average excess time, suggesting that participants who rated themselves as more likely to engage in risky behaviors made riskier timeframe decisions. Interestingly, for both risk behaviors and perceptions, the recreational subscales had the strongest associations with errand running performance. One possibility is that the virtual nature of the errand running task resembled a videogame-like experience and felt recreational to participants. In sum, the results of the DOSPERT scales suggested that risk-taking behavior and perceptions were associated with the decision-making strategies of participants.

However, the BART score, which is a behavioral measure of risk-taking, showed no link to errand running performance. One possibility is that the outcome of the BART is highly aleatory in nature. That is, while the probability that the balloon will pop increases with every pump, one cannot predict when the balloon will pop. On the other hand, while there is some aleatoriness associated with the errand running task regarding the random waiting time at each storefront, confidence in one's navigation abilities paired with knowledge of the environment adds a degree of epistemic certainty to decision-making about timeframe estimates. Thus, the nature of the risk involved in the BART and errand running task are quite different. Furthermore, it is still unclear what risky decision-making construct the BART measures. Performance on the BART and other gambling tasks have shown to load onto different constructs (Buelow & Blaine, 2015).

In a similar vein, wayfinding anxiety was linked to errand running performance, such that participants who rated themselves as having greater anxiety about navigation took longer to run errands, selected longer timeframes, and had more excess time, on average. These results suggest that the reason for anxiety about wayfinding may be in part due to weaker navigation ability. The latter two results suggest that wayfinding anxiety is also related to the decision-making process during timeframe estimation insofar as those with higher wayfinding anxiety overestimate how long it will take to complete errands to minimize losses incurred by arriving late. Overall, wayfinding anxiety appears to play an important role in errand running behavior, both regarding navigation time and decision-making.

Perceptions of uncertainty about running errands were also linked to errand running behavior. Greater perceptions of epistemic uncertainty were associated with a greater number of successful trials. However, given that the EARS was administered in the second session, it is unclear whether perceiving the time required to run errands as knowable in advance led participants to select more accurate timeframes, or that successfully completing more trials led participants to believe the time required to run errands was knowable in advance. Thus, the nature of the relationship between perceptions about epistemic uncertainty and successful errand running is unclear. Perceptions about aleatoriness were associated with smaller timeframes and less excess time. This finding is confusing as one might expect navigators to overcompensate with longer timeframes to account for random factors interfering with successful errand running. However, it is possible that navigators decided to hedge their bets on greater rewards per trial considering that random factors could reduce their chances of success anyway. Overall, the link between epistemic and aleatory uncertainty and errand running is unclear from the results of this study.

Overall, there was little evidence that the uncertainty about waiting times (narrow vs. wide trials) affected decision-making. Although some significant effects were observed, mean differences between narrow and wide trials were small and likely significant due to large sample sizes. It is possible that differences in mean waiting time between storefronts would have a larger impact on decision-making, as the manipulation here was of the range around a constant mean of 10 seconds. Thus, on average, navigators would expect the waiting time to be the same across the narrow and wide conditions.

On the other hand, gender differences in errand running performance were large and robust. Males tended to take less time to complete errands, select shorter timeframes, and have less excess time, on average, than females. However, the reason for these differences is unclear, as there was not a gender difference in route-finding time, which tasked participants with wayfinding from one location to another using the quickest possible route. While there was a monetary benefit to selecting shorter timeframes and completing errands faster, females may have been less inclined to select shorter timeframes for a couple of reasons. Firstly, as found in previous research, females reported higher levels of wayfinding anxiety (Lawton & Kallai, 2002). Secondly, females reported higher perceptions of risk and less likelihood of engaging in risky behaviors, which has also been supported by previous research (Harris & Jenkins, 2006). Lastly, females rated their sense of direction lower than males, which has also been reported in previous research (Hegarty et al., 2006). Thus, women may have selected longer timeframes in response to greater anxiety, avoidance of risk, a lower perception of their navigational competency, or a combination of these factors. However, there were no gender differences in any of the spatial ability measures reported here, which paired with the lack of a gender difference in average route-finding time, suggests that males and females were equally capable of errand running performance as measured in the current paradigm.

Future Directions

The errand running task presented here has shown strong internal consistency, reliability, and validity. While wayfinding tasks have been used extensively to study knowledge of spatial environments, decision-making during navigation is a ripe area for future research. This errand running task may be useful in further understanding how navigators make decisions under risk and uncertainty. Here, a few directions for future research are discussed.

As mentioned, errand running performance and decision-making was similar across narrow and wide trials. One possibility is the mean waiting time was constant across trial types, but the variability was manipulated instead. In other words, the average waiting time was equal across narrow and wide trials. Future studies could manipulate both the average waiting time and waiting time range to see whether there is an interaction between these two variables. It is likely that participants would be sensitive to large discrepancies between average waiting time. However, how navigators would respond to larger average waiting times that also have a high degree of variability such that the minimum is less than the minimum of the range associated with the smaller average waiting time is an interesting question.

In the current paradigm, participants learned the virtual shopping mall environment via free exploration during a wayfinding task. That is, while the participant was instructed to find a particular target storefront, they were not guided along any predetermined routes as is common in the dual-solution paradigm which has been applied to this environmental setup (Boone et al., 2018). In future experiments, participants could be led along a predetermined route when learning the environment such that participants encounter every storefront once along the route. Errand lists could be manipulated such that the order of errand destinations either coincide with the order in which the storefronts were passed along the learning route or not. Whether navigators would stick to learned routes or take novel shortcuts depending on the alignment of errand lists with the learning route is an interesting question. Furthermore, participants could be given agency over the ordering of the errand list itself. For instance, navigators who prefer route-based strategies may structure errand lists to align with the learning route, while survey-strategists may optimize errand lists in consideration of possible shortcuts.

Conclusion

Research on navigational decision-making has been relatively scarce. Here, we presented a novel errand running paradigm for studying navigational decision-making during a ubiquitous navigation scenario. The errand running task was shown to have high internal consistency, test-retest reliability, and construct validity. The field is ripe with new directions of study using this errand running task, including route-based learning versus free exploration, preset versus custom errand lists, and manipulation of parameters governing uncertainty during errand running. Furthermore, the task can be implemented remotely, allowing for quick, large-scale data collection.

REFERENCES

Boone, A. P., Maghen, B., & Hegarty, M. (2019). Instructions matter: individual differences in navigation strategy and ability. *Memory & Cognition, 47,* 1401-1414.

Boone, A. P., Gong, X., & Hegarty, M. (2018). Sex differences in navigation strategy and efficiency. *Memory & Cognition, 46*, 909-922.

Blais, A., & Weber, E. U. (2006). A domain-specific risk-taking (dospert) scale for adult populations. *Judgment and Decision Making, 1*(1), 33-47.

Buelow, M. T., & Blaine, A. L. (2015). The assessment of risky decision making: a factor analysis of performance on the iowa gambling task, balloon analogue risk task, and Columbia card task. *Psychological Assessment, 27*(3), 777-785.

Burgess, N. (2008). Spatial cognition and the brain. *Annals of the New York Academy of Sciences, 1124,* 77-97.

Burgess, P. W., Alderman, N., Evans, J., & Wilson, B. (1998). The ecological validity of tests of executive function. *Journal of International Neuropsychological Society, 4*, 547-558.

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L., Dawson, D. R., Anderson, N. D., Gilbert, S. J., Dumontheil, I., & Channon, S. (2006). The case for the development and use of "ecologically valid" measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12,* 194-209.

Burns, S., Dawson, D. R., Perra, J., & Vas, A. K. (2019). Development, reliability, and validity of the multiple errands test home version (met-home) in adults with stroke. *The American Journal of Occupational Therapy, 73*(3), 1-10.

Chrastil, E. R. (2013). Neural evidence supports a novel framework for spatial navigation. *Psychonomic Bulletin & Review, 20,* 208-227.

Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review, 19,* 1-23.

Chrastil, E. R., & Warren, W. H. (2013). Active and passive spatial learning in human navigation: acquisition of survey knowledge. *Journal of Experiment Psychology, 39*(5), 1520-1537.

Chrastil, E. R., & Warren, W. H. (2014). From cognitive maps to cognitive graphs. *PLoS ONE, 9*(11): e112544. Doi:10.1371/journal.pone.0112544

Chrastil, E. R., & Warren, W. H. (2015). Active and passive spatial learning in human navigation: acquisition of graph knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41*(4), 1162-1178.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: a review. *Journal of Environmental Psychology, 24,* 329-340.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain.* Unpublished doctoral dissertation. McGill University, Montreal, Canada.

Coutrot, A., Manley, E., Goodroe, S., Gahnstrom, C., Filomena, G., Yesiltepe, D., Dalton, R. C., Wiener, J. M., Hölscher, C., Hornberger, M., & Spiers, H. J. (2022). Entropy of city street networks linked to future spatial navigation ability. *Nature,* 1-26. Doi:10.1038/s41586-022-04486-7

Coutrot, A., Silva, R., Manley, E., de Cothi, W., Sami, S., Bohbot, V. D., Wiener, J. M., Hölscher, C., Dalton, R. C., Hornberger, M., & Spiers, H. J. (2018). Global determinants of navigation ability. *Current Biology, 28,* 2861-2866.

Dawson, D. R., Anderson, N. D., Burgess, P., Cooper, E., Krpan, K. M., & Stuss, D. T. (2009). Further development of the multiple errands test: standardized scoring, reliability, and ecological validity for the baycrest version. *Archives of Psychical Medicine and Rehabilitation, 90,* S41-S51.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using g*power 3.1: tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Fox, C. R., & Ülkümen, G. (2011). Distinguishing two dimensions of uncertainty. Brun W., Keren G., Kirkeben G., Montgomery H., (eds). *Perspectives on Thinking, Judging, and Decision Making: A Tribute to Karl Halvor Teigen* (Universitetsforlaget, Oslo, Norway), 21-35.

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: contributions, challenges, and practical recommendations. *Consciousness & Cognition, 69,* 14-25.

Giraudo, M., & Pailhous, J. (1994). Distortions and fluctuations in topographic memory. *Memory & Cognition, 22*(1), 14-26.

Hazen, N. (1982). Spatial exploration and spatial knowledge: individual and developmental differences in very young children. *Child Development, 53,* 826-833.

He, Q., McNamara, T. P., Bodenheimer, B., & Klippel, A. (2019). Acquisition and transfer of spatial knowledge during wayfinding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(8), 1364-1386.

He, Q., McNamara, T. P., & Brown, T. I. (2019). Manipulating the visibility of barriers to improve spatial navigation efficiency and cognitive mapping. *Scientific Reports, 9*:11567, 1-12. Doi:10.1038/s41598-019-48098-0

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*, 151-176.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30,* 425-447.

Hegarty, M., & Waller, D. A. (2005). Individual differences in spatial abilities. In P. Shah, & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121-169). Cambridge University Press.

Huffman, D. J., & Ekstrom, A. D. (2018). Which way is the bookstore? A closer look at the judgments of relative directions task. *Spatial Cognition & Computation: An Interdisciplinary Journal*, *19*(2), 93-129.

Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, *52*, 93-129.

Jost, L., & Jansen, P. (2020). A novel approach to analyzing all trials in chronometric mental rotation and description of a flexible extended library of stimuli. *Spatial Cognition & Computation, 20*(3), 234-256.

Kessels, R. P. C., van den Berg Carla Ruis, E., & Brands, A. M. A. (2008). The backward span of the corsi block-tapping task and its association with the wais-iii digit span. *Assessment, 15*(4), 426-434.

Knight, C., Alderman, N., & Burgess, P. W. (2002). Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation, 12*(3), 231-355.

Krijnen, J. M. T., Ülkümen, G., Bogard, J. E., & Fox, C. R. (2022). Lay theories of financial well-being predict political and policy message preferences. *Journal of Personality and Social Psychology: Personality Processes and Individual Differences, 122*(2), 310-336.

Lauriola, M., Levin, I. P., & Hart, S. S. (2007). Common and distinct factors in decision making under ambiguity and risk: a psychometric study of individual differences. *Organizational Behavior and Human Decision Processes, 104,* 130-149.

Lawton, C. A. (1994). Gender differences in way-finding strategies: relationship to spatial ability and spatial anxiety. *Sex Roles, 30,* 765-779.

Lawton, C. A., & Kallai, J. (2002). Gender differences in wayfinding strategies and anxiety about wayfinding: a cross-cultural comparison. *Sex Roles, 47,* 389-401.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (BART). *Journal of Experimental Psychology: Applied, 8*(2), 75-84.

Lloyd, R. (1989). Cognitive maps: encoding and decoding information. *Annals of the Association of American Geographers, 79*(1), 101-124.

Magasi, S., Gohil, A., Burghart, M., & Wallisch, A. Understanding measurement properties. In: Law M., Baum C., Dunn W., (eds) Measuring occupational performance: supporting best practice in occupational therapy. 3rd ed. Thorofare: SLACK Incorporated; 2017. p 29-41.

Marchette, S. A., Bakker, A., & Shelton, A. L. (2011). Cognitive mappers to creatures of habit: differential engagement of place and response learning mechanisms predicts human navigational behavior. *The Journal of Neuroscience, 31*(43), 15264-15268.

McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Sala, S. D., & Maline, A. B. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence, 10*(4), 375-383.

McNamara, T. P. (2003). How are the locations of objects in the environment represented in memory? In: Freksa C., Brauer W., Habel C., Wender K.F. (eds) Spatial Cognition III. Spatial Cognition 2002. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 2685. Springer, Berlin, Heidelberg.

Meilinger, T. (2008). The network of reference frames theory: a synthesis of graphs and cognitive maps. In C. Freksa, N. S. Newcombe, P. Gärdenfors, & S. Wölfl (Eds.) *Spatial Cognition VI* (pp.334-360). Berlin: Springer.

Moeser, S. D. (1988). Cognitive mapping in a complex building. *Environment & Behavior, 20*(1), 21-49.

Montello, D. R., Waller, D., Hegarty, M., & Richardson, A. E. (2004). Spatial memory of real environments, virtual environments, and maps. In G. L. Allen, & D. Haun (Eds.), *Remembering where: advances in understanding spatial memory* (pp. 251-285). Mahwah, NJ: Lawrence Erlbaum Associates.

Muller, R. U., Stead, M., & Pach, J. (1996). The hippocampus as a cognitive graph. *Journal of General Physiology, 107,* 663-694.

Münzer, S., Fehringer, B. C.O.F., & Kühl, T. (2016). Validation of a 3-factor structure of spatial strategies and relations to possession and usage of navigational aids. *Journal of Environmental Psychology, 47,* 66-78.

Newman, E. L., Caplan, J. B., Kirschen, M. P., Korolev, I. O., Sekuler, R., & Kahana, M. J. (2007). Learning your way around town: how virtual taxicab drivers learn to use both layout and landmark information. *Cognition, 104*, 231-253.

Newman, P. M., & McNamara, T. P. (2021). Integration of visual landmark cues in spatial memory. *Psychological Research.* Doi:10.1007/s00426-021-01581-8

Newman, P. M., & McNamara, T. P. (2021). A computational cognitive model of judgments of relative direction. *Cognition, 209,* 1-16. Doi:10.1016/j.cognition.2020.104559

O’Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map.* Oxford, U.K.: Oxford University Press, Clarendon Press.

Peer, M., Brunec, I. K., Newcombe, N. S., & Epstein, R. A. (2021). Structuring knowledge with cognitive maps and cognitive graphs. *Trends in Cognitive Science, 25*(1), 37-54.

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain & Cognition, 66,* 260-264.

Pluck, G. (2019). Preliminary validation of a free-to-use, brief assessment of adult intelligence for research purposes: the matrix matching test. *Psychological Reports, 122*(2), 709-730.

Rand, D., Weiss, P. L., & Katz, N. (2009). Validation of the virtual met as an assessment tool for executive functions. *Neuropsychological Rehabilitation, 19*(4), 583-602.

Richardson, A. E., Montello, D. R., & Hegarty, M. (1999). Spatial knowledge acquisition from maps, and from navigation in real and virtual environments. *Memory & Cognition, 27,* 741-750.

Rossano, M. J., West, S. R., Robertson, T. J., Wayne, M. C., & Chase, R. B. (1999). The acquisition of route and survey knowledge from computer models. *Journal of Environmental Psychology, 19,* 101-115.

Rotenberg, S., Ruthralingam, M., Hnatiw, B., Neufeld, K., Yuzwa, K. E., Arbel, I., & Dawson, D. R. (2020). Measurement properties of the multiple errands test: a systematic review. *Archives of Physical Medicine and Rehabilitation, 101,* 1628-1642.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727-741.

Shelton, A. L., & Gabrieli, J. D. E. (2004). Neural correlates of individual differences in spatial learning strategies. *Neuropsychology, 18*(3), 442-449.

Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large environments. In H. W. Reese (Ed.), *Advances in child development and behavior* (pp. 9-55). New York, NY: Academic Press.

Tannenbaum, D., Fox, C. R., & Ülkümen, G. (2017). Judgment extremety and accuracy under epistemic vs. aleatory uncertainty. *Management Science, 63*(2), 497-518.

Taylor, H. A., & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory & Language, 35,* 371-391.

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology, 14,* 560-589.

Ülkümen, G., Fox, C. R., & Malle, B. F. (2016). Two dimensions of subjective uncertainty: clues from natural language. *Journal of Experimental Psychology: General.* Doi:10.1037/xge0000202

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599-604.

Walters, D. J., Ülkümen, G., Tannenbaum, D., & Erner, C. (2022). Investor behavior under epistemic versus aleatory uncertainty. Available at SSRN: Doi:10.2139/ssrn.3695316

Weber, E. U., Blais, A., & Betz, N. E. (2002). A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15,* 263-290.

Weisberg, S. M., & Ekstrom, A. D. (2021). Hippocampal volume and navigational ability: the map(ping) is not to scale. *Neuroscience & Biobehavioral Reviews, 126*(6), 102-112.

Weisberg, S. M., & Newcombe, N. S. (2018). Cognitive maps: some people make them, some people struggle. *Current Directions in Psychological Science*, *27*(4), 220-226.

Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Russell, A. E. (2014). Variations in cognitive maps: understanding individual differences in navigation. *Journal of Experimental Psychology*, *40*(3), 669-682.

Yu, S., Boone, A. P., Chuanxiuyue, H., Davis, R. C., Hegarty, M., Chrastil, E. R., & Jacobs, E. G. (2021). Age-related changes in spatial navigation are evidence by midlife and differ by sex. *Psychological Science,* 1-13. Doi:10.1177/0956797620979185

APPENDIX

Appendix A: Errand Running Task Trial Lists

Wayfinding Targets			Pre-generated Waiting Times		
Original Environment	Mirror Environment	Waiting Time Range (sec)	Attempt 1	Attempt 2	Attempt 3
1. Dry Cleaners	1. Limbo Lounge	8-12	11	9	8
2. Butcher Shop	2. Big O Tires	8-12	10	8	8
3. Clothing Store	3. Costume Party	8-12	10	9	9
4. Coffee Store	4. Hughe’s Auto Sales	8-12	11	9	10
5. Candle Shop	5. Video Services	8-12	10	9	9
6. Food Market	6. Mike’s Restaurant	2-18	18	8	7
7. Burger City	7. Bake Shop	2-18	4	2	14
8. Cookie Shop	8. Computers Store	2-18	15	18	3
9. Art Gallery	9. Tobacco Shop	2-18	11	11	6
10. Piano Store	10. Pharmacy	2-18	16	15	17
11. Toy Store	11. Gift Shop	8-12	12	10	11
12. Travel Shop	12. House of Pizza	2-18	8	12	16

Table A1. Displays the wayfinding target storefronts during the learning phase and the order in which they were visited by the participants. The waiting times at each storefront for each attempt were pre-generated from each storefront’s associated waiting time range such that all participants experienced the same waiting times at each.

Errand Lists

Trial	Original Env.	Mirror Env.	Trial Type	Pre-generated Waiting Times (sec)
1	1. Burger City 2. Cookie Shop 3. Travel Shop	1. Bake Shop 2. Computers Store 3. House of Pizza	Wide	14 4 17
2	1. Candle Shop 2. Coffee Store 3. Dry Cleaners	1. Video Services 2. Hughe's Auto Sales 3. Limbo Lounge	Narrow	9 9 10
3	1. Art Gallery 2. Burger City 3. Food Market	1. Tobacco Shop 2. Bake Shop 3. Mike's Restaurant	Wide	12 5 15
4	1. Butcher Shop 2. Clothing Store 3. Toy Store	1. Big O Tires 2. Costume Party 3. Gift Shop	Narrow	11 8 8
5	1. Burger City 2. Travel Shop 3. Piano Store	1. Bake Shop 2. House of Pizza 3. Pharmacy	Wide	2 6 4
6	1. Candle Shop 2. Clothing Store 3. Butcher Shop	1. Video Services 2. Costume Party 3. Big O Tires	Narrow	10 9 11
7	1. Food Market 2. Travel Shop 3. Burger City	1. Mike's Restaurant 2. House of Pizza 3. Bake Shop	Wide	14 12 17
8	1. Toy Store 2. Candle Shop 3. Dry Cleaners	1. Gift Shop 2. Video Services 3. Limbo Lounge	Narrow	8 9 8
9	1. Art Gallery	1. Tobacco Shop	Wide	2

	2. Piano Store 3. Burger City	2. Pharmacy 3. Bake Shop		9 12
10	1. Coffee Store 2. Dry Cleaners 3. Toy Store	1. Hughe’s Auto Sales 2. Limbo Lounge 3. Gift Shop	Narrow	12 8 8
11	1. Cookie Shop 2. Piano Store 3. Travel Shop	1. Computers Store 2. Pharmacy 3. House of Pizza	Wide	10 9 16
12	1. Dry Cleaners 2. Butcher Shop 3. Toy Store	1. Limbo Lounge 2. Big O Tires 3. Gift Shop	Narrow	10 12 10
13	1. Food Market 2. Art Gallery 3. Cookie Shop	1. Mike’s Restaurant 2. Tobacco Shop 3. Computers Store	Wide	5 10 11
14	1. Dry Cleaners 2. Butcher Shop 3. Candle Shop	1. Limbo Lounge 2. Big O Tires 3. Video Services	Narrow	11 8 10
15	1. Food Market 2. Travel Shop 3. Piano Store	1. Mike’s Restaurant 2. House of Pizza 3. Pharmacy	Wide	18 6 18
16	1. Butcher Shop 2. Candle Shop 3. Dry Cleaners	1. Big O Tires 2. Video Services 3. Limbo Lounge	Narrow	8 8 9
17	1. Art Gallery 2. Piano Store 3. Travel Shop	1. Tobacco Shop 2. Pharmacy 3. House of Pizza	Wide	6 8 3

18	1. Coffee Store 2. Toy Store 3. Butcher Shop	1. Hughe's Auto Sales 2. Gift Shop 3. Big O Tires	Narrow	11 9 11
19	1. Piano Store 2. Burger City 3. Art Gallery	1. Pharmacy 2. Bake Shop 3. Tobacco Shop	Wide	15 9 14
20	1. Coffee Store 2. Candle Shop 3. Toy Store	1. Hughe's Auto Sales 2. Video Services 3. Gift Shop	Narrow	9 12 9
21	1. Burger City 2. Food Market 3. Travel Shop	1. Bake Shop 2. Mike's Restaurant 3. House of Pizza	Wide	10 18 4
22	1. Dry Cleaners 2. Butcher Shop 3. Coffee Store	1. Limbo Lounge 2. Big O Tires 3. Hughe's Auto Sales	Narrow	8 9 12
23	1. Travel Shop 2. Burger City 3. Food Market	1. House of Pizza 2. Bake Shop 3. Mike's Restaurant	Wide	3 2 15
24	1. Dry Cleaners 2. Coffee Store 3. Toy Store	1. Limbo Lounge 2. Hughe's Auto Sales 3. Gift Shop	Narrow	12 9 12
25	1. Food Market 2. Travel Shop 3. Cookie Shop	1. Mike's Restaurant 2. House of Pizza	Wide	17 5 11

		3. Computers Store		
26	1. Candle Shop 2. Clothing Store 3. Dry Cleaners	1. Video Services 2. Costume Party 3. Limbo Lounge	Narrow	8 9 11
27	1. Burger City 2. Food Market 3. Art Gallery	1. Bake Shop 2. Mike’s Restaurant 3. Tobacco Shop	Wide	16 6 5
28	1. Butcher Shop 2. Toy Store 3. Dry Cleaners	1. Big O Tires 2. Gift Shop 3. Limbo Lounge	Narrow	12 12 9
29	1. Piano Store 2. Travel Shop 3. Art Gallery	1. Pharmacy 2. House of Pizza 3. Tobacco Shop	Wide	16 17 13
30	1. Coffee Store 2. Candle Shop 3. Butcher Shop	1. Hughe’s Auto Sales 2. Video Services 3. Big O Tires	Narrow	8 9 10
31	1. Cookie Shop 2. Art Gallery 3. Piano Store	1. Computers Store 2. Tobacco Shop 3. Pharmacy	Wide	12 7 9
32	1. Toy Store 2. Butcher Shop 3. Clothing Store	1. Gift Shop 2. Big O Tires 3. Costume Party	Narrow	9 11 8

Table A2. Ordered list of every errand running trial in the original environment used in all experiments, along with their trial types and pre-generated waiting times. Narrow trials contain storefronts with associated waiting time ranges of 8-12 seconds (Wide: 2-18 sec).

Appendix B: Santa Barbara Sense of Direction Scale (Hegarty et al., 2002)

This questionnaire consists of several statements about your spatial and navigational abilities, preferences, and experiences. After each statement, you should click a box to indicate your level of agreement with the statement. Click the first box if you strongly agree, last box if you strongly disagree, or a box in between if your agreement is intermediate. Click the middle box if you neither agree nor disagree.

1. I am very good at giving directions.
2. I have a poor memory for where I left things.
3. I am very good at judging distances.
4. My “sense of direction” is very good.
5. I tend to think of my environment in terms of cardinal directions (N,S,E,W).
6. I very easily get lost in a new city.
7. I enjoy reading maps.
8. I have trouble understanding directions.
9. I am very good at reading maps.
10. I don’t remember routes very well while riding as a passenger in a car.
11. I don’t enjoy giving directions.
12. It’s not important to me to know where I am.
13. I usually let someone else do the navigational planning for long trips.
14. I can usually remember a new route after I have traveled it only once.
15. I don’t have a very good “mental map” of my environment.

Appendix C: Domain-Specific Risk-Taking Scale (Blais & Weber, 2006)

Risk-Taking Behaviors Scale

For each of the following statements, please indicate the likelihood that you would engage in the described behavior if you were to find yourself in that situation. Provide a rating from “Extremely Unlikely” to “Extremely Likely”, with the middle representing “Not Sure”.

Risk Perception Scale

People often see some risk in situations that contain uncertainty about what the outcome or consequences will be and for which there is the possibility of negative consequences. However, the riskiness is a very personal and intuitive notion, and we are interested in your gut level assessment of how risky each situation or behavior is. For each of the following statements, please indicate how risky you perceive each situation. Provide a rating from “Not at all Risky” to Extremely Risky”.

1. Admitting that your tastes are different from those of a friend. (S)
2. Going camping in the wilderness. (R)
3. Betting a day’s income at the horse races. (F)
4. Investing 10% of your annual income in a moderate growth mutual fund. (F)
5. Drinking heavily at a social function. (H/S)
6. Taking some questionable deductions on your income tax return. (E)
7. Disagreeing with an authority figure on a major issue. (S)
8. Betting a day’s income at a high-stake poker game. (F)
9. Having an affair with a married man/women. (E)
10. Passing off somebody else’s work as your own. (E)
11. Going down a ski run that is beyond your ability. (R)
12. Investing 5% of your annual income in a very speculative stock. (F)

13. Going whitewater rafting at high water in the spring. (R)

14. Betting a day's income on the outcome of a sporting event. (F)

15. Engaging in unprotected sex. (H/S)

16. Revealing a friend's secret to someone else. (E)

17. Driving a car without wearing a seat belt. (H/S)

18. Investing 10% of your annual income in a new business venture. (F)

19. Taking a skydiving class. (R)

20. Riding a motorcycle without a helmet. (H/S)

21. Choosing a career that you truly enjoy over a more secure one. (S)

22. Speaking your mind about an unpopular issue in a meeting at work. (S)

23. Sunbathing without sunscreen. (H/S)

24. Bungee jumping off a tall bridge. (R)

25. Piloting a small plane. (R)

26. Walking home along at night in an unsafe area of town. (H/S)

27. Moving to a city far away from your extended family. (S)

28. Starting a new career in your mid-thirties. (S)

29. Leaving your young children along at home while running an errand. (E)

30. Not returning a wallet your found that contains $200. (E)

Note: E = Ethical, F = Financial, H/S = Health/Safety, R = Recreational, and S = Social.

Appendix D: Wayfinding Anxiety Scale (Lawton & Kallai, 2002)

This questionnaire contains statements relating to navigation scenarios. After each statement, click on a box indicating how anxious you would be in the described navigation scenario, ranging from “Not at all Anxious” to “Very Anxious”.

1. Deciding which direction to walk in an unfamiliar city or town after coming out of a train/bus/metro station or parking garage.
2. Finding my way to an appointment in an unfamiliar area of a city or town.
3. Leaving a store that I have been to for the first time and deciding which way to turn to get to a destination.
4. Finding my way back to a familiar area after realizing I have made a wrong turn and become lost while traveling.
5. Finding my way in an unfamiliar shopping mall, medical center, or large building complex.
6. Finding my way out of a complex arrangement of offices that I have visited for the first time.
7. Trying a new route that I think will be a shortcut, without a map.
8. Pointing in the direction of a place outside that someone wants to get to and has asked for directions when I am in a windowless room.

Appendix E: Epistemic-Aleatory Rating Scale (Walters et al., 2022)

In this questionnaire, please consider the scenario below, and rate how much you agree (or disagree) with the following statements about the scenario on a scale from “Not at all” to Very much”.

Consider the task of running three errands in your come town or city and returning home afterwards. Imagine that you want to tell your friend to meet you at home around the same time you arrive back home after your run all your errands.

The approximate time to tell your friend to meet you at your home …

1. … is knowable in advance, given enough information.
2. … is something that becomes more predictable with additional knowledge.
3. … is something that well-informed people would agree on.
4. … Is determined by chance factor.
5. … could play out in different ways on similar occasions.
6. … is something that has an element of randomness.

Note: Items 1-3 measure perceptions about epistemicness and items 4-6 measure perceptions about aleatoriness.