



# A meta-analysis of sex differences in human navigation skills

Alina Nazareth<sup>1</sup> · Xing Huang<sup>2</sup> · Daniel Voyer<sup>3</sup> · Nora Newcombe<sup>4</sup>

Published online: 3 July 2019  
© The Psychonomic Society, Inc. 2019

## Abstract

There are inconsistent reports regarding behavioral sex differences in the human navigation literature. This meta-analysis quantifies the overall magnitude of sex differences in large-scale navigation skills in a variety of paradigms and populations, and examines potential moderators, using 694 effect sizes from 266 studies and a multilevel analytic approach. Overall, male participants outperform female participants, with a small to medium effect size ( $d = 0.34$  to  $0.38$ ). The type of task, the type of dependent variable and the testing environment significantly contribute to variability in effect sizes, although there are only a few situations in which differences are either nonexistent or very large. Pointing and recall tasks (and the deviation scores associated with them) show larger sex differences than distance estimation tasks or learning to criterion. Studies with children younger than 13 years showed much smaller effect sizes ( $d = .15$ ) than older age groups. We discuss the implications of these findings for understanding sex differences in human spatial navigation and identify avenues for future navigation research.

**Keywords** Meta-analysis · Sex difference · Navigation · Spatial skills

Wayfinding or goal-directed navigation is important for the survival of most mobile species. Against that backdrop, it may seem surprising that humans differ in their navigation skills. Nevertheless, there is considerable individual variation in wayfinding accuracy and confidence, across a variety of experimental paradigms (Newcombe, 2018). We do not yet fully understand the correlates and causes of such individual variation, but sex of participants may be one dimension that matters. There are frequently cited instances of male advantage in the literature (Dabbs, Chang, Strong, & Milun, 1998; Gagnon et al., 2018; Lawton, 1994; Lawton & Kallai, 2002). However, there are also reports of no sex differences (Driscoll, Hamilton, Yeo, Brooks & Sutherland, 2005; Herman, Kail, & Siegel, 1979; O’Laughlin & Brubaker, 1998; Saucier et al.,

2002; Schmitzer-Torbert, 2007) and even studies showing a female advantage on navigation tasks under certain conditions (Burigat & Chittaro, 2007). These variations in results raise the question of whether sex differences exist in human spatial navigation. However, in popular perception and the popular press, there is little doubt; it is common to see headlines such as “Study says that men have the better sense of direction” (Gregoire, 2015).

A meta-analysis can bring together the empirical literature, integrating reports with varying sample sizes and demographic characteristics, and a wide range of methodologies (testing environments, experimental procedures, use of technology, learning perspectives, and outcome measures). Previous reviews do not address the core question of quantifying behavioral sex differences in human navigation and examining potential moderators, although they do provide valuable information. The most relevant prior effort is a narrative literature review, in which Coluccia and Louse (2004) found that, while more than 50% of papers did not find significant sex differences, the remainder found better performance by males, and no studies reported females doing better. This pattern of evidence was suggestive of a male advantage, but in the absence of a systematic meta-analysis, it is impossible to know for sure, let alone to estimate the size of sex differences and the role of potential moderators in any effects. In addition, in the years since the Collucia and Louse review, there has

**Electronic supplementary material** The online version of this article (<https://doi.org/10.3758/s13423-019-01633-6>) contains supplementary material, which is available to authorized users.

✉ Alina Nazareth  
alina.nazareth@gmail.com

<sup>1</sup> Temple University, Philadelphia, PA, USA

<sup>2</sup> University of New Brunswick Fredericton, Fredericton, NB, Canada

<sup>3</sup> Psychology, University of New Brunswick, Fredericton, NB, Canada

<sup>4</sup> Psychology, Temple University, Philadelphia, PA, USA

been a great deal more navigation research. For example, the paper by Burigat and Chittaro (2007) showing a female advantage came out later.

There are two other relevant reviews, both of which are quantitative, although they address somewhat different questions. Jonasson (2005) conducted a meta-analysis of sex differences in rodent navigation, which can be seen as relevant to human differences given our common mammalian heritage. The analysis showed that sex differences in rodents varied with rearing environment, training protocols, and navigation task. Crucially, there was evidence of species variation, suggesting that there is no basic mammalian pattern. Mice actually demonstrated a small female advantage in the water maze task. In another review relevant to our purpose, Boccia, Nemmi, and Guariglia (2014) conducted a meta-analysis of 66 fMRI studies of the neural substrates involved in human navigation. They found interesting sex-linked variations in brain activation moderated by environment familiarity and egocentric and allocentric strategies. However, their review did not address the question of behavioral sex differences in navigation accuracy.

Accordingly, our purpose was to provide a comprehensive meta-analysis of behavioral sex differences in human spatial navigation. We aimed to quantify the overall magnitude of sex differences and examine potential moderators of the effects. To conduct this synthesis, we began with defining spatial navigation and then specified potential moderators.

## Defining wayfinding

There are a wide variety of cognitive processes involved in navigation, and experimenters have developed many different small-scale and large-scale paradigms and self-report measures to investigate individual differences (Wiener, Büchner, & Hölscher, 2009). There does not seem to be an agreed-upon definition of what constitutes “spatial navigation.” Coluccia and Louse (2004) defined navigation as a “complex of all the skills used for locating themselves with respect to a point of reference or an absolute system of coordinates, (p. 329).” However, this definition encompasses spatial processes at different scales (see Montello, 1993, 2010), despite substantial evidence that navigation in large-scale environments is fundamentally different from small-scale tasks in which all objects are intervisible (Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Learmonth, Newcombe, Sheridan, & Jones, 2003; Montello, 1993; Padilla, Creem-Regehr, Stefanucci & Cashdan, 2017).

For the purpose of the present meta-analysis, we defined navigation as skills involved in encoding spatial relations at environmental scale and transforming them to orient and navigate in environmental space. This definition includes spatial navigation with external representations (e.g., maps), but

excludes reorientation tasks conducted in small spaces and small-scale mazes with target locations visible from a single vantage point. The only exception to the *scale* criterion for our review was the virtual water-maze task (Astur, Ortiz, & Sutherland, 1998), based on the Morris water-maze task (Morris, 1981). This task is a well-accepted measure in the animal spatial cognition literature, where it has been shown to engage hippocampal processes that support cognitive mapping (Astur, Taylor, Mamelak, Philpott, & Sutherland, 2002). It has been widely used to study individual differences in the use of proximal and distal cues and has a known neural substrate implicated in navigation (Daugherty, Bender, Yuan, & Raz, 2015; Packard & McGaugh, 1996). Further, the hidden platform task occludes the target to be reached and thus can be arguably included within the environment scale definition.

## Potential moderators

A systematic approach to moderator identification is likely to decrease the risk of Type I error in meta-analytic results (Lipsey & Wilson, 2001). Accordingly, we identified empirically supported or theoretically relevant factors *a priori* for consideration in the moderator analysis. Table 1 lists the measure-level variables that we identified for later coding. Below, we discuss the empirical and theoretical basis that led us to select these factors as potential moderators.

**Task goal** There are large variations in the task goal used to assess a participant’s navigation skills. We list these task goals in eight distinct categories (see Table 1). These eight categories may demonstrate large, small, or negligible sex differences, depending on the cognitive processes involved in completing the task goal. For example, Jansen-Osmann, Schmid, and Heil (2007) found no sex difference on a route-recall task, whereas map reading showed a large male advantage. In contrast, Allen and Wittenborg (1998) found a significant male advantage for route recall, but not for a distance task. This type of inconsistency across studies suggests that inclusion of task goals as a moderator is important. In addition, task goals form the basic level from which all other task requirements are defined, and thus we used this moderator as the basic unit of analysis in subgroup analyses.

**Perspective** An important difference in navigation tasks pertains to the type of visual representation of the environment—route or survey. In a route perspective, individuals experience the environment from a first-person perspective (i.e., they locate target objects in relation to their own position). In contrast, in a survey perspective, individuals see the entire spatial layout from a bird’s-eye view. The inclusion of perspective as a moderator is particularly important in light of research suggesting that men show a preference for a Euclidean orientation

**Table 1** Coding scheme for moderators in the present meta-analysis

Moderator	Categories
<b>Task goal</b>	(1) <i>Pointing to target</i> (e.g., point to a learned/familiar target: Pointing) (2) <i>Retracing learned route</i> by recall or recognition (e.g., maze exit, walk to a target location: Recall/recognition) (3) <i>Orienteering with cardinal directions</i> (e.g., orienteering based on N/S/E/W: Cardinal directions) (4) <i>Estimating distance</i> (e.g., estimate the distance to a target building: Distance) (5) <i>Positioning targets or landmarks</i> (e.g., map arrangement/model-building: Landmark position) (6) <i>Reading maps</i> (e.g., navigate by reading a map: Maps) (7) <i>Navigating with verbal instructions</i> (e.g., navigate with verbal directions: Verbal instructions) (8) <i>Reaching learning criterion</i> (Learning)
<b>Perspective</b>	(1) <i>Route</i> (e.g., walking through a city) (2) <i>Survey</i> (e.g., flying over a city, reading a map) (3) <i>Both</i> (e.g., using a map to walk through a city)
<b>Outcome measure</b>	(1) <i>Accuracy</i> (e.g., number of correct turns during route recall) (2) <i>Degrees of deviation</i> (e.g., degree error in pointing to a target location) (3) <i>Distance</i> (e.g., distance traveled on a particular route) (4) <i>Response time</i> (e.g., time taken to reach a target location)
<b>Route direction</b>	(1) <i>Forward</i> (e.g., move towards a learned/familiar target) (2) <i>Backward</i> (e.g., go back to the starting locations i.e., opposite to the learning direction) (3) <i>Free choice</i>
<b>Route selection</b>	(1) <i>Free choice</i> (e.g., any route to find the target location) (2) <i>Exact way</i> (e.g., same learned path to target location) (3) <i>Shortcut</i> (e.g., most direct way to the target location)
<b>Timing condition</b>	(1) <i>Untimed</i> (e.g., unlimited time for task completion) (2) <i>Timed</i> (e.g., limited time for task completion)
<b>Cues</b>	(1) <i>Distal</i> (e.g., nontarget salient landmarks that cannot be directly reached through learned routes) (2) <i>Proximal</i> (e.g., nontarget salient landmarks that may be reached through learned routes or through nonsignificant detours)
<b>Environment</b>	(1) <i>Indoor</i> (e.g., routes in a building basement) (2) <i>Outdoor</i> (e.g., routes in a city) (3) <i>Indoor–outdoor</i> (e.g., routes through a building basement and into an open campus/field) (4) <i>Water maze</i>
<b>Familiarity</b>	(1) <i>New</i> (e.g., find maze exit with no prior learning/learning criterion) (2) <i>Learned</i> (e.g., repeated learning in an environment before testing) (3) <i>Familiar</i> (e.g., point to a local landmark)
<b>Testing medium</b>	(1) <i>Real</i> (e.g., routes in a campus/city) (2) <i>Virtual</i> (e.g., routes/mazes on a computer screen) (3) <i>Symbolic</i> (e.g., routes on a map/model)
<b>Feedback</b>	(1) <i>Immediate</i> (e.g., target location revealed in each trial) (2) <i>No feedback</i> (e.g., no indication of performance)
<b>Hints</b>	(1) <i>Hints</i> (e.g., hints provided during testing trials) (2) <i>No hints</i> (e.g., no hints provided during testing)
<b>Device assistance</b>	(1) <i>No device</i> (e.g., no assistive device during testing) (2) <i>Device</i> (e.g., compass/map during testing)
<b>Learning interval</b>	(1) <i>Immediate</i> (e.g., testing begins after learning criterion is met) (2) <i>Short</i> (e.g., testing begins within 24 hours of learning) (3) <i>Long</i> (e.g., testing begins after 24 hours of learning)

In all moderators, the categories “not applicable” or “not reported” were also added when relevant

strategy using cardinal directions and distances, but that women prefer a landmark strategy that relies on a sequence of turns and proximal cues (Dabbs, Chang, Strong, & Milun, 1998; Lawton, 1994; Lawton & Kallai, 2002). This observation raises the possibility that men and women may differentially benefit from a particular perspective. There are also navigation studies that use a combination of route and survey perspectives, and we categorized these studies into a third group. At a more global, theoretical level, perspective can be seen as

a central defining component of spatial navigation, possibly because a change in perspective is a component of many navigation tasks (Wolbers & Hegarty, 2010). The importance of perspective as a moderator therefore manifests itself both at the empirical and theoretical level.

**Outcome measure** Outcome measures are often presented as reflecting different dimensions of performance. For example, in mental rotation, Lohman (1986) presented response time as

reflecting speed of processing, whereas accuracy was believed to reflect level of ability. From this perspective, findings of a larger male advantage with accuracy than with response time in a spatial navigation task by Malinowski and Gillespie (2001) emphasizes sex differences in level of ability. In the context of spatial navigation, methods of scoring other than accuracy and response time have also been used to quantify performance. With this in mind, we therefore categorized outcome measures into four categories—accuracy, degrees of deviation, distance, and response time. Examining potential variations in the magnitude of sex differences in outcome measures is crucial to assess current navigation metrics and biases built into them and how these might affect the magnitude of sex differences.

**Route direction** Depending on task goals and outcome measures, participants may need to navigate routes in the same direction as initially learned or in opposite directions. Potential variations in sex differences based on route direction could be indicative of differences in working memory and strategy flexibility. For example, individuals who depend on a rigid sequence of cues and turns may perform better in the forward direction than in the reversal condition. Given the male advantage in visuospatial working memory (Voyer, Voyer, & Saint Aubin, 2017), the magnitude of sex difference might be larger during route reversal as a result of the added working memory load. Hence, we included route directions as a potential moderator and categorized movement into three groups—forward, backward, and free choice.

**Route selection** In the research we retrieved in the literature, participants were often asked to select a specific route to meet the task goal. These route selection options could be categorized into three groups—free choice, exact way, and shortcut. Empirical findings from Choi, McKillop, Ward, and L'Hirondelle (2006) showed that asking participants to use a shortcut can produce a male advantage, whereas following the exact way that was learned can produce a female advantage. Furthermore, Boone, Gong, and Hegarty (2018) reported that men are more likely than women to use shortcuts, whereas women prefer the learned route. Therefore, we included route selection as a moderator in our analysis.

**Timing** Research suggests that anxiety and self-doubt negatively influence the ability to encode important spatial features and consequently navigation performance (Saucier et al., 2002). Women generally report higher spatial anxiety than men (Huang & Voyer, 2017; Lawton, 1994; Lawton & Kallai, 2002) and lower self-confidence in spatial tasks (Huang & Voyer, 2017; O'Laughlin & Brubaker, 1998). This suggests that women may be affected differentially by the imposition of time constraints. For example, it is possible that the male advantage in spatial navigation could be larger

under time constraints compared with unconstrained conditions, as has been reported for mental rotation (Voyer, 2011). Hence, we included timing conditions as a potential moderator with two categories—untimed and timed.

**Cues** Environments may also differ in the amount and diversity of visual cues available in them. Visual cues used to orient and navigate can be broadly divided into two main categories—proximal and distal (Chai & Jacobs, 2010; Knierim & Hamilton, 2011; O'Keefe & Nadel, 1978; Sandstrom, Kauffman & Huettel, 1998; Vorhees & Williams, 2006). Proximal and distal cues can be differentiated by the effect on parallax when individuals change their location in an environment, with the latter cue type providing a more stable bearing than the former (Knierim & Hamilton, 2011). Most work manipulating the role of visual cues, however, has been restricted to small-scale navigation paradigms with few exceptions (Padilla et al., 2017), even though large-scale environments are also obviously rich in both proximal and distal cues. Existing research suggests a male superiority in environments with limited or directional landmark cues (Astur et al., 2004; Barkley & Gabriel, 2007; Cánovas, Garcia, & Cimadevilla, 2011; Moffat, Hampson, & Hatzipantelis, 1998) and no sex differences in environments rich with proximal cues (Astur, Tropp, Sava, Constable, & Markus, 2004; Saucier et al., 2002). Sex differences in human visual cue processing, mentioned by many authors (Astur, Ortiz & Sutherland, 1998; Barkley & Gabriel, 2007; Chai & Jacobs, 2009, 2010; Choi et al., 2006; Sandstrom, Kauffman & Huettel, 1998) provide strong empirical support for the inclusion of cues as a moderator of sex differences in navigation.

The role of distal and proximal cues in spatial navigation also has theoretical implications, as these cues are seen as central to many navigation tasks, likely as a result of the much-documented link between the use of such cues and hippocampal functioning (Poulter, Hartley, & Lever, 2018). Therefore, a result showing an effect of cues on the magnitude of sex differences in spatial navigation could raise the possibility of differentiated hippocampal functions in men and women.

**Environment** The environment experienced by the participant while navigating can be seen as reflecting variation in the actual scale of the surroundings. Specifically, we categorized the environment as indoor, outdoor, or a combination of indoor–outdoor to refer to a limited, unlimited, or mixed-area scale environment, respectively. More importantly, recall that although some researchers might consider them small-scale tasks, water-maze tasks were included in our sample as they essentially provide the richness of large-scale tasks within a limited area by means of a manipulation of distal and proximal cues (Poulter et al., 2018). Therefore, this type of

task formed a fourth environment category. In view of the argument that water-maze tasks provide the clearest available evidence for sex differences in spatial navigation strategy selection (Boone et al., 2018), it is critical for us to consider this type of tasks as a central component of the environment moderator.

**Familiarity** Navigation tasks can be conducted in environments that are either learned specifically for the purpose of testing (e.g., Jansen-Osmann, Schmid, & Heil, 2007; Nazareth, Weisberg, Margulis, & Newcombe, 2018) or that are already familiar to the participant (Meilinger, Frankenstein, & Bühlhoff, 2013). Boccia, Nemmi, and Guariglia (2014) found that different neural correlates were involved in the processing of familiar and learned environments. It is therefore not surprising that Abu-Obeid (1998) found a significant male advantage in a new environment but not in a familiar one. Accordingly, we included familiarity as a potential moderator of sex differences in navigation.

**Testing medium** Coluccia and Louse (2004) reported that males outperformed females in about 57% of the research using “simulated” (i.e., virtual) environments, very similar to the 59% found for real environments. However, this value dropped to 42% for maps. More importantly, virtual and real environment showed no cases where females outperformed males, whereas this figure was around 18% for maps. Therefore, we aimed to determine whether Coluccia and Louse’s findings on this moderator would be reflected in smaller sex differences for map-like tasks compared with virtual or real environments in our comprehensive quantitative review. Accordingly, we categorized testing medium into virtual, real, and symbolic groups, where symbolic referred to the use of a map or a similar medium.

**Feedback and hints** There is evidence in the cognition literature of sex differences in attitudes and responses to failures and achievements (Dweck, 1986). Although not examined frequently in the navigation literature, the inclusion of feedback and/or hints from guides as moderators accounts for important differences in men’s and women’s response to performance feedback (Lenney, 1977) and their ability to use the feedback constructively for spatial updating. We therefore considered feedback and hints as two separate moderators. Feedback had two categories—immediate feedback and no feedback. Hints had two categories—hints and no hints—based on whether participants were given hints during the task goal.

**Device assistance** We know that device assistance from a map provides a survey perspective of the environment and consequently may provide an additional support to individuals—primarily women—who rely on a route strategy (Dabbs,

Chang, Strong, & Milun, 1998; Lawton, 1994; Lawton & Kallai, 2002), as is described in more detail in the section on perspective. Hence, device assistance was used as a potential moderator with two categories—device and no device.

**Learning intervals** Learning intervals were broken down into three categories—immediate (no gap between learning and testing phases), short (less than 24 hours between both phases), and long (more than 24 hours gap). Forgetting occurs over time, as a century of memory research shows, but spatial memory also seems to undergo consolidation during wakeful rest or sleep (Skaggs & McNaughton, 1996), in which it may actually strengthen. We are unaware of data suggesting that such “replay” differs by sex, but it seemed prudent to examine the question in our analysis, given the importance of these processes in the literature.

**Age** There is gradual age-related change in navigation skills in children between 6 and 12 years of age (e.g., Acredolo, Pick, & Olsen, 1975; Allen, Kirasic, Siegel, & Herman, 1979; Heth, Cornell, & Alberts, 1997; Laurance, Learmonth, Nadel, & Jacobs, 2003; Overman, Pate, Moore, & Peuster, 1996) with spatial representations maturing near adolescence (Liben, Myers, Christensen, & Bower, 2013). Further, the large body of literature on the influence of hormone levels on spatial skills (Brake & Lacasse, 2018; Lisofsky, Riediger, Gallinat, Lindenberger, & Kühn, 2016; Puts, McDaniel, Jordan, & Breedlove, 2008) highlights the importance of participant age during testing (e.g., puberty, pregnancy, menopause). Thus, in addition to the task-related variations, we investigated the moderating influence of participant age in sex differences in navigation skills.

## Current meta-analysis

The main aim of the current meta-analysis was to provide a summary of sex differences found in navigation research. To this end, we conducted an exhaustive search for relevant published and unpublished data collected in different countries with different sample populations and using different navigation paradigms. To our knowledge, no meta-analysis specifically quantifying human sex differences in navigation has been published to date, and we believe that the current meta-analysis serves to fill that gap in the navigation literature. It is important to clarify that the current meta-analysis excludes self-report measures of navigation skills as well as any navigation studies involving small-scale paradigms in which the entire testing environment can be viewed from a single vantage point. We also constrained the meta-analysis to include only studies with typical populations. Further, published research studies that failed to include numerical values for male and female sex differences and for which the corresponding



author was unable to provide these data were excluded from the meta-analysis.

Not surprisingly, many of the studies we retrieve included several outcome measures relevant to our research questions or provided nonindependent measurements of their participants under different experimental conditions relevant to the potential moderators we identified. In the context of fixed effects, random effects, or mixed effects meta-analysis, such nonindependent effect sizes would violate the assumptions underlying these types of analyses (Borenstein, Hedges, Higgins, & Rothstein, 2009) and make the results uninterpretable. Accordingly, for all data analyses, we used multilevel meta-analysis, as this method is particularly well suited to handle the hierarchical nature of a meta-analytic data set and nonindependent effect sizes (Raudenbush & Bryk, 2002).

## Method

### Literature search and study selection

A primary search was conducted in databases of PsycINFO, PsycARTICLES, and ERIC using various Boolean combinations of the search terms *wayfinding*, *spatial*, *navigation*, *orientation*, *maps*, *representation*, *cognition*, and *environment*. The search included all available records from the year 1803 (default lower limit in the search engine) to October 2017. We refrained from using *sex/gender differences* in our initial search to prevent prematurely excluding papers that did not have sex/gender differences as their central theme. Our filters excluded patents, reviews, books, magazine articles, and any non-peer-reviewed sources. These searches resulted in 10,663 nonoverlapping hits, which was initially reduced to 1,164 with the help of our inclusion criteria, detailed below. Foreign language articles including an English abstract were also included in the analyses. Theses and dissertations were considered as possible sources of unpublished material, but were excluded if the same data had been published. In such cases, only the published version was included in the meta-analysis. In addition, requests for unpublished data announcements were posted to the following LISTSERVs—Cognitive Development Society (CDS), Spatial Learning Network of the Spatial Intelligence and Learning Center (SILC), and Canadian Society for Brain, Behavior and Cognitive Science (CSBBS). Altogether, we received 47 responses to our LISTSERV announcement, although most of those were for published work.

Next, we adhered to the rules in Fig. 1 to exclude articles that did not meet our inclusion criteria but had not been excluded by our automated search engine filters. This process involved two of the four authors (a postdoctoral research fellow and a doctoral student) carefully reading through the title, abstract, and, on occasion, the entire article to determine

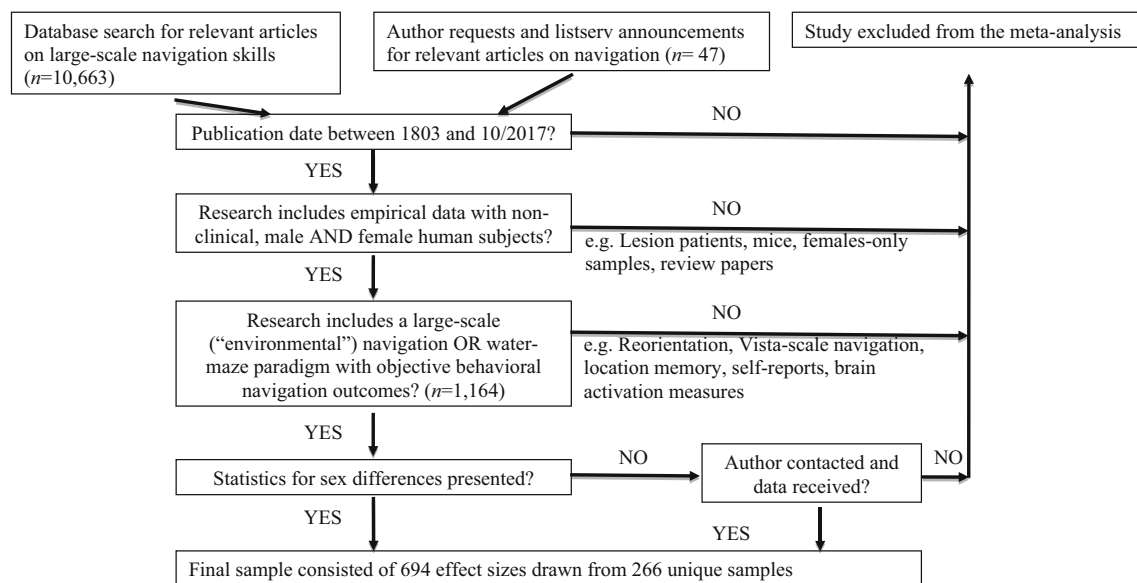
eligibility. These additional exclusions were papers that presented studies on nonhumans, robotics, unmanned devices, and self-report measures as well as tasks that involved line orientation, object location in a room/table/virtual environment, web navigation, language, haptic orientation, tactile, grating, locomotion, optic flow, categorization, spatial frequency, grasping, and reach planning, and so forth, that did not fit our definition of spatial navigation. Studies that had either male-only or female-only human participants, as well as special populations (for example, brain-damaged, hearing, or visually impaired individuals) were excluded from our sample, although data from healthy control groups were included. Finally, papers that did not present original research (e.g., review papers) were excluded from the analyses. The next step was to contact authors of articles that cleared our screening process, but that did not report the information required for us to compute effect sizes relevant to sex differences. Papers from authors who did not respond to our e-mails or who no longer had access to the data were excluded from the meta-analysis.

Our final sample had 694 effect sizes drawn from 266 samples. Out of the total effect sizes, 80 came from unpublished research in English (70 from dissertations, 10 from one unpublished paper). For the remainder, 614 effect sizes were from papers published in English, and no papers came from work published in other languages. Furthermore, 293 out of the 694 effect sizes (42.2%) originated from the United States, 135 effect sizes (19.5%) were from Germany, 73 were from Canada (10.5%), and 64 were from the United Kingdom (9.2%). The remaining effect sizes (18.6%) were from a variety of other countries. The final sample of studies included in the meta-analysis is presented in the [Supplemental Material](#).

### Inclusion criteria

The following criteria were used to determine the inclusion eligibility of a study in the meta-analysis:

1. The study should involve empirical research with nonclinical, male AND female human populations. For papers that investigated both nonhuman and human subjects, we included data for the human subjects only. For papers that included a healthy control when investigating clinical/special populations, we included data for the healthy controls only.
2. The study should include at least one objective navigation performance outcome in a task that met our definition of spatial navigation (presented in the Introduction). Importantly, this criterion meant that studies using only self-reports/surveys as measures of navigation performance were excluded.
3. The navigation paradigm should be categorized at an “environmental” scale (see Montello, 1993, for a description



**Fig. 1** Flowchart illustrating the selection process of articles included in the meta-analysis

of scales). At an environmental scale, a participant needs to move some distance within the space in order to obtain information about the spatial properties of the real or virtual environment (i.e., all spatial information cannot be obtained from a single vantage point in the environment). The only exception to this rule was the water maze task (e.g., Chamizo, Artigas, Sansa, & Banterla, 2011), as justified in the Introduction.

### Coding of study variables

As a starting point, we developed a coding template that captured crucial methodological aspects of the studies along with the moderators of interest. Our coding template had categories for a number of variables not necessarily considered as moderators to provide as complete a picture as possible for each study, with an eye on later data interpretation. Therefore, the coding template involved the following study characteristics: authors, year of publication, author ID, sample ID (a crucial variable for multilevel analysis), publication status, mean age of sample, experiment number, sample origin, number of males, number of females, study location, task goals, testing medium, outcome measure(s), route direction, timing conditions, perspective, locomotion, route direction, environment, familiarity, test interval, hints, feedback, device assistance, cues, and the calculated effect size(s). From this larger set, we considered the variables at sample level and measure level detailed in the Introduction section to identify factors that might moderate sex differences in wayfinding/navigation skills. Sample-level variables reflect variables inherent to the samples themselves, such as mean age. Measure-level

variables are those that are inherent to spatial navigation tasks, such as medium and outcome measure.

**Sample-level variables** Undergraduate students represented 53.5% of the total effect sizes, and children represented 9.7% of the total effect sizes. The remaining effect sizes represented a wide range of participant ages, justifying our use of mean age of the participants the moderator of interest as a continuous variable and a categorical variable (less than 13, 13–17, 18–29, 30–50, 50–older).

We also considered the sample-level variable year of publication. This variable is easily obtained and is routinely considered in meta-analyses of cognitive sex differences as a means to investigate Feingold's (1988) claim that cognitive sex differences are decreasing in magnitude in recent years.

**Measure-level variables** Table 1 presents a list of measure-level variables along with their respective categories. We have also included examples to clarify our classification process. In total, we had 14 measure-level variables consisting mostly of two to three groupings each. The task goals variable had eight categories, which was the highest number of categories in any variable. This large number of task goals testifies to the wide range of skills measured in the navigation literature.

To ensure coding validity, the detailed coding template mentioned earlier was used as a strong guideline. As an initial step before final coding, two of the authors of the present report (Coder 2 experienced in meta-analyses; Coder 1 experienced in navigation research and fully trained in meta-analytic coding by Coder 2) independently coded 25 studies accounting for 59 effects sizes for a total of 1,711 entries. This coding process involved 29 variables (again, not all considered in moderator analyses): authors, year of publication,

author ID, sample ID (a crucial variable for multilevel analysis), publication status, mean age of sample, experiment number, sample origin, number of males, number of females, study location, task goals, testing medium, outcome measure, route direction, timing conditions, perspective, locomotion, route selection, environment, familiarity, learning interval, guide assistance, feedback, device assistance, cues and the calculated effect size. This independent coding resulted in 173 disagreements, representing 1,538 agreements over 1,711 entries (29 variables  $\times$  59 effect sizes), for a .899 agreement rate (Cohen's kappa = .798). The two coders had extensive discussions to elucidate points of disagreement, and Coder 1 then proceeded with coding the remainder of the studies. At completion of coding, Coder 2 independently coded a new set of 50 studies (accounting for 77 effect sizes) from the final sample. In this case, the total of 2,233 entries (29 variables  $\times$  77 effect sizes) produced only nine disagreements, resulting in an interrater reliability of 99.6% (2,224 agreements/2,233 total entries; Cohen's kappa = .992). This high interrater reliability clearly reflects the validity of the final coding.

**Measure of effect size** The effect size measure was the standardized mean difference calculated as the mean for males minus that for females divided by the pooled standard deviation (Cohen's *d*; Cohen, 1988). The assumption is that men would perform better than women based on the literature presented so far. Thus, a positive effect size reflects a male advantage, and a negative effect size reflects a female advantage in spatial orientation tasks. The effect size calculation was based on Cohen's (1988) formula when means and standard deviations were available, which was the case for 372 out of the 694 effect sizes (53.6%). The remaining cases were available with an inferential statistic (typically *t* test, *p*, *r*, or *F*), thus, the formulae presented by Lipsey and Wilson (2001) were used. In all cases, effect sizes were computed by the calculator provided on David Wilson's webpage ([http://mason.gmu.edu/~dwilsonb/downloads/ES\\_Calculator.xls](http://mason.gmu.edu/~dwilsonb/downloads/ES_Calculator.xls)). Following recommendations by Hedges and Becker (1986), a small sample correction was applied to all effect sizes. When an effect size was not significant and no means or inferential statistics values were presented, authors were contacted by e-mail for more information. Out of the 19 authors who were contacted for that purpose, six replied and provided usable data. For the remainder, as suggested by Rosenthal (1991), we kept an effect size of zero to avoid excluding relevant work. Note, however, that in some of the cases in the table, zero was the actual effect size value.

## Data analysis

As is a typical goal of most meta-analyses focusing on sex differences, we aimed to quantify the overall magnitude of sex differences in spatial navigation and to identify variables that

might moderate these sex differences. A valid examination of specific tasks and potential moderators required us to retrieve multiple effect sizes that are nonindependent. Using these effect sizes in a fixed or random effects meta-analysis would violate the assumption that effect sizes should be independent (Borenstein et al., 2009), and this would distort the statistical analyses (Bateman & Jones, 2003). Accordingly, we relied on the multilevel linear modeling (MLM) approach to meta-analysis, as it does not require independence of effect sizes and it easily handles the type of hierarchical design represented in meta-analysis (Raudenbush & Bryk, 2002). As the standard error calculated for each effect size in a meta-analysis reflects an estimate of the variance for individual effect sizes (see Borenstein et al., 2009), multilevel meta-analysis represents a "variance-known" hierarchical linear model resulting in the precision weighted estimates of effect sizes typical of meta-analytic results (Raudenbush & Bryk, 2002).

As a starting point, and similar to the approach used by Voyer, Voyer, and Saint Aubin (2017), we computed an overall analysis and moderator analysis on the whole sample. The overall multilevel analysis was computed by examining the data organized in two levels: effect sizes nested within samples. This overall structure reflected 694 effect sizes (Level 1) nested within 266 samples (Level 2). This large number of Level 1 and Level 2 units had the advantage to maximize power for the identification of significant moderators and to provide a more complete documentation of the overall findings in the available data. The variables task, year of publication, mean age of the sample, age coded categorically, testing medium, outcome measure, route direction, timing conditions, perspective, route selection, environment, familiarity, learning interval, hints, feedback, device assistance, and cues were considered in the moderator analysis.

As a second step, after demonstrating that the different tasks differed at some basic level (as reflected in the finding that they produced effect sizes of different magnitude), we performed a moderator analysis separately for each task. This additional set of analyses was required to recognize the fact that some of the moderators might be confounded with task. For example, in most cases, pointing tasks generally produce a deviation score as an outcome variable. These analyses also used the multilevel approach considering that all the task groupings included nonindependent effect sizes.

All meta-analytic computations were performed with the *metafor* package in the R statistical software (Viechtbauer, 2010). Effect sizes were treated as random effects whereas moderators were treated as fixed effects. As previously mentioned, the observed values obtained in this approach reflect precision weighted estimates of effect sizes (Raudenbush & Bryk, 2002). In addition, significance testing for multilevel models used robust standard error for added precision, as they can be easily obtained with the *robust* command in the *metafor* package. However, as a small number of Level 2



clusters (i.e., samples) can bias the calculation of robust standard errors, the appropriate correction built-in the *robust* procedure in *metafor* was implemented. Note that, with *robust*, an *F* test is reported (instead of the more common between-groups *Q* test) when robust standard errors are used.

Categorical independent variables were dummy coded into  $k - 1$  dichotomous vectors (where  $k$  represents the number of categories) for consideration in the analysis, whereas continuous moderators were mean centered. In all moderator analyses, moderators were examined one at a time in models as there was no a priori basis to justify the examination of multifactor models or interactions. In addition, only effects significant with  $p < .05$  are presented in the Results section. This means that any moderator that is not mentioned in the results was nonsignificant.

## Results

A preliminary analysis was conducted to identify outliers. Following recommendations by Tabachnick and Fidell (2007), we defined outliers as effect sizes values that were more than 3.29 standard deviations away from the grand mean. Five outliers were identified based on this criterion. However, as such a number of outliers should be expected in a comprehensive sample, they were preserved as is for the sake of completeness, although they are identified by a star (\*) in Table 1. The final sample, therefore, consisted of 694 effect sizes drawn from 266 independent samples, reflecting combined results from 9,435 males and 9,570 females.

### Overall meta-analysis

**Overall sex differences in spatial navigation** A null model where the test of significance for the intercept is examined (Raudenbush & Bryk, 2002) provided data on the overall magnitude of sex differences in spatial navigation based on the current sample of studies. This analysis produced a mean estimated  $d$  of 0.341 (95% confidence interval (CI) [0.302, 0.380]), indicating that males significantly outperformed females on spatial navigation tasks,  $z = 17.05$ ,  $p < .001$ . Having considered this initial finding, it is important to remember that when authors reported sex differences as nonsignificant but provided no information for effect size coding, we entered an effect size of zero for these studies. Of course, we contacted authors to obtain clarifications but were still left with 62 cases where the effect size was coded as zero because of the lack of additional data. Accordingly, the estimate presented above might underestimate the actual magnitude of sex differences in spatial navigation. With this in mind, we removed these 62 effect sizes and conducted a second overall analysis. In this analysis, we found a mean estimated  $d$  of 0.381 (95% CI [0.340, 0.423]). Therefore, it might be more appropriate to

state that the true estimate of sex difference in spatial navigation is found within a range from 0.341 to 0.381. In any case, the remainder of the analyses preserved all 694 effect sizes in an attempt to provide a report on the complete data set. Regardless of which sample is used, however, results also showed that the overall effect was heterogeneous,  $Q(693) = 1,473.11$ ,  $p < .001$ ,  $I^2 = 50.4\%$  (for the complete data set). This fact suggests that the overall estimate of effect size is not representative of the sample of effect sizes. Accordingly, the examination of potential moderators was undertaken to attempt to account for this variability.

**Moderators of sex differences in the overall sample** The moderator analysis revealed that task goal accounted for significant variance in effect sizes,  $F(7, 258) = 4.38$ ,  $p < .001$ . Estimated effect sizes for this variable are presented in Table 2. The finding that none of the 95% confidence intervals contain zero indicates that a significant male advantage was observed for all task categories. In addition, multiple comparisons among means based on the robust standard errors and using the Tukey HSD method at the .05 level showed that recall and pointing tasks produced significantly larger effects than distance, learning, and verbal instructions tasks. No other differences achieved significance (all  $ps > .057$ ).

Outcome measures was also a significant moderator of the effect sizes,  $F(3, 262) = 7.04$ ,  $p < .001$ , with estimated effect sizes presented again in Table 2. As we have seen in all cases so far, none of the confidence intervals contained zero, indicating a significant male advantage in all categories. Tukey HSD multiple comparisons showed that deviation scores produced significantly larger effects than accuracy and distance measures, whereas response time produced larger effects than distance. No other differences achieved significance (all  $ps > .11$ ).

Timing condition was also found to contribute significantly to variance in effect sizes,  $F(1, 264) = 4.54$ ,  $p = .034$ , with estimated effect sizes also presented in Table 2. Examination of Table 2 indicates that, based on confidence intervals, both timed and untimed administration produce a significant male advantage, although it appears that the sex differences are larger for timed than for untimed tasks.

Environment produced a significant effect,  $F(3, 262) = 4.01$ ,  $p = .008$ , with estimated effect sizes presented in Table 2. As none of the confidence intervals contained zero, a significant male advantage was found in all categories. Tukey HSD multiple comparisons showed that water mazes and combination of indoor–outdoor environments produced significantly larger effects than studies that used exclusively indoor or outdoor environments. No other differences achieved significance (smallest  $p = .56$ ).

Finally, even though mean age of the participants as a continuous variable failed to account for significant variance ( $p = .94$ ), age defined categorically produced a significant

**Table 2** Summary for significant moderators in the overall meta-analysis

Moderator	Number of effects ( <i>k</i> )	Number of samples ( <i>N</i> )	Estimated mean <i>d</i>	95% confidence interval
<b>Task goal</b>				
Pointing	122	71	0.365	[0.294, 0.436]
Recall/recognition	298	131	0.386	[0.332, 0.439]
Cardinal directions	18	14	0.441	[0.160, 0.722]
Distance	14	10	0.229	[0.140, 0.318]
Landmark position	36	25	0.285	[0.176, 0.393]
Maps	62	31	0.349	[0.207, 0.491]
Verbal instructions	32	15	0.219	[0.102, 0.335]
Learning	112	63	0.247	[0.164, 0.329]
<b>Outcome measure</b>				
Accuracy	326	176	0.307	[0.254, 0.359]
Deviation	123	84	0.440	[0.375, 0.505]
Distance	96	66	0.250	[0.169, 0.331]
Response time	149	100	0.388	[0.302, 0.473]
<b>Timing conditions</b>				
Untimed	583	225	0.315	[0.268, 0.362]
Timed	111	47	0.459	[0.347, 0.572]
<b>Environment</b>				
Indoor	336	130	0.332	[0.275, 0.389]
Outdoor	329	132	0.336	[0.283, 0.389]
Indoor–outdoor	12	5	0.546	[0.383, 0.709]
Water maze	17	3	0.488	[0.380, 0.596]
<b>Age categories</b>				
Below 13	69	30	0.145	[0.055, 0.244]
13–17	24	7	0.577	[0.351, 0.803]
18–29	489	187	0.365	[0.317, 0.413]
30–49	81	34	0.314	[0.199, 0.430]
50 and above	31	10	0.423	[0.258, 0.588]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero. The fact that, in many cases, the total number of samples (*N*) shown in this table for each moderator exceeds the actual number of samples retrieved reflects the nonindependence of the effect sizes

contribution to variance,  $F(4, 261) = 4.74$ ,  $p = .001$ . Once more, the estimated mean effect sizes are presented in Table 2. Again, none of the confidence intervals contained zero, indicating a significant male advantage in all age categories. Tukey HSD multiple comparisons showed that studies sampling participants below the age of 13 produced significantly smaller effects than participants in all other age groups. In addition, the 30 to 49 category produced a smaller effect than the 13 to 17 category. No other differences achieved significance (all  $ps > .07$ ).

### Task subgroup analysis

As a starting point to the task subgroup analyses, we examined the overall homogeneity statistic within each subgroup to ensure that there was significant heterogeneity in the effect sizes

that required explanation by moderators before proceeding. Results of this analysis showed that learning and distance tasks goals produced homogeneous effect sizes ( $p = .330$ ,  $I^2 = 22.8\%$  for learning,  $p = .474$ ,  $I^2 = < 0.01\%$  for distance). Accordingly, the effect sizes presented in Table 2 for these two tasks goals can be considered representative of the state of affairs. However, for the remaining tasks, significant heterogeneity was observed (recall/recognition:  $p < .001$ ,  $I^2 = 47.4\%$ ; cardinal directions:  $p < .001$ ,  $I^2 = 75.8\%$ ; landmark position:  $p < .001$ ,  $I^2 = 53.9\%$ ; maps:  $p < .001$ ,  $I^2 = 71.3\%$ ; pointing:  $p < .001$ ,  $I^2 = 41.2\%$ ; verbal instructions:  $p = .015$ ,  $I^2 = 21.8\%$ ). Accordingly, moderator analyses proceeded separately for recall/recognition, cardinal directions, pointing, landmarks, maps, and verbal tasks, as presented in the following sections. However, to provide a clearer picture of what specific moderators accounted for significant variability, the

results are structured as a function of moderator in what follows.

**Perspective** Results showed that perspective was a significant moderator only in landmark position tasks,  $F(2, 11) = 9.21$ ,  $p = .001$ . Estimated means, presented in Table 3, show that only the route perspective category produced a significant male advantage. Tukey HSD comparisons showed that a route perspective produced larger effects than a combination of route-survey perspectives. No other comparisons produced significant differences among the mean estimated effect sizes (smallest  $p = .24$ ).

**Outcome measure** Outcome measure was a significant moderator for all the tasks goals examined in the subgroup analysis (largest  $p = .006$ ), with the estimates presented in Table 4. On pointing tasks, all the categories except response time indicated a significant male advantage. Tukey HSD comparisons showed that a deviation measures produced larger sex differences than response time and distance measures. No other differences achieved significance (smallest  $p = .054$ ).

For recall/recognition tasks, the estimates indicated a significant male advantage in all categories. Tukey HSD comparisons showed that response time measures produced a significantly larger effect than accuracy and distance measures. No other differences achieved significance (smallest  $p = .25$ ).

For cardinal direction tasks, all the categories except response time indicated a significant male advantage. Tukey HSD comparisons showed that accuracy and deviation measures produced larger sex differences than response time measures. No other differences achieved significance ( $p = .56$ ).

For map tasks, the estimates presented in Table 4 indicated a significant male advantage on accuracy and deviations scores, but not on distance and response-time measures. Tukey HSD comparisons showed that deviation measures produced significantly larger effects than distance measures. No other differences achieved significance (smallest  $p = .071$ ).

Estimates relevant to landmark position tasks showed that only the “other” category, combining one each for distance and response time measures, produced no significant male

**Table 3** Results of the analysis for perspective as a moderator

		<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Task goals	Categories			
Landmark position	Route	30	21	0.29 [0.15, 0.43]
	Survey	3	2	0.09 [−0.25, 0.42]
	Both	3	2	−0.05 [−0.13, 0.04]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero

**Table 4** Results of the analysis for outcome measure as a moderator

Task goal	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Pointing	Accuracy	9	7	0.37 [0.16, 0.58]
	Deviation	93	65	0.37 [0.28, 0.46]
	Distance	2	1	0.22 [0.17, 0.27]
	Response time	18	15	0.11 [−0.08, 0.30]
Recall/recognition	Accuracy	133	85	0.35 [0.27, 0.43]
	Deviation	14	12	0.41 [0.20, 0.63]
	Distance	60	43	0.28 [0.17, 0.39]
	Response time	91	57	0.52 [0.42, 0.63]
Cardinal direction	Accuracy	9	8	0.57 [0.29, 0.84]
	Deviation	4	3	0.69 [0.23, 1.14]
	Response time	5	5	0.21 [−0.03, 0.46]
Map	Accuracy	45	28	0.42 [0.21, 0.63]
	Deviation	4	2	0.67 [0.39, 0.95]
	Distance	8	5	0.23 [−0.29, 0.74]
	Response time	5	5	0.40 [−0.31, 1.22]
Landmark position	Accuracy	32	23	0.21 [0.08, 0.33]
	Deviation	2	1	0.57 [0.45, 0.69]
	Other	2*	2	0.77 [−0.03, 1.58]
Verbal instructions	Accuracy	18	14	0.23 [0.02, 0.43]
	Deviation	3	2	0.14 [−0.74, 1.01]
	Distance	3	2	−0.27 [−0.57, 0.03]
	Response time	8	8	0.23 [0.01, 0.45]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in parentheses. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero. \* Reflects one effect size each for distance and response time

advantage, despite seemingly large effects. Tukey HSD comparisons showed a larger effect for deviation than accuracy measures. There were no other significant differences among the effect sizes (smallest  $p = .15$ ).

Finally, on verbal instructions tasks, accuracy and response-time measures indicated a significant male advantage, whereas deviation and distance scores did not. Tukey HSD comparisons showed that accuracy and response time produced larger sex differences than distance measures did. No other differences achieved significance (smallest  $p = .22$ ).

**Route direction** Route direction was a significant moderator of effect sizes for pointing tasks,  $F(2, 68) = 3.19$ ,  $p = .048$ , and verbal instructions tasks,  $F(1, 13) = 19.93$ ,  $p < .001$ , with estimates presented in Table 5. For pointing, results showed a significant male advantage when participants went forward on the learned route or when such instructions were not applicable, but not for the backward route. Tukey HSD comparisons showed that a backward route produced significantly smaller effects than the “not applicable” category. No other difference achieved significance (smallest  $p = .09$ ).

**Table 5** Results of the analysis for route direction as a moderator

Task goals	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Pointing	Forward	102	59	0.33 [0.25, 0.41]
	Backward	5	5	0.15 [−0.04, 0.35]
	Not applicable	15	9	0.51 [0.30, 0.71]
Verbal instructions	Forward	29	13	0.12 [−0.04, 0.27]
	Not applicable	3	2	0.45 [0.41, 0.49]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero

For verbal instruction tasks, the estimates presented in Table 5 indicated a significant male advantage only for the “Not applicable” category. Direct interpretation of the estimated mean seen in Table 5 suggests that sex differences were larger when route direction was not applicable compared with when the forward direction was followed.

**Route selection** Route selection was only a significant moderator in cardinal direction tasks,  $F(3, 10) = 361.61$ ,  $p < .001$ . Relevant estimates, seen in Table 6, indicate that the male advantage was not significant only when a free choice was allowed. Tukey HSD multiple comparisons showed that shortcuts produced significantly larger effects than in all other groups. Free choice and exact way also produced smaller effects than when this moderator was not applicable. No other differences achieved significance (all  $ps > .055$ ).

**Timing condition** Timing conditions was a significant moderator in map tasks,  $F(1, 29) = 8.16$ ,  $p = .008$ , and verbal instructions tasks,  $F(1, 13) = 34.18$ ,  $p < .001$ , with estimates presented in Table 7. For map tasks, the results reflected the overall finding that although the male advantage was significant regardless of timing conditions, it was significantly larger when a timed administration was used. In contrast, on verbal instructions tasks, timed conditions resulted in a female advantage, whereas untimed conditions produced a male advantage.

**Table 6** Results of the analysis for route selection as a moderator

Task Goals	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Cardinal Directions	Free choice	3	2	0.11 [−0.26, 0.48]
	Exact way	2	2	0.46 [0.42, 0.51]
	Shortcut	2	1	1.07 [0.70, 1.44]
	Not applicable	11	9	0.53 [0.22, 0.84]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero

**Table 7** Results of the analysis for timing condition as a moderator

Task goals	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Map	Untimed	57	29	0.37 [0.17, 0.57]
	Timed	5	3	0.78 [0.51, 1.05]
Verbal instructions	Untimed	28	14	0.22 [0.07, 0.38]
	Timed	4	1	−0.20 [−0.34, −0.06]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero

**Environment** Environment was a significant moderator of effect sizes in pointing tasks,  $F(2, 68) = 9.60$ ,  $p < .001$ , and cardinal direction tasks,  $F(2, 11) = 18.34$ ,  $p < .001$ , with estimates presented in Table 8. In both tasks goals, the male advantage was significant regardless of environmental conditions. For pointing, Tukey HSD comparisons showed that testing in both environments (indoor–outdoor) produced significantly larger effects than testing indoor or outdoor singly. No other difference achieved significance (smallest  $p = .51$ ).

For cardinal direction tasks, Tukey HSD comparisons showed that water-maze environments produced significantly larger sex differences than indoor or outdoor environments did. The difference between these last two categories did not achieve significance ( $p = .65$ ).

**Familiarity** Familiarity was only a significant moderator for cardinal direction tasks,  $F(2, 11) = 22.57$ ,  $p < .001$ , with the estimates presented in Table 9, and a significant male advantage was found for all categories. Tukey HSD comparisons showed that familiar locations produced larger sex differences than unfamiliar ones did, whereas remaining differences did not achieve significance (smallest  $p = .07$ ).

**Feedback** The presence of feedback was a significant moderator for pointing tasks,  $F(2, 28) = 9.68$ ,  $p < .001$ , and map tasks,  $F(2, 28) = 32.56$ ,  $p < .001$ . Estimates are presented in

**Table 8** Results of the analysis for environment as a moderator

Task goals	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Pointing	Indoor	50	27	0.36 [0.24, 0.48]
	Outdoor	64	42	0.31 [0.22, 0.40]
	Indoor–outdoor	8	2	0.63 [0.51, 0.75]
Cardinal directions	Indoor	6	5	0.40 [0.04, 0.76]
	Outdoor	10	8	0.50 [0.21, 0.78]
	Water maze	2	1	1.07 [0.72, 1.42]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero



**Table 9** Results of the analysis for familiarity as a moderator

Task goals	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Cardinal direction	New	8	8	0.39 [0.13, 0.47]
	Learned	8	5	0.58 [0.22, 0.93]
	Familiar	2	1	1.09 [0.75, 1.25]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero

Table 10. For pointing, the estimated effect sizes indicated a significant male advantage in all categories. Tukey HSD comparisons showed that immediate feedback and the “not reported” category produced significantly larger effects than no feedback. The difference between the first two categories did not achieve significance ( $p = .43$ ).

For map tasks, a significant male advantage was found under immediate or no feedback, but not for the “not reported” categories. Tukey HSD comparisons showed that immediate feedback produced significantly larger effects than no feedback or the “not reported” category did. The difference between these last two categories did not achieve significance ( $p = .97$ ).

**Learning interval** Learning interval only produced a significant contribution to variance in recall/recognition tasks,  $F(3, 127) = 4.03$ ,  $p = .009$ . As seen in the estimates presented in Table 11, only immediate testing produced a significant male advantage when a delay was applied. Studies where the interval was not applicable also produced a significant male advantage. Formal Tukey HSD comparisons showed that effects in long time intervals were significantly smaller than those in immediate interval or the “not applicable” case. No other differences achieved significance (smallest  $p = .20$ ).

**Age** Age defined categorically was also a significant moderator for pointing tasks,  $F(3, 67) = 6.82$ ,  $p = .007$ , and landmark

**Table 10** Results of the analysis for feedback as a moderator

Task goals	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Pointing	Immediate	9	3	0.59 [0.42, 0.75]
	No feedback	107	62	0.31 [0.23, 0.39]
	Not reported	6	6	0.69 [0.49, 0.88]
Map	Immediate	4	2	0.94 [0.81, 1.07]
	No feedback	53	25	0.38 [0.16, 0.60]
	Not reported	5	5	0.37 [−0.01, 0.79]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero

**Table 11** Results of the analysis for learning interval as a moderator

Task Goals	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Recall/Recognition	Immediate	260	110	0.38 [0.32, 0.45]
	Short	4	3	0.42 [−0.13, 0.98]
	Long	2	2	0.04 [−0.14, 0.23]
	Not applicable	32	16	0.41 [0.26, 0.56]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero

position tasks,  $F(2, 22) = 7.53$ ,  $p = .003$ . Estimated mean effect sizes are presented in Table 12. For pointing, results indicated a significant male advantage in all the represented age categories except in the ages 30 to 49 category. Tukey HSD multiple comparisons showed that the 13 to 17 category produced significantly larger effects than all other groups. In addition, the less than 13 category produced a smaller effect than the 18 to 29 category. No other differences achieved significance (all  $ps > .16$ ).

In landmark position tasks, all the represented age categories showed a significant male advantage. Tukey HSD comparisons showed a significantly larger effect for the 18 to 29 and 30 to 49 age samples than for samples below the age of 13. The 18 to 29 and 30 to 49 groupings did not differ from each other ( $p = .71$ ).

## Publication bias and the file drawer problem

Despite our best efforts to obtain unpublished work, the present meta-analysis consists mostly of data obtained from published studies. In such a case, it is often assumed that the final sample might not be representative of the entire population of studies in existence (Rosenthal, 1979). Such a situation raises the possible influence of the “file-drawer problem” (Sterling, 1959), suggesting that studies producing nonsignificant

**Table 12** Results of the analysis for age categories as a moderator

Task goals	Categories	<i>k</i>	<i>N</i>	<i>d</i> [95% CI]
Pointing	Less than 13	6	3	0.12 [0.01, 0.24]
	13–17	5	5	0.69 [0.41, 0.97]
	18–29	100	52	0.37 [0.29, 0.46]
	30–49	11	9	0.20 [−0.04, 0.43]
Landmark position	Less than 13	3	3	0.03 [0.001, 0.05]
	18–29	26	18	0.27 [0.10, 0.45]
	30–49	7	4	0.32 [0.09, 0.55]

The table presents the number of effect sizes (*k*) and the mean weighted *d* for each moderator category with the 95% confidence interval (CI) in brackets. The mean weighted effect size is significantly different from zero with  $p < .05$  if the 95% CI for *d* does not include zero

results or in the unexpected direction (a female advantage in our case) have a lower probability of publication. This putative publication bias has the potential to affect any meta-analytic results so that by including mostly published studies, meta-analytic findings might exaggerate the magnitude of the effect under consideration.

The simplest way to examine the potential influence of the file-drawer problem is to compare the mean estimated effect sizes for samples obtained from published and unpublished research. We therefore proceeded with such an analysis in the overall sample. We also divided research into three rather than two categories. Of course, whether a data source was clearly published or unpublished formed two of the categories. However, theses and dissertation were considered as a third category because their status is uncertain in relation to the file-drawer problem. Specifically, one reason why they are not published might simply be that the author of the thesis did not pursue publication. This resulted in 614 published effect sizes (239 samples), 70 effect sizes from theses (24 samples), and only 10 effect sizes from unpublished sources (three samples). With this in mind, the analysis using publication status as moderator showed no significant influence of publication status,  $F(2, 263) = 0.72, p = .488$ . This suggests no evidence of a publication bias in the present sample.

One might sensibly argue that the small number of unpublished studies that we were able to obtain reduces the value of our examination of publication status as a moderator in testing for a publication bias. Based on this argument, and despite potential drawbacks, the Egger, Davey Smith, Schneider, and Minder (1997) approach was used as a further way to test a potential publication bias in the present sample. This method makes the assumption that studies with a small sample and a small effect size are less likely to get published (Borenstein et al., 2009). Therefore, in the presence of publication bias, plotting precision (the inverse of the variance;  $y$ -axis) against effect size ( $x$ -axis) would produce an asymmetrical distribution with few values on the bottom left-hand side of the plot, where small samples and negative effects would belong. Accordingly, the present data are shown in such a funnel plot in Fig. 2. A visual inspection of Fig. 2 reveals no sign of asymmetry. However, visual examination of the plot is not sufficient, and Egger et al. have formalized this process mathematically.

Specifically, the Egger et al. (1997) method allows examination of a possible publication bias by regressing the standard normal deviate for the effect size on precision. If there is no publication bias, the regression line should run through the origin and the intercept of the regression equation should not be significantly different from zero. Egger et al. recommended a significance level of .10 to maximize power. Following the approach proposed by Viechtbauer (see <http://stats.stackexchange.com/questions/155693/metafor-package-bias-and-sensitivity-diagnostics>), the inverse standard error

was used as a moderator in the multilevel analysis with the standard normal deviate for the effect sizes as outcome variable. Results of the Egger et al. test showed that the intercept was not significantly larger than zero at  $p < .10$ , with an intercept estimate of 0.065 (90% CI [-0.242, 0.375]). Therefore, the Egger et al.'s approach failed to support the presence of a publication bias in our data. This finding was further supported by results of a trim and fill analysis, an approach in which data points missing as a result of funnel plot asymmetry are imputed (see Duval & Tweedie, 2000). Specifically, this analysis indicated that no imputed data points were required and confirmed that the overall effect size remained unchanged from the original analysis.

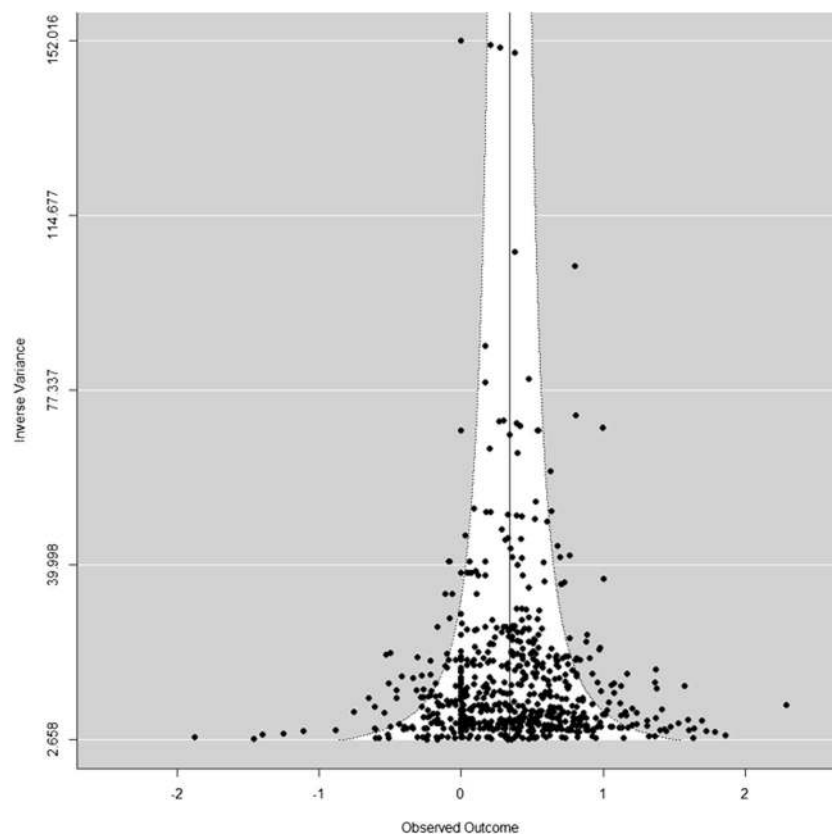
## Discussion

The present meta-analysis aimed to summarize the available literature on sex differences in human spatial navigation and to examine potential moderators of these effects. The results of the analyses are summarized in Table 13. These results should be seen as a guide to forward thinking in this area, although it is important to remember that, because of their quasi-experimental nature, meta-analytic results do not allow causal conclusions. Accordingly, our discussion provides speculations intended to stimulate empirical assessment in future work.

## Overall results

The overall effect for our 694 effect sizes was estimated as  $d$  of 0.34 (up to 0.38 if effects sizes coded as zero are omitted). In Cohen's (1988) classification, this effect would be considered small to medium. Interestingly, it is in line with the overall effect reported by Voyer, Voyer, and Bryden (1995) in their examination of small-scale spatial ability sex differences (overall  $d$  of 0.37). In addition, the effect size is in line with the range of sex differences reported in a large international study using the game *Sea Hero Quest* (Coutrot et al., 2018). This study had an extremely large sample size, but was reliant on self-reports of sex and was vulnerable to factors such as more than one player using the game. Thus, alignment in those results with those in the meta-analysis supports the validity of the conclusion that there are sex differences, albeit of small to moderate size.

Furthermore, results showed that a publication bias is unlikely to account for the present findings. To put these estimates in context, consider that for a  $d$  of 0.34, the distribution of men and women in spatial navigation overlaps by approximately 86%, which, at first glance, might seem to be a large overlap. However, it also means that about 64% of men score above the mean of women (and therefore, 36% of men scored below the mean of women), and thus the difference is



**Fig. 2** Funnel plot of precision (inverse variance) as a function of Cohen's  $d$  (observed outcome) for the whole sample

arguably not negligible. In a more practical fashion, this means that in any pair composed of a randomly selected man and a randomly selected woman, the man would have about a 64% of probability to score above the woman, and a 36% probability to score below her. Navigation training in regular schooling or in informal activities might be helpful to narrow sex differences, in view of the malleability of spatial abilities more generally (Uttal et al., 2013), although navigation-specific training needs development. There have only been a few efforts to improve mapping and wayfinding skills, although what prior efforts exist have seen some success (Kastens & Liben, 2007; Nazareth, Newcombe, Shipley, Velazquez & Weisberg, *in press*). The goal of such training would be not only to narrow the sex gap but also to benefit individuals of both sexes in the long term.

### Moderator analyses

The discussion of moderator analyses builds on Table 13. Readers should refer to Tables 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13 for the relevant summary values.

**Task goal** Task goal was a significant moderator of sex differences in navigation skills. Although the numerically largest effects were found in cardinal direction tasks (based on a relatively small number of effect sizes; see Table 2), statistical

results showed only that recall and pointing tasks produced significantly larger effects than distance, learning, and verbal instructions tasks, with no other differences achieving significance. The smaller sex difference for distance tasks might reflect a floor effect as both men and women tend to perform poorly on such tasks as a result of their inherent difficulty (Newcombe, 1985). For verbal instructions, the smaller sex difference might reflect the female advantage in language tasks and courses (Voyer & Voyer, 2014), providing them with some amount of compensation in relation to men's performance. Finally, for learning tasks, although the male advantage is significant, the reduced magnitude of the effect size might reflect the possibility that women can catch up with men in paradigms that allow for learning.

**Outcome measure** Outcome measure was a significant moderator of sex differences in navigation skills in the overall analysis (see Table 2). In addition, it was a significant moderator for all task goals considered in the subgroup analysis (see Table 4). For all task goals except for verbal instructions, deviation scores produced the largest male advantage, perhaps because deviation scores provide an unusually sensitive measure of precision and accuracy. For verbal instructions, accuracy and response time produced the largest male advantage. Although distance measures did not produce a significant sex difference under verbal instructions, the fact that they reflect a

**Table 13** Summary of results for significant moderators

Moderator	Significant on:	Comments
Perspective	Landmark position	Only route view produced a significant male advantage, not survey or both combined.
Outcome variable	Overall, recall, maps, pointing, cardinal directions, landmark position, verbal instructions	Deviation measures generally larger than other scores, except on verbal. Aside from this finding, the specific pattern of difference among means varies depending on the task.
Route direction	Pointing, verbal	Significant male advantage found for retracing the learned route or when not applicable in pointing; not significant for learned route in verbal.
Route selection	Cardinal directions	Male advantage significant in all cases, except for a free choice.
Timing conditions	Overall, map, verbal	Male advantage larger with timed rather than untimed conditions in overall and map, female advantage for timed in verbal.
Environment	Overall, pointing, cardinal directions	Sex effect significant in all categories. For pointing and overall, the “both” category produced larger effects than indoor and outdoor environments separately; for cardinal directions and overall, water maze also produced larger effects than single environments.
Familiarity	Cardinal directions	Significant male advantage in all categories but larger for familiar than new locations.
Feedback	Maps, pointing	Immediate feedback and “not reported” produced significantly larger effects than no feedback.
Learning interval	Recall	Immediate testing and studies where it was not relevant produced a significant male advantage, not longer time intervals.
Age categories	Overall, pointing, landmark position	Fine details vary across tasks but the <13 grouping tends to produce the smallest sex differences.

trend favoring women is intriguing, especially considering that this is one of only three effect sizes with a negative sign in the tables of results. Conclusions on this finding are limited by the small number of relevant effect sizes ( $k = 3$ ; see Table 4). However, the investigation of distance measures with verbal instructions tasks might provide a fruitful avenue in future work to better understand why women might process navigation tasks efficiently in this context.

**Route direction** Route direction was a significant moderator for pointing and verbal instruction task goals (see Table 5). However, the results with route direction are not particularly informative because the reverse route condition was not used as manipulation with verbal instructions. For pointing, a similar problem arises, with only five of 107 studies using a reversed direction. The findings show that sex differences are small and not significant under such conditions. One possible way to account for this finding is that the additional cognitive load required for a route followed in the reversed direction might reach a point where it even exceeds the men’s ability to handle the extra information, resulting in floor effects.

**Timing condition** Timing condition was a significant moderator of sex differences in navigation skills in the overall sample (see Table 2). It was also significant in maps and verbal instructions tasks (see Table 7). Except for verbal instructions, the male advantage was larger under timed than under untimed conditions. The role of timing condition might suggest the presence of sex differences in speed of processing for complex spatial information. However, it is important to keep in mind that timed tasks accounted for only 16% of effect sizes in the overall sample (111 out of 694) and only 8% for map tasks (five out of 62). Furthermore, the findings for verbal instructions are the opposite of what would be expected, with a significant female advantage observed for timed tasks, and a significant male advantage for untimed conditions. The findings for verbal instructions are questionable, however, because all four effect sizes for the timed condition come from the same sample in one study (Ishikawa & Kiyomoto, 2008). It is also important to consider that all effect sizes for distance and pointing were untimed, whereas for other categories the percentage of timed effect sizes was as follows: recall/recognition = 29.5%; cardinal directions = 27.8%; landmark position = 13.9%; learning = 3.6%; and verbal instructions = 12.5%. The small number of effect sizes from timed tasks in most of these task categories suggests that it is premature to draw strong conclusions concerning the role of timing conditions on sex differences in spatial navigation.

**Environment** Environment was a significant moderator of sex differences in the overall sample as well as for pointing tasks (see Tables 3 and 9). Sex differences were largest when testing involved a combination of indoor and



outdoor environments. Water-maze tasks stood out for producing larger sex differences than indoor environments or outdoor environments. There were no significant differences in effects between the water maze and the combined indoor–outdoor environments.

The large sex differences in combined indoor–outdoor environments might reflect the complexity of such tasks, adding to instances of high task complexity promoting larger sex differences in spatial tasks (Coluccia & Louse, 2004; Heil & Jansen-Osmann, 2008). Environments combining indoor and outdoor settings likely involve switching between egocentric and allocentric wayfinding strategies. Therefore, large sex differences in this context might partly reflect the male advantage in the ability to alternate between strategies (Wang & Carr, 2014).

**Feedback** Feedback significantly moderated task goals, specifically for maps and pointing tasks (see Table 10). Immediate feedback increased the magnitude of the sex difference. However, the finding that the largest effect for the pointing tasks is for the “not reported” category undermines this conclusion. In fact, it is clear that more research manipulating feedback is required considering that studies with immediate feedback reflected only 6.5% of effect sizes for maps and 7.4% for pointing.

**Age** Age was a significant moderator of sex differences in the overall sample (see Table 2) as well as in pointing and landmark position tasks (see Table 11). It is readily apparent from the data presented in the tables that the less than 13 years age group produced the smallest effect sizes, with a clear increase in magnitude for the 13 to 17 years category. Considering that adolescence is a time for increased independent navigational range (Anooshian & Young, 1981) with sex differences in how far and frequently children travel away from home, experiential and social norms may play a role in promoting sex differences in navigation skills in the latter age group. However, it is important to note that the 18 to 29 years old group is overrepresented in the retrieved literature, reflecting 70.5% of the effect sizes examined here (see Table 2). The 13 to 17 years old group is interesting because it produced the largest effect size presented in Table 2; however, it is also the grouping with the smallest sample ( $k = 24$ ). These data emphasize the need for more life-span developmental research on spatial navigation.

**Moderators significant in only one task** Finally, a number of moderators only accounted for significant variance for one of the task categories. For instance, perspective accounted for significant variance only in landmark position tasks (see Table 3.) However, this finding might once again reflect a limited number of studies and does not warrant a lengthy discussion.

Route selection only achieved significance for cardinal direction (see Table 6), showing that the male advantage was largest when a shortcut was required as part of the task response. In fact, the effect size of 1.07 for that category is the largest in all our tables. Speculatively, the use of a shortcut might require deeper processing of the route and results in a better cognitive map, suggesting that higher depth of processing advantages males. However, this reflects another case of a result based on a small number of effect sizes ( $k = 2$ ). The most parsimonious conclusion here is therefore that this is a finding that requires replication in many more studies before efforts are expanded to explain it.

Familiarity with the testing environment moderated cardinal directions (see Table 9), showing a significantly larger male advantage for familiar compared with new locations. This finding could be a side effect of the very small number of effect sizes ( $k = 2$ ) for familiar settings in cardinal direction tasks.

Finally, learning interval was a significant moderator only for recall/recognition tasks (see Table 11). Longer time intervals did not produce a significant male advantage, whereas immediate time intervals and testing where learning interval was not relevant did. We can speculate that, for longer intervals, the memory load exceeded even the male’s abilities and produced a floor effect. Short time intervals did not produce a significant male advantage despite a medium effect size ( $d = 0.42$ ). However, this category had a very broad confidence interval as a result of imprecise estimates accounted for by small sample sizes. Nevertheless, in terms of actual magnitude, short intervals produced similar effects to what we found with immediate recall, thereby supporting the memory load account to some extent.

### Nonsignificant moderators of special importance

Although we found that a large number of variables were significant moderators of sex differences in spatial navigation, some of the moderators failed to achieve significance despite our expectations. A few of those are particularly noteworthy because of their theoretical or practical implications. Specifically, despite the theoretical importance often assigned to cue types (proximal, distal; Padilla, Creem-Regehr, Stefanucci, & Cashdan, 2017), this moderator was not significant. This finding contradicts the assumption of differential hippocampal function in men and women in processing environmental cues. This finding is consistent with our earlier finding related to the environment moderator. Specifically, it is reasonable to assume that an indoor environment (e.g., closed basement maze) may offer more proximal cues and fewer distal cues in comparison to an outdoor environment (e.g., university campus). Given that there were no significant differences in effects found between indoor or outdoor testing environments, the null effect of cues should not come as a

surprise. However, the significantly larger sex differences in combined indoor–outdoor settings may point to sex differences in flexibility in cue processing rather than the ability to use one or the other cue and should be further investigated.

It is also noteworthy that year of publication failed to account for significant variability in effect sizes both in the overall sample and in the separate analysis for each task goals. On the surface, this suggests that the magnitude of sex differences in spatial navigation is unaffected by social changes associated with year of publication (e.g., Feingold, 1988). However, in considering this finding, it is also crucial to keep in mind that year of publication reflected a limited range in our sample, from 1977 to 2018 (at least limited from a statistical perspective), despite our search parameters including research published since 1803 (the lower limit for the search engine by default). This range limitation would have adverse effects on the likelihood of obtaining a significant relation between year of publication and the magnitude of effect sizes, as is always the case in correlational designs (Tabachnick & Fidell, 2007). Accordingly, it would be premature to draw definite conclusions on the current null finding for the moderating effect of year of publication on sex differences in spatial navigation.

## Limitations

Of course, any comprehensive meta-analysis is not without limitations. In particular, throughout the discussion section we mentioned moderator categories where there were too few effect sizes to allow solid conclusions. All these cases reflect areas that require more research and should encourage researchers to direct their efforts to elucidate factors accounting for sex differences in human spatial navigation. In particular, more studies examining cardinal direction and distance task goals might be warranted considering that these are the categories where there are the fewest effect sizes in our sample (see Table 2). Similarly, it might be worthwhile to conduct more studies comparing timed and untimed conditions in the same experiment.

Future research should also consider manipulating the amount of feedback provided to participants and examine sex differences in improvement in navigation performance that may result from this manipulation. The inability of low-performers to use feedback to self-correct one's cognitive map of the environment may present an opportunity for training interventions. Finally, large-scale navigation research has been abundantly tested with undergraduate psychology students who represent not only a very specific demographic but also a specific stage in neural development. The emergence of sex difference around the age of 13 years emphasizes the need for more developmental research in the 13 to 17 year age group.

The fact that we were only able to find a small number of effect sizes from clearly unpublished work ( $k = 10$ ) is also a

limitation of our analysis, although it is a very common problem for meta-analysis. We found statistical reassurance in the findings that publication status was not a significant moderator and the Egger et al.'s (1997) test produced no evidence of publication bias. These findings are most likely a consequence of our sampling of much research that did not aim primarily at examining sex differences in spatial navigation. Accordingly, we are quite confident that our results are a valid reflection of the current state of affairs for sex differences in human spatial navigation.

## Conclusions

The present meta-analysis provided the first comprehensive quantitative review of sex differences in human spatial navigation. Overall, the take home message from our results is that the male advantage in human spatial navigation is small to moderate and does not vary that much, with few exceptions. It is particularly noteworthy that the effect sizes were generally small for children. This observation is congruent with a recent meta-analysis of the development of sex-differences in mental rotation (Lauer, Yhang, & Lourenco, 2019). Of course, an increasing male advantage with age could reflect either emerging biological constraints or the cumulative effect of environmental opportunities and expectations. But the pattern constrains the search for causal explanations, in that any satisfactory explanation needs to predict the pattern. For example, had this pattern not been observed, an environmental explanation would have seemed less persuasive. Future work should investigate fine-grained hypotheses, such as relations of daily exploration patterns to growth in wayfinding skills. Cross-cultural research can provide a special purchase on these questions, given that gender roles vary culturally. Studies along these lines are beginning to appear, both using anthropological techniques (e.g., Davis & Cashdan, *in press*; Wood et al., 2019) and the large-scale use of wayfinding games (Coutrot et al., 2018).

An important observation arising from our results is that, in many cases, significant effects of moderators, especially when occurring in task subgroups, were compromised by small numbers of effect sizes. When such findings were theoretically unanticipated and did not have clear interpretations, they should be seen only as potentially intriguing but preliminary. Another caveat is that some of the effects of moderators may arise in cases where testing conditions or task factors promoted floor or ceiling effects, and these were often compounded by the presence of few effect sizes in the sample. Such findings create “effects” that are theoretically uninteresting and may make a male advantage either more or less pronounced.

We have emphasized in the discussion the cases that require more empirical research either on their own right or to address issues relevant to small sample sizes or floor/ceiling

effects. In this way, we hope that the research presented here will allow researchers to investigate promising avenues in their future work on spatial navigation and in their efforts to document how performance in such tasks is affected by sex.

**Acknowledgements** This meta-analytic project was funded by the Natural Sciences and Engineering Research Council of Canada grant (RGPIN-2016-03635) awarded to Dr. Daniel Voyer and the NSF Spatial Intelligence and Learning Center grant (SBE -1041707) awarded to Dr. Nora Newcombe. The authors have no competing interests in the current manuscript.

## References

- Articles included in the meta-analysis are denoted by a star (\*) and they are cited in supplemental material.
- \*Abu-Obeid, N. (1998). Abstract and scenographic imagery: The effect of environmental form on wayfinding. *Journal of Environmental Psychology, 18*, 159–173. <https://doi.org/10.1006/jevp.1998.0082>
  - \*Acevedo, S. F., Piper, B. J., Craytor, M. J., Benice, T. S., & Raber, J. (2010). Apolipoprotein E4 and sex affect neurobehavioral performance in primary school children. *Pediatric Research, 67*, 293–299. <https://doi.org/10.1203/PDR.0b013e3181cb8e68>
  - Acredolo, L. P., Pick, H. L., & Olsen, M. G. (1975). Environmental differentiation and familiarity as determinants of children's memory for spatial location. *Developmental Psychology, 11*, 495–501. <https://doi.org/10.1037/h0076667>
  - \*Allahyar, M. (2003). *Individual differences in the use of strategy in spatial orientation: Acquiring route and configural knowledge in virtual environments* (Order No. 3090956). Available from ProQuest Dissertations & Theses Global. (305274467)
  - Allen, G. L., Kirasic, K. C., Siegel, A. W., & Herman, J. F. (1979). Developmental issues in cognitive mapping: The selection and utilization of environmental landmarks. *Child Development, 50*, 1062–1070. <https://doi.org/10.2307/1129332>
  - \*Allen, G. L., & Willenborg, L. J. (1998). The need for controlled information processing in the visual acquisition of route knowledge. *Journal of Environmental Psychology, 18*, 419–427. <https://doi.org/10.1006/jevp.1998.0079>
  - \*Allison, C., Redhead, E. S., & Chan, W. (2017). Interaction of task difficulty and gender stereotype threat with a spatial orientation task in a virtual nested environment. *Learning and Motivation, 57*, 22–35. <https://doi.org/10.1016/j.lmot.2017.01.005>
  - Anooshian, L. J., & Young, D. (1981). Developmental changes in cognitive maps of a familiar neighborhood. *Child Development, 52*, 341–348. <https://doi.org/10.2307/1129248>
  - \*Astur, R. S., Purton, A. J., Zaniwski, M. J., Cimadevilla, J., & Markus, E. J. (2016). Human sex differences in solving a virtual navigation problem. *Behavioural Brain Research, 308*, 236–243. <https://doi.org/10.1016/j.bbr.2016.04.037>
  - Astur, R. S., Ortiz, M. L., & Sutherland, R. J. (1998). A characterization of performance by men and women in a virtual Morris water task: A large and reliable sex difference. *Behavioural Brain Research, 93*, 185–190. [https://doi.org/10.1016/S0166-4328\(98\)00019-9](https://doi.org/10.1016/S0166-4328(98)00019-9)
  - Astur, R. S., Taylor, L. B., Mamelak, A. N., Philpott, L., & Sutherland, R. J. (2002). Humans with hippocampus damage display severe spatial memory impairments in a virtual Morris water task. *Behavioural Brain Research, 132*, 77–84. [https://doi.org/10.1016/S0166-4328\(01\)00399-0](https://doi.org/10.1016/S0166-4328(01)00399-0)
  - Astur, R. S., Tropp, J., Sava, S., Constable, R. T., & Markus, E. J. (2004). Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation. *Behavioural Brain Research, 151*, 103–115. <https://doi.org/10.1016/j.bbr.2003.08.024>
  - \*Bakdash, J. Z. (2010). *Guided navigation impairs spatial knowledge: Using aids to improve spatial representations* (Order No. 3451456). Available from ProQuest Dissertations & Theses Global. (862324878)
  - Barkley, C. L., & Gabriel, K. I. (2007). Sex differences in cue perception in a visual scene: Investigation of cue type. *Behavioral Neuroscience, 121*, 291–300. <https://doi.org/10.1037/0735-7044.121.2.291>
  - \*Barrash, J. (1994). Age-related decline in route learning. *Developmental Neuropsychology, 10*, 189–201. <https://doi.org/10.1080/87565649409540578>
  - \*Baskaya, A., Wilson, C., & Özcan, Y. Z. (2004). Wayfinding in an unfamiliar environment: Different spatial settings of two polyclinics. *Environment and Behavior, 36*, 839–867. <https://doi.org/10.1177/0013916504265445>
  - \*Basten, K., Meilinger, T., & Mallot, H. A. (2012). Mental travel primes place orientation in spatial recall. In C. Stachniss, K. Schill, & D. Uttal (Eds.), *Spatial cognition VIII: Lecture notes in computer science* (Vol. 7463, pp. 378–385). Berlin, Germany: Springer.
  - Bateman, I., & Jones, L. P. (2003). Contrasting conventional with multi-level modeling approaches to meta-analysis: Expectation consistency in UK woodland recreation values. *Land Economics, 79*, 235–258. <https://doi.org/10.2037/3146869>
  - \*Blacker, K. J., Weisberg, S. M., Newcombe, N. S., & Courtney, S. M. (2017). Keeping track of where we are: Spatial working memory in navigation. *Visual Cognition, 25*, 691–702. <https://doi.org/10.1080/13506285.2017.1322652>
  - \*Blanch, R. J., Brennan, D., Condon, B., Santosh, C., & Hadley, D. (2004). Are there gender-specific neural substrates of route learning from different perspectives?. *Cerebral Cortex, 14*, 1207–1213. <https://doi.org/10.1093/cercor/bhh081>
  - Boccia, M., Nemmi, F., & Guariglia, C. (2014). Neuropsychology of environmental navigation in humans: Review and meta-analysis of fMRI studies in healthy participants. *Neuropsychology Review, 24*, 236–251. <https://doi.org/10.1007/s11065-014-9247-8>
  - \*Boone, A. P., Gong, X., & Hegarty, M. (2018). Sex differences in navigation strategy and efficiency. *Memory & Cognition, 46*, 909–922. <https://doi.org/10.3758/s13421-018-0811-y>
  - Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2009). *Introduction to meta-analysis*. <https://doi.org/10.1002/9780470743386>
  - \*Bosco, A., Longoni, A. M., & Vecchi, T. (2004). Gender effects in spatial orientation: Cognitive profiles and mental strategies. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 18*, 519–532. <https://doi.org/10.1002/acp.1000>
  - Brake, W. G., & Lacasse, J. M. (2018). Sex differences in spatial navigation: The role of gonadal hormones. *Current opinion in behavioral sciences, 23*, 176–182. <https://doi.org/10.1016/j.cobeha.2018.08.002>
  - \*Broadbent, H. J., Farran, E. K., & Tolmie, A. (2014). Egocentric and allocentric navigation strategies in Williams syndrome and typical development. *Developmental Science, 17*, 920–934. <https://doi.org/10.1111/desc.12176>
  - \*Broadbent, H. J., Farran, E. K., & Tolmie, A. (2015). Sequential egocentric navigation and reliance on landmarks in Williams syndrome and typical development. *Frontiers in Psychology, 6*, 216. <https://doi.org/10.3389/fpsyg.2015.00216>
  - \*Brunyé, T. T., Gardony, A., Mahoney, C. R., & Taylor, H. A. (2012). Going to town: Visualized perspectives and navigation through virtual environments. *Computers in Human Behavior, 28*, 257–266. <https://doi.org/10.1016/j.chb.2011.09.008>



- \*Burigat, S., & Chittaro, L. (2007). Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human-Computer Studies*, 65, 945–958. <https://doi.org/10.1016/j.ijhcs.2007.07.003>
- \*Burkitt, J., Widman, D., & Saucier, D. M. (2007). Evidence for the influence of testosterone in the performance of spatial navigation in a virtual water maze in women but not in men. *Hormones and Behavior*, 51, 649–654. <https://doi.org/10.1016/j.yhbeh.2007.03.007>
- \*Burte, H., & Montello, D. R. (2017). How sense-of-direction and learning intentionality relate to spatial knowledge acquisition in the environment. *Cognitive Research: Principles and Implications*, 2, 18. <https://doi.org/10.1186/s41235-017-0057-4>
- Cánovas, R., García, R. F., & Cimadevilla, J. M. (2011). Effect of reference frames and number of cues available on the spatial orientation of males and females in a virtual memory task. *Behavioural Brain Research*, 216, 116–121. <https://doi.org/10.1016/j.bbr.2010.07.026>
- \*Cashdan, E., Marlowe, F. W., Crittenden, A., Porter, C., & Wood, B. M. (2012). Sex differences in spatial cognition among Hadza foragers. *Evolution and Human Behavior*, 33, 274–284. <https://doi.org/10.1016/j.evolhumbehav.2011.10.005>
- \*Castelli, L., Corazzini, L. L., & Geminiani, G. C. (2008). Spatial navigation in large-scale virtual environments: Gender differences in survey tasks. *Computers in Human Behavior*, 24, 1643–1667. <https://doi.org/10.1016/j.chb.2007.06.005>
- Chai, X. J., & Jacobs, L. F. (2009). Sex differences in directional cue use in a virtual landscape. *Behavioral Neuroscience*, 123, 276–283. <https://doi.org/10.1037/a0014722>
- \*Chai, X. J., & Jacobs, L. F. (2010). Effects of cue types on sex differences in human spatial memory. *Behavioural Brain Research*, 208, 336–342. <https://doi.org/10.1016/j.bbr.2009.11.039>
- \*Chamizo, V. D., Artigas, A. A., Sansa, J., & Banterla, F. (2011). Gender differences in landmark learning for virtual navigation: The role of distance to a goal. *Behavioural Processes*, 88, 20–26. <https://doi.org/10.1016/j.beproc.2011.06.007>
- \*Chen, C. H., Chang, W. C., & Chang, W. T. (2009). Gender differences in relation to wayfinding strategies, navigational support design, and wayfinding task difficulty. *Journal of Environmental Psychology*, 29, 220–226. <https://doi.org/10.1016/j.jenvp.2008.07.003>
- \*Choi, J., McKillop, E., Ward, M., & L'Hirondelle, N. (2006). Sex-specific relationships between route-learning strategies and abilities in a large-scale environment. *Environment and Behavior*, 38, 791–801. <https://doi.org/10.1177/0013916506287004>
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd). Hillsdale: Erlbaum.
- Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, 24, 329–340. <https://doi.org/10.1016/j.jenvp.2004.08.006>
- \*Coluccia, E., Iosue, G., & Brandimonte, M. A. (2007). The relationship between map drawing and spatial orientation abilities: A study of gender differences. *Journal of Environmental Psychology*, 27, 135–144. <https://doi.org/10.1016/j.jenvp.2006.12.005>
- \*Corbin, I. R. (2014). *Sex differences in wayfinding in a life-sized, multi-level environment* (Unpublished dissertation). Royal Military College of Canada, Kingston.
- \*Cornell, E. H., Heth, C. D., & Broda, L. S. (1989). Children's wayfinding: Response to instructions to use environmental landmarks. *Developmental Psychology*, 25, 755–764. <https://doi.org/10.1037/0012-1649.25.5.755>
- \*Cornell, E. H., Heth, C. D., & Rowat, W. L. (1992). Wayfinding by children and adults: Response to instructions to use look-back and retrace strategies. *Developmental Psychology*, 28, 328–336. <https://doi.org/10.1037/0012-1649.28.2.328>
- Coutrot, A., Silva, R., Manley, E., De Cothi, W., Sami, S., Bohbot, V. D., . . . Spiers, H. J. (2018). Global determinants of navigation ability. *Current Biology*, 28(17), 2861–2866.
- Cubukcu, E., & Nasar, J. L. (2005). Relation of physical form to spatial knowledge in largescale virtual environments. *Environment and Behavior*, 37, 397–417. <https://doi.org/10.1177/0013916504269748>
- \*Cutmore, T. R., Hine, T. J., Maberly, K. J., Langford, N. M., & Hawgood, G. (2000). Cognitive and gender factors influencing navigation in a virtual environment. *International Journal of Human-Computer Studies*, 53, 223–249. <https://doi.org/10.1006/ijhc.2000.0389>
- Dabbs, J. M., Jr, Chang, E. L., Strong, R. A., & Milun, R. (1998). Spatial ability, navigation strategy, and geographic knowledge among men and women. *Evolution and Human Behavior*, 19, 89–98. [https://doi.org/10.1016/S1090-5138\(97\)00107-4](https://doi.org/10.1016/S1090-5138(97)00107-4)
- \*Dahmani, L., Ledoux, A. A., Boyer, P., & Bohbot, V. D. (2012). Wayfinding: The effects of large displays and 3-D perception. *Behavior Research Methods*, 44, 447–454. <https://doi.org/10.3758/s13428-011-0158-9>
- Daugherty, A. M., Bender, A. R., Yuan, P., & Raz, N. (2015). Changes in search path complexity and length during learning of a virtual water maze: Age differences and differential associations with hippocampal subfield volumes. *Cerebral Cortex*, 26, 2391–2401. <https://doi.org/10.1093/cercor/bhv061>
- \*Daugherty, A. M., & Raz, N. (2017). A virtual water maze revisited: Two-year changes in navigation performance and their neural correlates in healthy adults. *NeuroImage*, 146, 492–506. <https://doi.org/10.1016/j.neuroimage.2016.09.044>
- \*Davies, C. (2002). When is a map not a map? Task and language in spatial interpretation with digital map displays. *Applied Cognitive Psychology*, 16, 273–285. <https://doi.org/10.1002/acp.786>
- Davis, H. E., & Cashdan, E. (in press). Spatial cognition, navigation, and mobility among children in a forager-horticulturalist population, the Tsimane of Bolivia. *Cognitive Development*.
- \*Devlin, A. S., & Bernstein, J. (1995). Interactive wayfinding: Use of cues by men and women. *Journal of Environmental Psychology*, 15, 23–38. [https://doi.org/10.1016/0272-4944\(95\)90012-8](https://doi.org/10.1016/0272-4944(95)90012-8)
- \*Driscoll, I., Hamilton, D. A., Yeo, R. A., Brooks, W. M., & Sutherland, R. J. (2005). Virtual navigation in humans: The impact of age, sex, and hormones on place learning. *Hormones and Behavior*, 47, 326–335. <https://doi.org/10.1016/j.yhbeh.2004.11.013>
- \*Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040–1048. <https://doi.org/10.1037/0003-066X.41.10.1040>
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634. <https://doi.org/10.1136/bmj.315.7109.629>
- \*Farran, E. K., Courbois, Y., Van Herwegen, J., & Blades, M. (2012a). How useful are landmarks when learning a route in a virtual environment? Evidence from typical development and Williams syndrome. *Journal of Experimental Child Psychology*, 111, 571–586. <https://doi.org/10.1016/j.jecp.2011.10.009>
- \*Farran, E. K., Courbois, Y., Van Herwegen, J., Cruickshank, A. G., & Blades, M. (2012b). Colour as an environmental cue when learning a route in a virtual environment: Typical and atypical development. *Research in Developmental Disabilities*, 33, 900–908. <https://doi.org/10.1016/j.ridd.2011.11.017>
- \*Farran, E. K., Formby, S., Daniyal, F., Holmes, T., & Van Herwegen, J. (2016). Route-learning strategies in typical and atypical development; eye tracking reveals atypical landmark selection in Williams syndrome. *Journal of Intellectual Disability Research*, 60, 933–944. <https://doi.org/10.1111/jir.12331>
- \*Farran, E. K., Purser, H. R., Courbois, Y., Ballé, M., Sockeel, P., Mellier, D., & Blades, M. (2015). Route knowledge and configural knowledge in typical and atypical development: A comparison of sparse



- and rich environments. *Journal of Neurodevelopmental Disorders*, 7, 37. <https://doi.org/10.1186/s11689-015-9133-6>
- Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, 43, 95–103. <https://doi.org/10.1037/0003-066X.43.2.95>
- \*Ferguson, A. M., Maloney, E. A., Fugelsang, J., & Risko, E. F. (2015). On the relation between math and spatial ability: The case of math anxiety. *Learning and Individual Differences*, 39, 1–12. <https://doi.org/10.1016/j.lindif.2015.02.007>
- \*Gagnon, K. T., Cashdan, E. A., Stefanucci, J. K., & Creem-Regehr, S. H. (2016). Sex differences in exploration behavior and the relationship to harm avoidance. *Human Nature*, 27, 82–97. <https://doi.org/10.1007/s12110-015-9248-1>
- \*Gagnon, K. T., Thomas, B. J., Munion, A., Creem-Regehr, S. H., Cashdan, E. A., & Stefanucci, J. K. (2018). Not all those who wander are lost: Spatial exploration patterns and their relationship to gender and spatial memory. *Cognition*, 180, 108–117. <https://doi.org/10.1016/j.cognition.2018.06.020>
- \*Galati, A., Weisberg, S., Newcombe, N., & Avraamides, M. N. (2015). Individual differences in spatial ability influence the effect of gesturing on navigation and spatial memory. *Proceedings of Gesture and Speech in Interaction—4th edition (GESPIN 4)*, 119–124.
- \*Galati, A., Weisberg, S. M., Newcombe, N. S., & Avraamides, M. N. (2018). When gestures show us the way: Co-thought gestures selectively facilitate navigation and spatial memory. *Spatial Cognition & Computation*, 18, 1–30. <https://doi.org/10.1080/13875868.2017.1332064>
- \*Galea, L. A., & Kimura, D. (1993). Sex differences in route-learning. *Personality and Individual Differences*, 14, 53–65. [https://doi.org/10.1016/0191-8869\(93\)90174-2](https://doi.org/10.1016/0191-8869(93)90174-2)
- \*Glicksohn, J., Balmor-Braun, I., Bar-Ziv, J., & Myslobodsky, M. S. (1998). Is spatial orientation influenced by the calcification of intracranial structures?. *International Journal of Neuroscience*, 96, 73–85. <https://doi.org/10.3109/00207459808986459>
- \*Goldiez, B. F. (2004). Techniques for assessing and improving performance in navigation and wayfinding using mobile augmented reality. *Dissertation Abstracts International*, 66(02).
- Gregoire, C. (2015). Study says that men have the better sense of direction. *Huffington Post*, Retrieved from [https://www.huffingtonpost.com/entry/men-women-navigation-study\\_us\\_5665b14ce4b08e945ff01357](https://www.huffingtonpost.com/entry/men-women-navigation-study_us_5665b14ce4b08e945ff01357)
- \*Grön, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., & Riepe, M. W. (2000). Brain activation during human navigation: Gender-different neural networks as substrate of performance. *Nature Neuroscience*, 3, 404–408. <https://doi.org/10.1038/73980>
- \*Grübel, J., Thrash, T., Hölscher, C., & Schinazi, V. R. (2017). Evaluation of a conceptual framework for predicting navigation performance in virtual reality. *PLOS ONE*, 12, e0184682. <https://doi.org/10.1371/journal.pone.0184682>
- \*Gugerty, L., & Brooks, J. (2004). Reference-frame misalignment and cardinal direction judgments: Group differences and strategies. *Journal of Experimental Psychology: Applied*, 10, 75–88. <https://doi.org/10.1037/1076-898X.10.2.75>
- \*Hamburger, K., & Röser, F. (2014). The role of landmark modality and familiarity in human wayfinding. *Swiss Journal of Psychology*, 73, 205–213. <https://doi.org/10.1024/1421-0185/a000139>
- \*Hamilton, D. A., & Sutherland, R. J. (1999). Blocking in human place learning: Evidence from virtual navigation. *Psychobiology*, 27, 453–461. <https://doi.org/10.3758/BF03332140>
- \*Hardt, O., Hupbach, A., & Nadel, L. (2009). Factors moderating blocking in human place learning: The role of task instructions. *Learning & Behavior*, 37, 42–59. <https://doi.org/10.3758/LB.37.1.42>
- \*Harrison, C. R. (2000). *Gender and menstrual cycle effects in human spatial cognition* (Order No. 9971089). Available from ProQuest Dissertations & Theses Global. (304599985)
- \*Head, D., & Isom, M. (2010). Age effects on wayfinding and route learning skills. *Behavioural Brain Research*, 209, 49–58. <https://doi.org/10.1016/j.bbr.2010.01.012>
- \*Hedge, C., Weaver, R., & Schnall, S. (2017). Spatial learning and wayfinding in an immersive environment: The digital fulldome. *Cyberpsychology, Behavior, and Social Networking*, 20, 327–333. <https://doi.org/10.1089/cyber.2016.0399>
- Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research on gender differences. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Progress through meta-analysis*. Baltimore: The Johns Hopkins University Press.
- \*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. <https://doi.org/10.1016/j.intell.2005.09.005>
- \*Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence*, 30, 425–447. [https://doi.org/10.1016/S0160-2896\(02\)00116-2](https://doi.org/10.1016/S0160-2896(02)00116-2)
- Heil, M., & Jansen-Osmann, P. (2008). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones?. *Quarterly Journal of Experimental Psychology*, 61(5), 683–689.
- \*Hemmer, I., Hemmer, M., Neidhardt, E., Obermaier, G., Uphues, R., & Wrenger, K. (2015). The influence of children's prior knowledge and previous experience on their spatial orientation skills in an urban environment. *Education*, 3(13), 43, 184–196. <https://doi.org/10.1080/03004279.2013.794852>
- Herman, J. F., Kail, R. V., & Siegel, A. W. (1979). Cognitive maps of a college campus: A new look at freshman orientation. *Bulletin of the Psychonomic Society*, 13, 183–186. <https://doi.org/10.3758/BF0333350>
- \*Heth, C. D., Cornell, E. H., & Alberts, D. M. (1997). Differential use of landmarks by 8- and 12-year-old children during route reversal navigation. *Journal of Environmental Psychology*, 17, 199–213. <https://doi.org/10.1003/jevp.1997.0057>
- \*Holding, C. S., & Holding, D. H. (1989). Acquisition of route network knowledge by males and females. *The Journal of General Psychology*, 116, 29–41. <https://doi.org/10.1080/00221309.1989.9711108>
- \*Hölscher, C., Büchner, S. J., Meilinger, T., & Strube, G. (2009). Adaptivity of wayfinding strategies in a multi-building ensemble: The effects of spatial structure, task requirements, and metric information. *Journal of Environmental Psychology*, 29, 208–219. <https://doi.org/10.1016/j.jenvp.2008.05.010>
- \*Hölscher, C., Tenbrink, T., & Wiener, J. M. (2011). Would you follow your own route description? Cognitive strategies in urban route planning. *Cognition*, 121, 228–247. <https://doi.org/10.1016/j.cognition.2011.06.005>
- \*Honda, A., & Nihei, Y. (2004). Sex differences in wayfinding behavior using well-or poorly-written route descriptions. *Tohoku Psychologica Folia*, 63, 15–24.
- Huang, X., & Voyer, D. (2017). Timing and sex effects on the “Spatial Orientation Test”: A World War II map reading test. *Spatial Cognition & Computation*, 17, 251–272. <https://doi.org/10.1080/13875868.2017.1319836>
- \*Hund, A. M. (2016). Visuospatial working memory facilitates indoor wayfinding and direction giving. *Journal of Environmental Psychology*, 45, 233–238. <https://doi.org/10.1016/j.jenvp.2016.01.008>
- \*Hund, A. M., & Gill, D. M. (2014). What constitutes effective wayfinding directions: The interactive role of descriptive cues and memory demands. *Journal of Environmental Psychology*, 38, 217–224. <https://doi.org/10.1016/j.jenvp.2014.02.006>

- \*Hund, A. M., & Minarik, J. L. (2006). Getting from here to there: Spatial anxiety, wayfinding strategies, direction type, and wayfinding efficiency. *Spatial Cognition and Computation*, 6, 179–201. [https://doi.org/10.1207/s15427633scc0603\\_1](https://doi.org/10.1207/s15427633scc0603_1)
- \*Hund, A. M., & Nazarczuk, S. N. (2009). The effects of sense of direction and training experience on wayfinding efficiency. *Journal of Environmental Psychology*, 29, 151–159. <https://doi.org/10.1016/j.jenvp.2008.05.009>
- \*Hund, A. M., & Padgitt, A. J. (2010). Direction giving and following in the service of wayfinding in a complex indoor environment. *Journal of Environmental Psychology*, 30, 553–564. <https://doi.org/10.1016/j.jenvp.2010.01.002>
- \*Ishikawa, T., Fujiwara, H., Imai, O., & Okabe, A. (2008). Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of Environmental Psychology*, 28, 74–82. <https://doi.org/10.1016/j.jenvp.2007.09.002>
- \*Ishikawa, T., & Kiyomoto, M. (2008). Turn to the left or to the west: Verbal navigational directions in relative and absolute frames of reference. In T. J. Cova, H. J. Miller, K. Beard, A. U. Frank, & M. F. Goodchild (Eds.), *International Conference on Geographic Information Science* (pp. 119–132). Berlin, Germany: Springer.
- \*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. <https://doi.org/10.1016/j.cogpsych.2005.08.003>
- \*Ishikawa, T., & Takahashi, K. (2014). Relationships between methods for presenting information on navigation tools and users' wayfinding behavior. *Cartographic Perspectives*, 75, 17–28. <https://doi.org/10.14714/CP75.82>
- \*Jansen, P., Schmelter, A., & Heil, M. (2010). Spatial knowledge acquisition in younger and elderly adults. *Experimental Psychology*, 57, 54–60. <https://doi.org/10.1027/1618-3169/a000007>
- \*Jansen-Osmann, P., & Fuchs, P. (2006). Wayfinding behavior and spatial knowledge of adults and children in a virtual environment: The role of landmarks. *Experimental Psychology*, 53, 171–181. <https://doi.org/10.1027/1618-3169.53.3.171>
- \*Jansen-Osmann, P., Schmid, J., & Heil, M. (2007). Wayfinding behavior and spatial knowledge of adults and children in a virtual environment: The role of the environmental structure. *Swiss Journal of Psychology*, 66, 41–50. <https://doi.org/10.1024/1421-0185.66.1.41>
- \*Jansen-Osmann, P., & Wiedenbauer, G. (2004). Wayfinding performance in and the spatial knowledge of a color-coded building for adults and children. *Spatial Cognition and Computation*, 4, 337–358. [https://doi.org/10.1207/s15427633scc0404\\_3](https://doi.org/10.1207/s15427633scc0404_3)
- Jonasson, Z. (2005). Meta-analysis of sex differences in rodent models of learning and memory: A review of behavioral and biological data. *Neuroscience and Biobehavioral Reviews*, 28, 811–825. <https://doi.org/10.1016/j.neubiorev.2004.10.006>
- Kastens, K. A., & Liben, L. S. (2007). Eliciting self-explanations improves children's performance on a field-based map skills task. *Cognition and Instruction*, 25(1), 45–74.
- \*Kirasac, K. C., Allen, G. L., & Siegel, A. W. (1984). Expression of configurational knowledge of large-scale environments: Students' performance of cognitive tasks. *Environment and Behavior*, 16, 687–712. <https://doi.org/10.1177/0013916584166002>
- Knierim, J. J., & Hamilton, D. A. (2011). Framing spatial cognition: Neural representations of proximal and distal frames of reference and their roles in navigation. *Physiological Reviews*, 91, 1245–1279. <https://doi.org/10.1152/physrev.00021.2010>
- \*Kober, S. E., & Neuper, C. (2011). Sex differences in human EEG theta oscillations during spatial navigation in virtual reality. *International Journal of Psychophysiology*, 79, 347–355. <https://doi.org/10.1016/j.ijpsycho.2010.12.002>
- \*Kong, X. Z., Pu, Y., Wang, X., Xu, S., Hao, X., Zhen, Z., & Liu, J. (2017). Intrinsic hippocampal-caudate interaction correlates with human navigation. *BioRxiv*, 116129. <https://doi.org/10.1101/116129>
- \*Korthauer, L. E., Nowak, N. T., Frahmmand, M., & Driscoll, I. (2017). Cognitive correlates of spatial navigation: associations between executive functioning and the virtual Morris Water Task. *Behavioural Brain Research*, 317, 470–478.
- \*Kozlowski, L. T., & Bryant, K. J. (1977). Sense of direction, spatial orientation, and cognitive maps. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 590–598. <https://doi.org/10.1037/0096-1523.3.4.590>
- \*Kushigian, R. H. (1998, March). Training for indoor wayfinding: The comparative effects of landmark, route, and configuration instruction. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 59(9–A), 3319
- Lauer, J. E., Yhang, E., & Lourenco, S. F. (2019). The development of gender differences in spatial reasoning: A meta-analytic review. *Psychological Bulletin*, 45(6), 537–565. <https://doi.org/10.1037/bul000019>
- Laurance, H. E., Learmonth, A. E., Nadel, L., & Jacobs, W. J. (2003). Maturation of spatial navigation strategies: Convergent findings from computerized spatial environments and self-report. *Journal of Cognition and Development*, 4, 211–238. <https://doi.org/10.1207/S15327647JCD0402-04>
- Lawton, C. A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles*, 30, 765–779. <https://doi.org/10.1007/BF015442>
- \*Lawton, C. A. (1996). Strategies for indoor wayfinding: The role of orientation. *Journal of Environmental Psychology*, 16, 137–145. <https://doi.org/10.1006/jev.1996.0011>
- \*Lawton, C. A., Charleston, S. I., & Zieles, A. S. (1996). Individual and gender-related differences in indoor wayfinding. *Environment and Behavior*, 28, 204–219. <https://doi.org/10.1177/0013916596282003>
- Lawton, C. A., & Kallai, J. (2002). Gender differences in wayfinding strategies and anxiety about wayfinding: A cross-cultural comparison. *Sex Roles*, 47, 389–401. <https://doi.org/10.1023/A:1021668>
- Learmonth, A. E., Newcombe, N. S., Sheridan, N., & Jones, M. (2003). Why size counts: Children's spatial reorientation in large and small enclosures. *Developmental Science*, 11, 414–426. <https://doi.org/10.1111/j.1467-7687.2008.00686.x>
- \*Lehnung, M., Haaland, V. Ø., Pohl, J., & Leplow, B. (2001). Compass-versus finger-pointing tasks: The influence of different methods of assessment on age-related orientation performance in children. *Journal of Environmental Psychology*, 21, 283–289. <https://doi.org/10.1006/jev.2001.0208>
- Lenney, E. (1977). Women's self-confidence in achievement settings. *Psychological Bulletin*, 84, 1–13. <https://doi.org/10.1037/0033-2909.84.1.1>
- \*Liben, L. S., Myers, L. J., Christensen, A. E., & Bower, C. A. (2013). Environmental-scale map use in middle childhood: Links to spatial skills, strategies, and gender. *Child Development*, 84, 2047–2063. <https://doi.org/10.1111/cdev.12090>
- \*Lingwood, J., Blades, M., Farran, E. K., Courbois, Y., & Matthews, D. (2015). The development of wayfinding abilities in children: Learning routes with and without landmarks. *Journal of Environmental Psychology*, 41, 74–80. <https://doi.org/10.1016/j.jenvp.2014.11.008>
- Lipsey, M. W., & Wilson, D. B. (2001). Practical meta-analysis. Thousand Oaks: SAGE.
- Lisofsky, N., Riediger, M., Gallinat, J., Lindenberger, U., & Kühn, S. (2016). Hormonal contraceptive use is associated with neural and affective changes in healthy young women. *NeuroImage*, 134, 597–606. <https://doi.org/10.1016/j.neuroimage.2016.04.042>
- \*Liu, I., Levy, R. M., Barton, J. J., & Iaria, G. (2011). Age and gender differences in various topographical orientation strategies. *Brain*

- Research, 1410, 112–119. <https://doi.org/10.1016/j.brainres.2011.07.005>
- \*Livingstone-Lee, S. A., Zeman, P. M., Gillingham, S. T., & Skelton, R. W. (2014). Navigational strategy may be more a matter of environment and experience than gender. *Learning and Motivation*, 45, 30–43. <https://doi.org/10.1016/j.lmot.2013.09.003>
- Lohman, D. F. (1986). The effect of speed-accuracy tradeoff on sex differences in mental rotation. *Perception and Psychophysics*, 39, 427–436. <https://doi.org/10.3758/BF03207071>
- \*Lövdén, M., Herlitz, A., Schellenbach, M., Grossman-Hutter, B., Krüger, A., & Lindenberger, U. (2007). Quantitative and qualitative sex differences in spatial navigation. *Scandinavian Journal of Psychology*, 48, 353–358. <https://doi.org/10.1111/j.1467-9450.2007.00582.x>
- \*Malinowski, J. C., & Gillespie, W. T. (2001). Individual differences in performance on a large-scale, real-world wayfinding task. *Journal of Environmental Psychology*, 21, 73–82. <https://doi.org/10.1006/jevp.2000.0183>
- \*Marchette, S. A., Yerramsetti, A., Burns, T. J., & Shelton, A. L. (2011). Spatial memory in the real world: Long-term representations of everyday environments. *Memory & Cognition*, 39, 1401–1408. <https://doi.org/10.3758/s13421-011-0108-x>
- \*McGuiness, D., & Sparks, J. (1983). Cognitive style and cognitive maps: Sex differences in representations of a familiar terrain. *Journal of Mental Imagery*, 7, 91–100.
- \*Meilinger, T., Frankenstein, J., & Bühlhoff, H. H. (2013). Learning to navigate: Experience versus maps. *Cognition*, 129, 24–30. <https://doi.org/10.1016/j.cognition.2013.05.013>
- \*Meilinger, T., Frankenstein, J., Simon, N., Bühlhoff, H. H., & Bresciani, J. P. (2016). Not all memories are the same: Situational context influences spatial recall within one's city of residency. *Psychonomic Bulletin & Review*, 23, 246–252. <https://doi.org/10.3758/s13423-015-0883-7>
- Meilinger, T., Frankenstein, J., Watanabe, K., Bühlhoff, H. H., & Hölscher, C. (2015). Reference frames in learning from maps and navigation. *Psychological Research*, 79, 1000–1008. <https://doi.org/10.1007/s00426-014-0629-6>
- Meilinger, T., Hölscher, C., Büchner, S. J., & Brösamle, M. (2006). How much information do you need? Schematic maps in wayfinding and self localisation. In *International Conference on Spatial Cognition* (pp. 381–400). Berlin, Germany: Springer.
- \*Meilinger, T., Knauff, M., & Bühlhoff, H. H. (2008). Working memory in wayfinding—A dual task experiment in a virtual city. *Cognitive Science*, 32, 755–770. <https://doi.org/10.1080/03640210802067004>
- \*Meilinger, T., Riecke, B. E., & Bühlhoff, H. H. (2014). Local and global reference frames for environmental spaces. *The Quarterly Journal of Experimental Psychology*, 67, 542–569. <https://doi.org/10.1080/17470218.2013.821145>
- \*Meneghetti, C., Borella, E., Gyselinck, V., & De Beni, R. (2012). Age-differences in environment route learning: The role of input and recall-test modalities in young and older adults. *Learning and Individual Differences*, 22, 884–890. <https://doi.org/10.1016/j.lindif.2012.04.006>
- \*Merrill, E. C., Yang, Y., Roskos, B., & Steele, S. (2016). Sex differences in using spatial and verbal abilities influence route learning performance in a virtual environment: A comparison of 6- to 12-year old boys and girls. *Frontiers in Psychology*, 7, 258. <https://doi.org/10.3389/fpsyg.2016.00258>
- \*Moffat, S. D., Hampson, E., & Hatzipantelis, M. (1998). Navigation in a “virtual” maze: Sex differences and correlation with psychometric measures of spatial ability in humans. *Evolution and Human Behavior*, 19, 73–87. [https://doi.org/10.1016/S1090-5138\(97\)00104-9](https://doi.org/10.1016/S1090-5138(97)00104-9)
- \*Moffat, S. D., Kennedy, K. M., Rodrigue, K. M., & Raz, N. (2007). Extrahippocampal contributions to age differences in human spatial navigation. *Cerebral Cortex*, 17, 1274–1282. <https://doi.org/10.1093/cercor/bhl036>
- \*Montello, D. R. (1991). Spatial orientation and the angularity of urban routes: A field study. *Environment and Behavior*, 23, 47–69. <https://doi.org/10.1177/0013916591231003>
- Montello, D. R. (1993). Scale and multiple psychologies of space. In *European conference on spatial information theory* (pp. 312–321). Springer, Berlin, Heidelberg.
- Montello, D. R. (2010). You are where? The function and frustration of you-are-here (YAH) maps. *Spatial Cognition & Computation*, 10, 94–104. <https://doi.org/10.1080/13875860903585323>
- \*Montello, D. R., Hölscher, C., Büchner, S. J., & Mavridou, M. (2016). *The role of prägnanz in apprehending architectural layouts: Coordinating egocentric and allocentric spatial knowledge*. Unpublished raw data.
- \*Montello, D. R., Lovelace, K. L., Golledge, R. G., & Self, C. M. (1999). Sex-related differences and similarities in geographic and environmental spatial abilities. *Annals of the Association of American Geographers*, 89, 515–534. <https://doi.org/10.1111/0004-5608.00160>
- \*Montello, D. R., & Pick, H. L., Jr. (1993). Integrating knowledge of vertically aligned large-scale spaces. *Environment and Behavior*, 25, 457–484. <https://doi.org/10.1177/0013916593253002>
- Morris, R. G. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation*, 12, 239–260. [https://doi.org/10.1016/0023-9690\(81\)90020-5](https://doi.org/10.1016/0023-9690(81)90020-5)
- Nazareth, A., Newcombe, N. S., Shipley, T. F., Velazquez, M., & Weisberg, S. M. (in press). Beyond small-scale spatial skills: Navigation skills and geoscience education. *Cognitive Research: Principles and Implications*.
- \*Nazareth, A., Weisberg, S. M., Margulis, K., & Newcombe, N. S. (2018). Charting the development of cognitive mapping. *Journal of Experimental Child Psychology*, 170, 86–106. <https://doi.org/10.1016/j.jecp.2018.01.009>
- \*Neidhardt, E., & Popp, M. (2010). Spatial tests, familiarity with the surroundings, and spatial activity experience: How do they contribute to children's spatial orientation in macro environments? *Journal of Individual Differences*, 31, 59–63. <https://doi.org/10.1027/1614-0001/a000010>
- \*Nelligan, B. D. (2016). *Understanding navigational success in humans* (Doctoral dissertation, Johns Hopkins University). Retrieved from <http://jhirlibrary.jhu.edu/handle/1774.2/40817>
- \*New, J., Krasnow, M. M., Truxaw, D., & Gaulin, S. J. (2007). Spatial adaptations for plant foraging: Women excel and calories count. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1626), 2679–2684. Retrieved from <http://www.jstor.org/stable/25249384>
- Newcombe, N. (1985). Methods for the study of spatial representation. In R. Cohen (Ed.), *The development of spatial cognition* (pp. 277–300). Hillsdale: Erlbaum.
- Newcombe, N. S. (2018). Individual variation in human navigation. *Current Biology*, 28, R1004–R1008. <https://doi.org/10.1016/j.cub.2018.04.053>
- \*Ngo, C. T., Weisberg, S. M., Newcombe, N. S., & Olson, I. R. (2016). The relation between navigation strategy and associative memory: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 663. <https://doi.org/10.1037/xlm0000193>
- \*Nori, R., Grandicelli, S., & Giusberti, F. (2009). Individual differences in visuo-spatial working memory and real-world wayfinding. *Swiss Journal of Psychology / Schweizerische Zeitschrift Für Psychologie / Revue Suisse De Psychologie*, 68, 7–16. <https://doi.org/10.1024/1421-0185.68.1.7>
- \*Nowak, N. T., Diamond, M. P., Land, S. J., & Moffat, S. D. (2014). Contributions of sex, testosterone, and androgen receptor CAG repeat number to virtual Morris water maze performance.



- Psychoneuroendocrinology*, 41, 13–22. <https://doi.org/10.1016/j.psyneuen.2013.12.003>
- \*Nowak, N. T., & Moffat, S. D. (2011). The relationship between second to fourth digit ratio, spatial cognition, and virtual navigation. *Archives of Sexual Behavior*, 40, 575–585. <https://doi.org/10.1007/s10508-010-9668-2>
- \*Oberholzer, Y. (2017.). *Does GPS rot your brain?: An investigation of spatial knowledge acquisition during GPS guided navigation* (Unpublished doctoral dissertation). Universität Zürich, Switzerland.
- \*Ohnishi, T., Matsuda, H., Hirakata, M., & Ugawa, Y. (2006). Navigation ability dependent neural activation in the human brain: An fMRI study. *Neuroscience Research*, 55, 361–369. <https://doi.org/10.1016/j.neures.2006.04.009>
- O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon Press.
- O'Laughlin, E. M., & Brubaker, B. S. (1998). Use of landmarks in cognitive mapping: Gender differences in self report versus performance. *Personality and Individual Differences*, 24, 595–601. [https://doi.org/10.1016/S0191-8869\(97\)00237-7](https://doi.org/10.1016/S0191-8869(97)00237-7)
- Overman, W. H., Pate, B. J., Moore, K., & Peuster, A. (1996). Ontogeny of place learning in children as measured in the radial arm maze, Morris search task, and open field task. *Behavioral Neuroscience*, 110, 1205–1228. <https://doi.org/10.1037/0735-7044.110.6.1206>
- Packard, M. G., & McGaugh, J. L. (1996). Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiology of Learning and Memory*, 65, 65–72. <https://doi.org/10.1006/nlme.1996.0007>
- \*Padgett, A. J., & Hund, A. M. (2012). How good are these directions? Determining direction quality and wayfinding efficiency. *Journal of Environmental Psychology*, 32, 164–172. <https://doi.org/10.1016/j.jenvp.2012.01.007>
- \*Padilla, L. M., Creem-Regehr, S. H., Stefanucci, J. K., & Cashdan, E. A. (2017). Sex differences in virtual navigation influenced by scale and navigation experience. *Psychonomic Bulletin & Review*, 24, 582–590. <https://doi.org/10.3758/s134>
- \*Pazzaglia, F., & Taylor, H. A. (2007). Perspective, instruction, and cognitive style in spatial representation of a virtual environment. *Spatial Cognition and Computation*, 7, 349–364. <https://doi.org/10.1080/13875860701663223>
- \*Pingel, T. J. (2010). *Strategic elements of route choice* (Order No. 3398843). Available from ProQuest Dissertations & Theses Global. (276104070)
- \*Piper, B. J., Acevedo, S. F., Edwards, K. R., Curtiss, A. B., McGinnis, G. J., & Raber, J. (2011). Age, sex, and handedness differentially contribute to neurospatial function on the Memory Island and Novel-Image Novel-Location tests. *Physiology & Behavior*, 103, 513–522. <https://doi.org/10.1016/j.physbeh.2011.03.024>
- \*Postma, A., Van Oers, M., Back, F., & Plukaard, S. (2012). Losing your car in the parking lot: Spatial memory in the real world. *Applied Cognitive Psychology*, 26, 680–686. <https://doi.org/10.1002/acp.2844>
- Poulter, S., Hartley, T., & Lever, C. (2018). The Neurobiology of mammalian navigation. *Current Biology*, 28, R1023–R1042. <https://doi.org/10.1016/j.cub.2018.05.050>
- \*Purser, H. R., Farran, E. K., Courbois, Y., Lemahieu, A., Mellier, D., Sockeel, P., & Blades, M. (2012). Short-term memory, executive control, and children's route learning. *Journal of Experimental Child Psychology*, 113, 273–285. <https://doi.org/10.1016/j.jecp.2012.06.005>
- \*Purser, H. R., Farran, E. K., Courbois, Y., Lemahieu, A., Sockeel, P., Mellier, D., & Blades, M. (2015). The development of route learning in Down syndrome, Williams syndrome and typical development: investigations with virtual environments. *Developmental Science*, 18, 599–613. <https://doi.org/10.1111/desc.12236>
- Puts, D. A., McDaniel, M. A., Jordan, C. L., & Breedlove, S. M. (2008). Spatial ability and prenatal androgens: Meta-analyses of congenital adrenal hyperplasia and digit ratio (2D: 4D) studies. *Archives of Sexual Behavior*, 37, 100–111. <https://doi.org/10.1007/s10508-007-9271-3>
- \*Rahman, Q., Sharp, J., McVeigh, M., & Ho, M. L. (2017). Sexual orientation-related differences in virtual spatial navigation and spatial search strategies. *Archives of Sexual Behavior*, 46, 1279–1294. <https://doi.org/10.1007/S10508-017-0986-5>
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd). Newbury Park: SAGE.
- \*Richardson, A. E., Powers, M. E., & Bousquet, L. G. (2011). Video game experience predicts virtual, but not real navigation performance. *Computers in Human Behavior*, 27, 552–560. <https://doi.org/10.1016/j.chb.2010.10.003>
- \*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. <https://doi.org/10.3758/BF03211566>
- \*Rizk-Jackson, A., Acevedo, S. F., Inman, D., Howieson, D., Benice, T. S., & Raber, J. (2006). Effects of sex on object recognition and spatial navigation in humans. *Behavioural Brain Research*, 173, 181–190. <https://doi.org/10.1016/j.bbr.2006.06.029>
- \*Rodgers, M. K., Sindone, J. A., III, & Moffat, S. D. (2012). Effects of age on navigation strategy. *Neurobiology of Aging*, 33, 202–e15. <https://doi.org/10.1016/j.neurobiolaging.2010.07.021>
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev.). Beverly Hills: SAGE.
- Rosenthal, R. (1979). The “file drawer problem” and the tolerance for null results. *Psychological Bulletin*, 86, 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- \*Ruddle, R. A., Volkova, E., Mohler, B., & Bühlhoff, H. H. (2011). The effect of landmark and body-based sensory information on route knowledge. *Memory & Cognition*, 39, 686–699. <https://doi.org/10.3758/s13421-010-0054-z>
- \*Sadalla, E. K., & Montello, D. R. (1989). Remembering changes in direction. *Environment and Behavior*, 21, 346–363. <https://doi.org/10.1177/0013916589213006>
- \*Sandberg, E. H., & Huttenlocher, J. (2001). Advanced spatial skills and advance planning: Components of 6-year-olds' navigational map use. *Journal of Cognition and Development*, 2, 51–70. [https://doi.org/10.1207/S15327647JCD0201\\_3](https://doi.org/10.1207/S15327647JCD0201_3)
- \*Sandstrom, N. J., Kaufman, J., & Huettel, S. A. (1998). Males and females use different distal cues in a virtual environment navigation task. *Cognitive Brain Research*, 6, 351–360. [https://doi.org/10.1016/S0926-6410\(98\)00002-0](https://doi.org/10.1016/S0926-6410(98)00002-0)
- \*Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., & Lin, N. (2019). Event memory uniquely predicts memory for large-scale space. *Memory & Cognition*, 1–17. <https://doi.org/10.3758/s13421-018-0860-2>
- \*Saucier, D. M., Green, S. M., Leason, J., MacFadden, A., Bell, S., & Elias, L. J. (2002). Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies? *Behavioral Neuroscience*, 116, 403–410. <https://doi.org/10.1037/0735-7044.116.3.403>
- \*Schinazi, V. R., Nardi, D., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2013). Hippocampal size predicts rapid learning of a cognitive map in humans. *Hippocampus*, 23, 515–528. <https://doi.org/10.1002/hipo.22111>
- \*Schmitz, S. (1997). Gender-related strategies in environmental development: Effects of anxiety on wayfinding in and representation of a three-dimensional maze. *Journal of Environmental Psychology*, 17, 215–228. <https://doi.org/10.1006/jevp.1997.0056>
- \*Schmitz, S. (1999). Gender differences in acquisition of environmental knowledge related to wayfinding behavior, spatial anxiety and self-



- estimated environmental competencies. *Sex Roles*, 41, 71–93. <https://doi.org/10.1023/A:1018837808724>
- \*Schmitzer-Torbert, N. (2007). Place and response learning in human virtual navigation: Behavioral measures and gender differences. *Behavioral Neuroscience*, 121, 277–290. <https://doi.org/10.1037/0735-7044.121.2.277>
- \*Schoenfeld, R., Lehmann, W., & Lepow, B. (2010). Effects of age and sex in mental rotation and spatial learning from virtual environments. *Journal of Individual Differences*, 31, 78–82. <https://doi.org/10.1027/1614-0001/a000014>
- \*Sensibaugh, T. K. (2017). *BIS and BAS sensitivity and spatial navigation ability* (Unpublished master's thesis). University of Wyoming, Laramie, WY.
- \*Shore, D. I., Stanford, L., MacInnes, W. J., Brown, R. E., & Klein, R. M. (2001). Of mice and men: Virtual Hebb—Williams mazes permit comparison of spatial learning across species. *Cognitive, Affective, & Behavioral Neuroscience*, 1, 83–89. <https://doi.org/10.3758/CABN.1.1.83>
- \*Silverman, I., Choi, J., Mackewn, A., Fisher, M., Moro, J., & Olshansky, E. (2000). Evolved mechanisms underlying wayfinding: Further studies on the hunter-gatherer theory of spatial sex differences. *Evolution and Human Behavior*, 21, 201–213. [https://doi.org/10.1016/S1090-5138\(00\)00036-2](https://doi.org/10.1016/S1090-5138(00)00036-2)
- \*Sjölander, M., Höök, K., Nilsson, L., & Andersson, G. (2005). Age differences and the acquisition of spatial knowledge in a three-dimensional environment: Evaluating the use of an overview map as a navigation aid. *International Journal of Human-Computer Studies*, 63, 537–564. <https://doi.org/10.1016/j.ijhcs.2005.04.024>
- Skaggs, W. E., & McNaughton, B. L. (1996). Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271, 1870–1873. <https://doi.org/10.1126/science.271.5257.1870>
- Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—Or vice versa. *Journal of the American Statistical Association*, 54, 30–34. <https://doi.org/10.2307/2282137>
- \*Sugimoto, M., & Kusumi, T. (2014). The effect of text continuity on spatial representation: Route versus survey perspective. *Cognitive Processing*, 15, 65–75. <https://doi.org/10.1007/s10339-013-0582-0>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th). Boston: Allyn & Bacon.
- \*Tan, D. S., Czerwinski, M. P., & Robertson, G. G. (2006). Large displays enhance optical flow cues and narrow the gender gap in 3-D virtual navigation. *Human Factors*, 48, 318–333. <https://doi.org/10.1518/001872006777724381>
- \*Tang, C., Wu, W., & Lin, C. (2009). Using virtual reality to determine how emergency signs facilitate way-finding. *Applied Ergonomics*, 40, 722–730. <https://doi.org/10.1016/j.apergo.2008.06.009>
- \*Tippett, W. J., Lee, J., Mraz, R., Zakzanis, K. K., Snyder, P. J., Black, S. E., & Graham, S. J. (2009). Convergent validity and sex differences in healthy elderly adults for performance on 3D virtual reality navigation learning and 2D hidden maze tasks. *CyberPsychology & Behavior*, 12, 169–174. <https://doi.org/10.1089/cpb.2008.0218>
- \*Tlauka, M., Brolese, A., Pomeroy, D., & Hobbs, W. (2005). Gender differences in spatial knowledge acquired through simulated exploration of a virtual shopping centre. *Journal of Environmental Psychology*, 25, 111–118. <https://doi.org/10.1016/j.jenvp.2004.12.002>
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A., Warren, C., & Newcombe, N. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139, 352–402. <https://doi.org/10.1037/a0028446>
- \*van Gerven, Dustin J. H., Schneider, A. N., Wuitchik, D. M., & Skelton, R. W. (2012). Direct measurement of spontaneous strategy selection in a virtual Morris water maze shows females choose an allocentric strategy at least as often as males do. *Behavioral Neuroscience*, 126, 465–478. <https://doi.org/10.1037/a0027992>
- \*Vashro, L., & Cashdan, E. (2015). Spatial cognition, mobility, and reproductive success in northwestern Namibia. *Evolution and Human Behavior*, 36, 123–129. <https://doi.org/10.1016/j.evolhumbehav.2014.09.009>
- \*Vashro, L., Padilla, L., & Cashdan, E. (2016). Sex differences in mobility and spatial cognition. *Human Nature*, 27, 16–34. <https://doi.org/10.1007/s12110-015-9247-2>
- \*Ventura, M., Shute, V., Wright, T., & Zhao, W. (2013). An investigation of the validity of the virtual spatial navigation assessment. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00852>
- \*Verdine, B. N. (2011). *Navigation experience in video game environments: Effects on spatial ability and map use skills* (Order No. 3479852). Available from ProQuest Dissertations & Theses Global. (898587381)
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48. Retrieved from <http://www.jstatsoft.org/v36/i03/>
- \*von Stülpnagel, R., & Steffens, M. C. (2013). Active route learning in virtual environments: Disentangling movement control from intention, instruction specificity, and navigation control. *Psychological Research*, 77, 555–574. <https://doi.org/10.1007/s00426-012-0451-y>
- Vorhees, C. V., & Williams, M. T. (2006). Morris water maze: Procedures for assessing spatial and related forms of learning and memory. *Nature Protocols*, 1, 848–858. <https://doi.org/10.1038/nprot.2006.116>
- Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin and Review*, 18, 267–277. <https://doi.org/10.3758/s13423-010-0042-0>
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140, 1174–1204. <https://doi.org/10.1037/a0036620>
- Voyer, D., Voyer, S. D. & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological bulletin*, 117(2), 250.
- Voyer, D., Voyer, S. D., & Saint-Aubin, J. (2017). Sex differences in spatial working memory: A meta-analysis. *Psychonomic Bulletin and Review*, 24, 307–334. <https://doi.org/10.3758/s13423-016-1085-7>
- \*Waller, D., & Greenauer, N. (2007). The role of body-based sensory information in the acquisition of enduring spatial representations. *Psychological Research*, 71, 322–332. <https://doi.org/10.1007/s00426-006-0087-x>
- \*Waller, D., Knapp, D., & Hunt, E. (2001). Spatial representations of virtual mazes: The role of visual fidelity and individual differences. *Human Factors*, 43, 147–158. <https://doi.org/10.1518/001872001775992561>
- Wang, L., & Carr, M. (2014). Working memory and strategy use contribute to gender differences in spatial ability. *Educational Psychologist*, 49, 261–282. <https://doi.org/10.1080/00461520.2014.960568>
- \*Wang, L., Mou, W., & Sun, X. (2014). Development of landmark knowledge at decision points. *Spatial Cognition & Computation*, 14, 1–17. <https://doi.org/10.1080/13875868.2013.784768>
- \*Ward, S. L., Newcombe, N., & Overton, W. F. (1986). Turn left at the church, or three miles north: A study of direction giving and sex differences. *Environment and Behavior*, 18, 192–213. <https://doi.org/10.1177/0013916586182003>
- \*Watson, T. L. (1999). Can I get there from here (and can I get back)? A study of abilities and wayfinding performance (aging, gender differences, spatial cognition, older adults). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 60(5–B), 2395.
- \*Weisberg, S. M., & Newcombe, N. S. (2014). A slippery directional slope: Individual differences in using slope as a directional cue.

- Memory & Cognition*, 42, 648–661. <https://doi.org/10.3758/s13421-013-0387-5>
- \*Weisberg, S. M., Badgio, D., & Chatterjee, A. (2018). Feel the way with a vibrotactile compass: Does a navigational aid navigation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 667–679. <https://doi.org/10.1037/xlm0000472>
- \*Weisberg, S. M., & Newcombe, N.S. (2016) How do (some) people make a cognitive map? Routes, places, and working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(5), 768–785. <https://doi.org/10.1037/xlm0000200>
- \*Wenczel, F., Hepperle, L., & von Stülpnagel, R. (2017). Gaze behavior during incidental and intentional navigation in an outdoor environment. *Spatial Cognition & Computation*, 17, 121–142. <https://doi.org/10.1080/13875868.2016.1226838>
- Wiener, J. M., Büchner, S. J., & Hölscher, C. (2009). Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Computation*, 9, 152–165. <https://doi.org/10.1080/13875860902906496>
- \*Wiener, J. M., de Condappa, O., Harris, M. A., & Wolbers, T. (2013). Maladaptive bias for extrahippocampal navigation strategies in aging humans. *Journal of Neuroscience*, 33, 6012–6017. <https://doi.org/10.1523/JNEUROSCI.0717-12.2013>
- \*Wiener, J. M., Kmecova, H., & de Condappa, O. (2012). Route repetition and route retracing: Effects of cognitive aging. *Frontiers in Aging Neuroscience*, 4, 7. <https://doi.org/10.3389/fnagi.2012.00007>
- \*Wiener, J. M., & Mallot, H. A. (2003). ‘Fine-to-coarse’ route planning and navigation in regionalized environments. *Spatial Cognition and Computation*, 3, 331–358. [https://doi.org/10.1207/s15427633scc0304\\_5](https://doi.org/10.1207/s15427633scc0304_5)
- \*Wiener, J. M., Schnee, A., & Mallot, H. A. (2004). Use and interaction of navigation strategies in regionalized environments. *Journal of Environmental Psychology*, 24, 475–493. <https://doi.org/10.1016/j.jenvp.2004.09.006>
- Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences*, 14, 138–146. <https://doi.org/10.1016/j.tics.2010.01.001>
- Wood, B. M., Harris, J. A., Vashro, L., Sayre, K. M., Raichlen, D. A., Pontzer, H., ... Cashdan, E. (2019). Hadza hunter-gatherers exhibit gender differences in space use and spatial cognition consistent with the ecology of male and female targeted foods. *American Journal of Physical Anthropology*, 168(S68), 273–274.
- \*Yasen, A. L., Raber, J., Miller, J. K., & Piper, B. J. (2015). Sex, but not apolipoprotein E polymorphism, differences in spatial performance in young adults. *Archives of Sexual Behavior*, 44, 2219–2226. <https://doi.org/10.1007/s10508-015-0497-1>
- \*Yuan, P., Daugherty, A. M., & Raz, N. (2014). Turning bias in virtual spatial navigation: Age-related differences and neuroanatomical correlates. *Biological Psychology*, 96, 8–19. <https://doi.org/10.1016/j.biopsycho.2013.10.009>
- \*Zancada-Menendez, C., Sampedro-Piquero, P., Meneghetti, C., Labate, E., Begega, A., & Lopez, L. (2015). Age differences in path learning: The role of interference in updating spatial information. *Learning and Individual Differences*, 38, 83–89. <https://doi.org/10.1016/j.lindif.2015.01.015>
- \*Zhong, J. Y. (2011). *Individual differences in navigational strategies and mental representations of large-scale environments*. Honors Thesis archived in the Central Library of the National University of Singapore, Singapore (Call no.: BF21-2011-49).
- \*Zwergal, A., Schöberl, F., Xiong, G., Pradhan, C., Covic, A., Werner, P., ... Dieterich, M. (2016). Anisotropy of human horizontal and vertical navigation in real space: Behavioral and PET correlates. *Cerebral Cortex*, 26, 4392–4404. <https://doi.org/10.1093/cercor/bhv213>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.