ORIGINAL ARTICLE

Integration of visual landmark cues in spatial memory

Phillip M. Newman[1] · Timothy P. McNamara[1]

Abstract

Over the past two decades, much research has been conducted to investigate whether humans are optimal when integrating sensory cues during spatial memory and navigational tasks. Although this work has consistently demonstrated optimal integration of visual cues (e.g., landmarks) with body-based cues (e.g., path integration) during human navigation, little work has investigated how cues of the same sensory type are integrated in spatial memory. A few recent studies have reported mixed results, with some showing very little benefit to having access to more than one landmark, and others showing that multiple landmarks can be optimally integrated in spatial memory. In the current study, we employed a combination of immersive and non-immersive virtual reality spatial memory tasks to test adult humans' ability to integrate multiple landmark cues across six experiments. Our results showed that optimal integration of multiple landmark cues depends on the difficulty of the task, and that the presence of multiple landmarks can elicit an additional latent cue when estimating locations from a ground-level perspective, but not an aerial perspective.

Introduction

Successful navigation is a critical function of any mobile organism as faulty navigation can lead to injury or even death. Thus, organisms require mechanisms by which they can remain oriented in an environment. One such mechanism is to utilize spatial cues (e.g., landmarks) that inform the organism of its location with respect to an internal or external reference frame. For example, a shopper might attempt to locate his or her car by recalling that it was parked near an oak tree. The oak tree serves as an environmental cue that provides relative information about the location of the car. The shopper may also recall that he or she walked diagonally to the left from the car to the entrance of the store. The person's internal sense of direction serves as a body-based cue. One strategy for successful navigation is to combine the information from different cues to obtain a more precise estimate of the car's true location. However, if the cues provide conflicting estimates, it may be better to choose one cue over the other.

Many studies have examined the ways by which humans privilege and integrate spatial cues during navigation (Bates & Wolbers, 2014; Butler et al., 2010; Chen et al., 2017; Cheng et al., 2007; Frissen et al., 2011; Kalia et al., 2013; McNamara & Chen, 2020; Nardini et al., 2008; Newman & McNamara, 2021; Petrini et al., 2016; Philbeck & O'Leary, 2005; Ratliff & Newcombe, 2008; Sjolund et al., 2018; Tcheang, Bulthoff, & Burgess, 2011; Twyman, Holden, & Newcombe, 2018; Wang & Mou, 2020; Wang, Mou, & Dixon, 2018; Xu, Regier, & Newcombe, 2017; Zhao & Warren, 2015a, b). Cheng et al. (2007) proposed that navigators weight and integrate spatial cues according to models of Maximum Likelihood Estimation (MLE). According to the MLE model, each cue provides a probability distribution for a target location, with less variable distributions representing more reliable cues. Weights are assigned to cues based on their relative reliabilities (i.e., more reliable cues receive more weight) and are inversely proportional to the response variance associated with a given cue. Single-cue estimates are linearly combined to obtain a statistically optimal (in the sense of minimizing variance) estimate of the target's location. The distribution of such optimal estimates is known as the optimal or combined distribution (in a Bayesian analysis, this distribution is referred to as the posterior distribution). Thus, MLE predicts that navigators optimally weight and integrate spatial cues during navigation according to cue reliability.

✉ Phillip M. Newman
   phillip.m.newman@vanderbilt.edu

[1] Department of Psychology, Vanderbilt University, 301 Wilson Hall, 111 21st Avenue South, Nashville, TN 37212, USA

In a typical cue integration experiment, participants attempt a spatio-perceptual task, with the number of available cues being manipulated (usually within subjects; Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002; Friedmann, Ludvig, & Legge, 2013; Girshick & Banks, 2009; Hillis et al., 2004; Jacobs, 1999; Oruç et al., 2003, Rohde et al., 2016). On some trials, both cues are available and are consistent (both-cue condition). On other trials, both cues are available but in conflict with one another, each indicating different estimates of a target (conflict condition). Critically, there are also trials for each of the single cues (single-cue conditions). Single-cue trials provide response distributions for each of the cues, which are used to compute cue reliabilities and predicted weights. The reliability of a given cue is equal to the inverse of its variance:

$$r = \frac{1}{\sigma^2}. \tag{1}$$

The optimal weights (W) for cues (A and B) are,

$$W_A = \frac{r_A}{(r_A + r_B)}, \tag{2}$$

$$W_B = \frac{r_B}{(r_A + r_B)}. \tag{3}$$

Note that W_A and W_B sum to 1. The optimal combination of the two cues is,

$$\mu_O = W_A \mu_A + W_B \mu_B. \tag{4}$$

The variance of the combined distribution is,

$$\sigma_O^2 = \frac{\sigma_A^2 \sigma_B^2}{\sigma_A^2 + \sigma_B^2}. \tag{5}$$

Note that the predicted, optimal variance is always less than or equal to the variances of the two single cues (i.e., more cues available allows greater precision). If navigators are optimally combining the cues, response variance for both-cue trials will equal the optimal estimate.

In a recent study, Sjolund et al. (2018; Experiment 1) showed that human navigators optimally integrated environmental (room geometry) and body-based cues during a homing task. The homing task required participants to follow a two-legged path marked by waypoints before attempting to return directly to the path origin using memory. The trials varied in the number of cues available to the participant (i.e., environmental or body based, or both), and whether cues were in conflict when both were presented. When both cues were presented and non-conflicting, response variability was reduced relative to the two single-cue conditions and was consistent with optimal integration. Furthermore, observed cue weights from the conflict condition were consistent with

predicted weights based on cue-relative reliability. Other studies using similar methods have shown that navigators can optimally combine body-based information with other visual cues, such as landmarks (e.g., Bates & Wolbers, 2014; Butler et al., 2010; Chen et al., 2017; Kalia et al., 2013; Nardini et al., 2008; Petrini et al., 2016; Tcheang et al., 2011; Zhao & Warren, 2015b) and optic flow (e.g., Fetsch, DeAngelis, & Angelaki, 2010; Fetsch et al., 2009, 2012).

However, cue combination studies in navigation have primarily focused on integration of cues between sensory modalities (i.e., visual and body-based cues). Some studies of cue competition suggest that visual and body-based cues are independent and do not compete for computational resources (Mou & Spetch, 2013; Shettleworth & Sutton, 2005). On the other hand, many cue competition studies examining the interplay of visual cues alone have demonstrated interference (e.g., blocking and overshadowing) between cues (Biegler & Morris, 1999; Chamizo, 2003; Chamizo et al., 1985; Hamilton & Sutherland, 1999; Hardt, Huprach, & Nadel, 2009; Jacobs et al., 1997, 1998; Prados, 2011; Rodrigo et al., 2005; Sánchez-Moreno et al., 1999).

Previous investigations into the use of separate spatial representations from two landmarks when recalling the location of a target have demonstrated a lack of cue integration (Baguley et al., 2006; Clark et al., 2013). For instance, Baguley et al. (2006) had participants learn the location of a target along a horizontal line with two individually presented landmarks. Participants in their study did not show improved performance when recalling the location of the target when both landmarks were presented at test compared to when only one was presented, suggesting that they were unable to integrate the information provided by both landmarks (Experiments 1 and 2). This pattern maintained even when participants learned the location of the target in the presence of both landmarks (Experiment 3). However, recent findings by Du et al. (2017) using a similar task found that participants optimally combined two landmark cues when estimating the location of a target on both horizontal and vertical axes.

The discrepancy in findings between Baguley et al. (2006) and Du et al. (2017) might be attributable to some key methodological differences. For example, Baguley et al. (2006) did not vary the absolute location of the landmarks and horizontal line on the computer screen, which may have allowed for participants to encode the target relative to the edges of the screen, while Du et al. (2017) varied the absolute location of the landmarks and horizontal (or vertical) line while keeping their relative distances constant. Baguley et al. (2006) also had participants learn many stimulus–target pairs during learning, requiring participants to encode more information than might have been possible. Du et al. (2017) instead trained participants to learn a single stimulus–target pair. Although Du et al. (2017) demonstrated optimal

combination of two landmarks, this result was only observed when participants learned the location of the target with both cues presented simultaneously.

Other evidence suggests that the use of multiple visual cues can lead to supra-optimal performance with those cues that is better than the sum of performance with each cue alone. Mou and Spetch (2013; Experiment 5) examined how humans combined visual cues during a spatial memory task. During a learning phase, participants studied an array of five objects arranged as a pentagon from an aerial perspective. The test involved a two-alternative forced-choice task in which participants responded to whether a target object had moved relative to the initial learning array. Humans can encode object locations as distance vectors between the target object and other objects (inter-object vectors), as well as between the target object and the viewer's body (body-object vectors; Klatzky, 1998; McNamara, 1986; Mou & McNamara, 2002; Mou & Spetch, 2013; Stevens & Coupe, 1978; Xiao et al., 2009). On some trials, participants had access to the entire array of objects during the test (both-cue trials). On other trials, participants either had access to the two closest objects or the two farthest objects to the target (close- and far-cue trials). These trials are analogous to single-cue trials such that their response distributions are combined to predict optimal cue integration. Mou and Spetch found that performance for the both-cue trials was better than optimal integration based on the close- and far-cue trials. They argued that the observed supra-optimal performance stemmed from an additional configural cue when all objects were present during the test. That is, when all objects were present during the test, participants had access to the inter-object vectors between the close and far objects and the target object, but they could also judge the location of the target object relative to the overall configuration. This supra-optimal effect was not observed for any of the other experiments investigating the integration of inter-object and body-object vectors, suggesting that these representations are likely governed by separate systems, and are akin to environmental and body-based cues (Burgess, 2008).

Previous work has shown that human navigators can use configural information during spatial search tasks (Jacobs et al., 1998; Spetch, Cheng, & MacDonald, 1996; Spetch et al. 1997). Spetch et al. (1997) had participants search for a target object in a grassy field. A 6×6 m square area was defined by four identical posts which served as landmarks. During training, the goal was always present and located in the center of the array of landmarks. During testing, participants were told that the goal would be present on some trials but not others, and that if they could not find the goal in a reasonable amount of time, they were to place a marker where the goal should be. Participants completed three test trials. One trial served as a control in which the landmarks were still arranged as vertices of a 6×6 m square

area. Another trial was a left–right expansion test in which landmarks were placed 12 m apart in the left–right dimension only, maintaining a distance of 6 m apart in the up-down dimension. Lastly, one trial was a diagonal expansion test in which landmarks were placed 12 m apart along both dimensions. On all three tests, participants searched in the center of the landmark arrays as opposed to using distance vectors from any of the individual landmarks. Thus, humans appeared to use configural information of landmarks as a spatial cue during navigation.

However, it remains unclear if humans combine configural information with individual landmark vectors according to the MLE framework during navigation. That is, do navigators show supra-optimal performance when the entire landmark array is present during navigation relative to the optimal combination of subsets of the array? Or, will navigators choose to only use the most reliable subset of cues, otherwise unable or refusing to integrate subsets? Experiments 1 and 2 were designed to address three hypotheses regarding this question. The *optimal integration hypothesis* predicts that navigators represent target locations relative to individual landmark vectors and combine these representations during retrieval. Previous work (e.g., Spetch et al., 1996, 1997) investigating the use of configural information has used arrays of identical landmarks, making individual landmark vectors unreliable. If the array is made up of unique landmarks, navigators may disregard configural information. Thus, the optimal combination hypothesis predicts optimal combination of subsets of the array. The *supra-optimal hypothesis* predicts that navigators combine individual landmark vectors with configural information. This hypothesis is consistent with work by Mou and Spetch (2013) showing that humans combine inter-object and configural cues during a two-alternative forced-choice spatial perception task. Importantly, the supra-optimal hypothesis posits that the configural information (the latent cue) is integrated with the landmarks in the manner specified by the MLE model. However, it is also possible that this latent cue might dominate, leading to supra-optimal performance by way of greater reliability (see General Discussion). The supra-optimal hypothesis predicts that response variability is reduced beyond the optimal combination of the subsets of the array. The *hierarchical hypothesis* predicts that navigators will choose to use the most reliable subset of cues during retrieval and predicts that response variability during both-cue trials will be equal to the response variability of the most reliable cue (Du et al., 2017).

Experiment 1

Participants completed a spatial memory task in immersive virtual reality. Participants first learned a target location by walking to a post in the presence of four unique landmarks arranged as vertices of a square. Participants then attempted to walk back to the location of the post from a different starting position. On some trials, the entire array of landmarks was present during the test. On other trials, only a subset of the landmarks was present. Response accuracy and response variability were assessed for each trial type, and optimal precision was predicted from response variability from the subset trials. If participants integrate configural information, response variability should be lower than predicted by optimal integration, consistent with the supra-optimal hypothesis. If participants represent the target location with respect to individual landmarks, response variability should be consistent with optimal integration, as predicted by the optimal integration hypothesis.

Methods

Participants

Undergraduate students ($N=25$; age $M=19.36$, $SD=1.04$; 13 females) from Vanderbilt University participated in exchange for credit in a psychology course. Previous cue combination studies in navigation (e.g., Bates & Wolbers, 2014; Chen et al., 2017; Sjolund et al., 2018) have used similar sample sizes, finding medium effect sizes (η_G^2s $= 0.11$–0.18) of cue condition on response variability. A G*Power analysis for repeated-measures ANOVA ($\alpha=0.05$, power$=0.95$, groups$=1$, measurements$=4$; Faul et al., 2009) showed that a sample size of 26 is sufficient to achieve $f=0.30$ (medium effect$=0.25$, large$=0.40$). Data for eight additional participants were excluded due to simulator sickness ($n=1$), failure to correctly follow experimental procedures ($n=3$), recognizing which landmarks belonged to a subset ($n=1$), response variability in at least one condition above the third quartile by three times the interquartile range ($n=1$), or equipment malfunction ($n=2$). A trial was considered an outlier if the response error fell above three times the interquartile range above the third quartile for a given cue condition. Less than 0.01% of trials were cut using this criterion.

Materials and procedure

The immersive virtual environment was rendered in Unity, a multiplatform game engine (https://unity.com/). The environment was displayed in the HTC Vive head-mounted
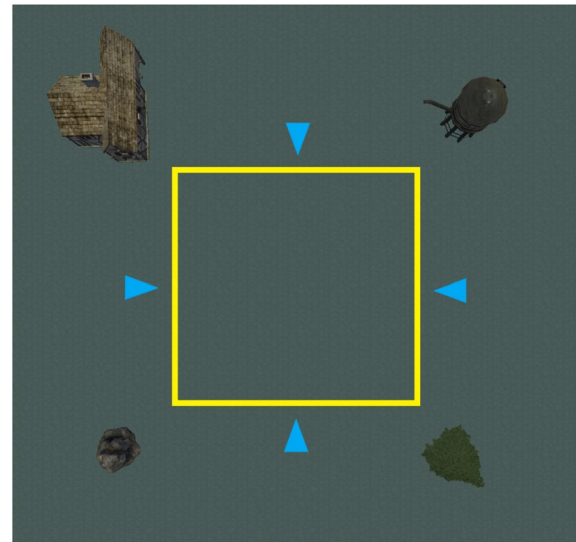

Fig. 1 Aerial perspective of the virtual environment in Experiment 1. The yellow box encompasses all possible locations for the target post. Blue triangles represent possible starting locations. The yellow box and blue triangles were not visible to participants (color figure online)

display (HMD) with a resolution of 1080×1200 per eye, refreshed at 90 Hz. The field-of-view of the HMD is approximately 110 degrees diagonally. Participants used HTC Vive's wireless controller to progress throughout the experiment. Position and orientation tracking were supported by HTC Vive's Lighthouse tracking system, with a 4×4 m tracking space. The size of the room was 7.3×8.5 m. The TPCast (https://www.tpcastvr.com/) supported wireless tracking of the HMD. With this approach, participants were able to physically rotate and walk throughout the virtual environment. The experiment was implemented on a computer with an Intel Core i7-6700K processor, 32 GB of RAM, and a NVIDIA GTX 1080 graphics card.

Numerous studies have demonstrated that experience and training with video games can enhance spatial abilities (see Uttal et al., 2013). Attempting to control for prior experience with video games, we administered a video game history and habits questionnaire to participants (originally developed by Boot et al., 2008). The survey asked participants about demographics, weekly time spent playing video games, when they first started playing video games, and what video game consoles they own. Only six participants reported playing video games at least 5 h a week, and only five participants reported being an active gamer. Therefore, we do not consider this metric any further. The survey also asked participants to describe any strategies used to complete the experimental task, and whether they noticed any patterns in the landmarks that were present during the test phase.
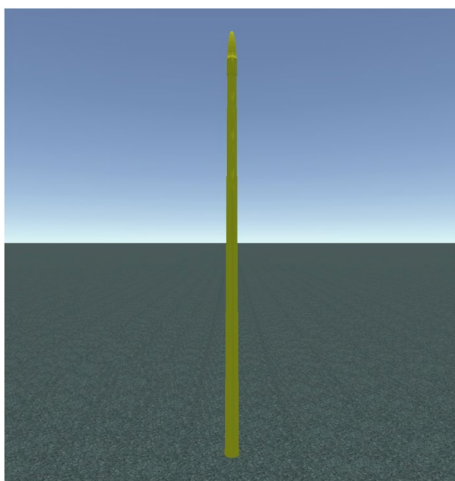
Fig. 2 Image of the target post

The virtual environment consisted of an infinite ground plane and four landmarks: A tree, rock, tower, and house. Landmarks were arranged as vertices of a square (Fig. 1), with adjacent landmarks 12 m apart. Yellow target posts (Fig. 2) 0.05 m in diameter appeared randomly within a 3.6 × 3.6 m area centrally superimposed between the landmarks (see Fig. 1).

Every trial comprised a learning phase, test phase, and resetting procedure before each phase, which kept participants within the VR tracking space. During the resetting procedure, a blue post and a red post were the only visible objects in the environment. Participants were instructed to walk to the blue post and turn to face the red post, and then press a button on the controller to begin the next phase. Participants began the learning phase at a randomly chosen starting location, each of which was half-way between and aligned with the two closest landmarks (see Fig. 1). During the learning phase, all landmarks were visible, and a yellow post marked the target location.[1] Participants were instructed to walk to the yellow post and take time to learn its location by looking around at the surrounding landmarks. Participants were told that some landmarks might or might not disappear during the test phase, so it was important to learn the location of the post relative to all the landmarks. Also, because of particular interest in the use of landmark cues, but not body-based cues, participants were told that they would never start at the same location during the test phase

as they did during the learning phase. When participants thought they had memorized the location of the yellow post, they pressed a button on the controller to complete another resetting procedure before starting the test phase.

Participants started the test phase at one of the remaining three starting locations (i.e., if participants started at the southern starting location during the learning phase, they could only start at the east, west, or north locations during the test phase), which was chosen randomly. During the test phase, the yellow post was no longer visible, and participants were instructed to walk to the remembered location of the yellow post. In the both-cue condition, all landmarks remained visible. In the subset-A condition, the tree and tower were no longer visible, leaving only the house and rock available. In the subset-B condition, the house and rock were no longer visible, leaving only the tree and tower available. Once participants were confident that they were standing at the location of the yellow post, they pressed a button on the controller to confirm their response and move on to the next trial. Participants completed a practice block with one of each trial type presented in a random order, followed by ten test blocks of three trials each, with one trial for each cue condition.[2]

Analyses

Because the target could take on random locations, the target location for each trial was treated as the origin and responses were aligned accordingly. We first analyzed response accuracy, defined as the mean Euclidean distance between each response location and the target location (origin). Following previous work (e.g., Chen et al., 2017; Nardini et al., 2008; Sjolund et al., 2018), the standard deviation was calculated for each condition, using the absolute distance of each response relative to the mean response location (see Appendix).[3] Using Eq. (5), optimal integration was calculated by

---

[1] Mou and Spetch (2013) showed that landmark cues can interfere with each other when presented simultaneously. In other words, landmarks contribute more individually when presented alone versus together, which would serve to underpredict optimal response variability during both-cue trials. Thus, all landmarks were displayed during the learning phase across all trial types to maintain consistency of potential interference.

[2] Variance reduction in the both-cue condition is not the only means to assess cue integration. Other studies (e.g., Du et al., 2017; Spetch et al., 1997) have employed expansion (conflict) trials in which the configuration of the landmarks is expanded along one or both dimensions. Expansion trials allow the experimenter to observe the weights that participants give to either cue, which can then be compared to the optimal weights predicted from relative cue reliability. Here, we did not employ such trials as the virtual space was constrained by the physical walls of the room. Bias in responding during expansion trials potentiate hazards of bumping into the walls of the room, and warning people before collision would confound the results by causing them to stop preemptively. We did not introduce expansion trials during the non-immersive Experiments 3–6 to maintain consistency in the basic design throughout.

[3] It is not possible to measure variable bias with only one response per target location, so constant bias must be assumed (if there is bias). The results are identical whether the standard deviation is computed relative to the aligned target locations or the mean response location.

combining the variances from the two subset conditions. We did not correct for multiple comparisons when conducting tests comparing model predictions to combined-cue performance as higher cost is assigned to falsely accepting the model (cf. Chen et al., 2017). Mauchly's test revealed that the assumption of sphericity was met for all repeated-measures ANOVAs reported. However, the Greenhouse–Geisser correction for departure from sphericity was still used as even non-significant departures from sphericity can influence within-subject effects. GG epsilon is reported for all repeated-measures ANOVAs and Cohen's d is reported for each comparison:

$$d = \frac{M_2 - M_1}{SD_{Pooled}}. \tag{6}$$

In addition to traditional inferential tests, the Bayes factor (BF) was computed for comparisons of response variability for the both-cue condition and optimal integration (Jarosz & Wiley, 2014). We considered a Bayes factor (null/alternative) greater than 3 as adequate evidence that performance in the both-cue condition did not differ from the optimal MLE prediction and conversely, a Bayes factor less than 1/3 as adequate evidence that observed and predicted performance differed. If the p-value did not reach significance and the Bayes factor was between 1 and 3, cues were considered to be combined near-optimally (cf. Chen et al., 2017). Following suggestions by Rouder et al. (2009), we used a central Cauchy distribution as the prior with scale r on effect size set to 0.707. This prior is the default setting in many current statistical packages for calculating the Bayes factor (e.g., *BayesFactor* package for the R Environment; Morey & Rouder, 2015; R Core Team, 2019). As shown by Rouder et al. (2009), changes in scale r seldom result in changes in interpreting the Bayes factor.

Results and discussion

Response accuracy

Response accuracy was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,48) = 8.60$, GG epsilon = 0.93, $p < 0.001$, $\eta_p^2 = 0.39$ (BF = 0.85). Planned comparisons showed that participants were more accurate in the both-cue condition ($M = 0.89$, $SD = 0.27$) than the subset-B condition ($M = 1.09$, $SD = 0.28$), $t(24) = 3.74$, $p = 0.001$, $d = 0.72$ (BF = 0.04). Participants were equally accurate in the both-cue and subset-A ($M = 0.95$, $SD = 0.21$) conditions, $t(24) = 1.31$, $p = 0.203$, $d = 0.26$ (BF = 2.21).
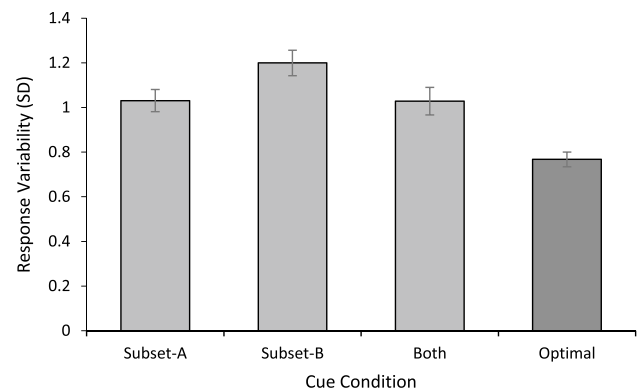

Fig. 3 Response variability (SD) as a function of cue condition in Experiment 1

Response variability

Response variability was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,48) = 7.03$, GG epsilon = 0.99, $p = 0.002$, $\eta_p^2 = 0.23$ (BF = 0.55) (Fig. 3). Planned comparisons revealed reduced response variability in the both-cue condition ($M = 1.03$, $SD = 0.31$) than the subset-B condition ($M = 1.20$, $SD = 0.28$), $t(24) = 3.12$, $p = 0.005$, $d = 0.62$ (BF = 0.11). Response variability in the both-cue condition was not significantly different than the subset-A condition ($M = 1.03$, $SD = 0.25$), $t(24) = 0.05$, $p = 0.959$, $d = 0.01$ (BF = 4.74). Response variability in the both-cue condition was significantly greater than optimal integration ($M = 0.77$, $SD = 0.17$), $t(24) = 5.59$, $p < 0.001$, $d = 1.12$ (BF < 0.01).

The results of Experiment 1 support the hierarchical hypothesis; that is, participants chose to use the most reliable of the two subsets (i.e., subset-A) when both subsets were presented during testing. Although we did not predict that either subset would be more reliable than the other, it was observed that participants tended to spend more time viewing the location of the target relative to the house than any other landmark. The house might have been a more reliable cue given its size and shape; the sharp edges of the house provide a salient reference point to which the location of the target can be encoded. Cue salience has been shown to be a critical component of cue reliability and weighting and can be determined by a multitude of factors such as a landmark's physical properties (Chen et al., 2017). Thus, it is not surprising that the landmarks composing each subset were not equal in this regard.

Although participants' response variabilities across cue conditions suggest a lack of cue integration, over a third of our participants ($n = 9$) reported utilizing a configural cue during encoding after the experimental session was over

Did you use any strategies while performing the task? If so, please describe them below. If not, just write "No".

I gauged the distance between the four landmarks by positioning myself ~~[crossed out]~~ on a grid of those landmarks.
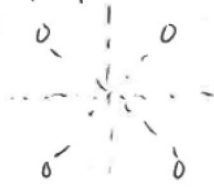
In terms of the four landmarks, these diagonal and horizontal lines were ~~[crossed out]~~ what ~~I positioned myself~~ I based my position off of.

Did you use any strategies while performing the task? If so, please describe them below. If not, just write "No".

I tried to imagine the landmarks as a rectangle, and I ~~[crossed out]~~ imagined the yellow pole as a mark on the grid in the rectangle. Then, in the test phase I tried to estimate the point on the "grid".

Fig. 4 Example strategy reports from configural strategy users

(Fig. 4). Thus, the results of Experiment 1 do not rule out the possibility that participants can integrate configural cues with individual landmark vectors to remember a target location. In Experiment 2, we drew inspiration from Spetch et al. (1996, 1997) and encouraged participants to utilize a configural cue strategy by equalizing cue salience across subsets and providing verbal instruction about the configural nature of the landmarks.

Experiment 2

Because only a minority of participants reported using configural strategies in Experiment 1, we attempted to prime participants to use a configural approach in Experiment 2 by eliminating the number of discriminant features across landmarks and by instructing participants to consider the configural structure of the landmark array. We also randomized the landmarks composing a subset across trials, with the only constraint being that two landmarks in a subset must be adjacent to one another.

Methods

Participants

Participants ($N = 20$; age $M = 22.95$, $SD = 5.41$; 10 females) recruited from the Vanderbilt community participated in exchange for monetary compensation. Although a smaller sample size than in Experiment 1, numerous experiments have demonstrated effects of cue combination in spatial navigation with similar sample sizes (e.g., Chen et al., 2017; Nardini et al., 2008; Petrini et al., 2016). Given the medium to large effect sizes found in Experiment 1, the sample size for the current experiment is justified. A trial was considered an outlier if the response error fell above three times the interquartile range above the third quartile for a given cue condition. Zero trials were cut using this criterion.

Materials and procedure

The immersive virtual environment was rendered in Unity. The environment was displayed in the HTC Vive Pro HMD with a resolution of 1440×1600 per eye, refreshed at 90 Hz, and a field-of-view of 110 degrees diagonally. Participants used HTC Vive's wireless controller to progress throughout the experiment. Position and orientation tracking were supported by HTC Vive's Lighthouse tracking system, with a 4×4 m tracking space. The size of the room was 7×5 m. Wireless tracking of the HMD was supported with the Vive Pro Wireless Adapter. The experiment was conducted on a computer with an Intel Core i9-9900K processor, 32 GB of RAM, and a GeForce RTX 2070 graphics card.

The virtual environment looked nearly identical to that used in Experiment 1, except that the landmarks were replaced with four white cylinders, distinguished by uniquely colored stripes (see Fig. 5). Configural cues might
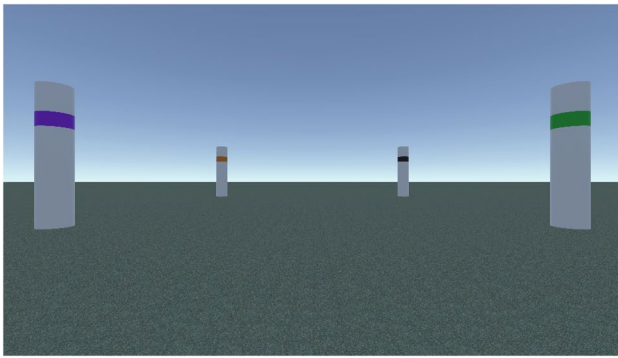
Fig. 5 Ground-level perspective of the virtual environment in Experiment 2

Fig. 6 Response variability (SD) as a function of cue condition in Experiment 2

have been used by navigators in Spetch et al.'s (1997) study because landmark identifiers were absent, making it impossible to represent vectorized relationships between the target and individual landmarks. Because we were interested in the combination of individual landmark vectors with configural cues, landmarks were created to be highly similar, to further encourage adoption of configural cues, while retaining identifying properties. Unlike Experiment 1, landmarks composing subsets were randomly selected for each trial, with the constraint that subsets were composed by adjacent landmarks.

Before the experiment began participants were told that the landmarks were arranged as a square, with each landmark representing the corner of the square, and that they should use this configuration to help them remember the location of the post. Based on strategy reports from Experiment 1, participants were told that one strategy is to represent the target location as a point on a grid defined by the landmarks. This instruction was repeated via textual display that appeared in the HMD at the beginning of every other block, starting with the first. The rest of the experimental procedure was identical to Experiment 1. Only three participants reported being active gamers and only two reported playing at least five hours a week, thus these variables are not considered further.

Results and discussion

Response accuracy

Response accuracy was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was not significant, $F(2,38) = 1.79$, GG epsilon $= 0.88$, $p = 0.186$, $\eta_p^2 = 0.09$ (BF $= 4.40$). Planned comparisons showed that participants were no more accurate in the both-cue condition ($M = 1.18$, $SD = 0.38$) than in either the subset-A ($M = 1.31$, $SD = 0.44$)
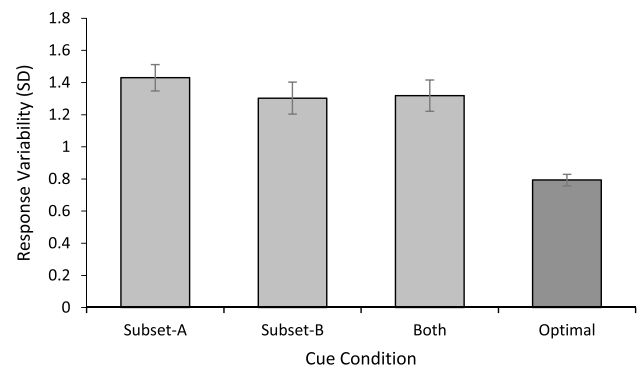
or subset-B conditions ($M = 1.18$, $SD = 0.37$), $t(19) = 1.55$, $p = 0.137$, $d = 0.35$ (BF $= 1.54$) and $t(19) = 0.11$, $p = 0.917$, $d = 0.02$ (BF $= 4.28$), respectively.

Response variability

Response variability was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was not significant, $F(2,38) = 1.41$, GG epsilon $= 0.94$, $p = 0.258$, $\eta_p^2 = 0.07$ (BF $= 4.97$) (Fig. 6). Planned comparisons showed that response variability in the both-cue condition ($M = 1.32$, $SD = 0.43$) was no different than response variability in either the subset-A ($M = 1.43$, $SD = 0.37$) or subset-B conditions ($M = 1.30$, $SD = 0.45$), $t(19) = 1.51$, $p = 0.015$, $d = 0.34$ (BF $= 1.62$) and $t(19) = 0.19$, $p = 0.849$, $d = 0.04$ (BF $= 4.23$), respectively. Response variability in the both-cue condition was significantly greater the optimal cue integration ($M = 0.79$, $SD = 0.16$), $t(19) = 5.78$, $p < 0.001$, $d = 1.29$ (BF < 0.01).

As in Experiment 1, these results support the hierarchical hypothesis. Response variability in the combined-cue condition was equal to the response variabilities from both subset conditions, suggesting that both subsets were equally reliable and that participants chose randomly amongst which landmarks to attend to during encoding. Despite encouragement to encode the target location relative to the landmark configuration, we did not find any evidence of integration of configural information with individual landmark subsets. Furthermore, we did not observe any evidence that landmark subsets were integrated at all. It is possible that cue integration is too difficult given the current task demands. For example, the target location is immersed within the landmark configuration rendering it impossible for participants to view all four landmarks from any given target location at once (from most locations, not even any three landmarks could be viewed simultaneously). Thus, encoding the location of the target relative to every landmark is highly

cognitively demanding. In Experiment 3, we simplified the task such that there were only two landmark cues and the target location could be viewed relative to both landmarks simultaneously.

Experiment 3

The results of Experiments 1 and 2 support the hierarchical hypothesis, such that participants tended to rely on only one of the landmark subsets when both were presented during testing. Participants may have been unable to integrate the two subsets because the task demands were too difficult (e.g., the entire configuration of landmarks could not be seen from any given orientation). In Experiment 3, we employed a simpler web-based spatial memory task that was completed online using a keyboard. One principal motivation for implementing this and following experiments online was to ensure the safety of participants during the COVID-19 pandemic. During the learning phase, participants viewed a target in the presence of two landmarks, a tower and a rocket-ship, both of which could be seen simultaneously. The learning phase was completed by moving to the target location using the arrow keys. Then, the participant was transported to another location during the testing phase in which only the landmark closest to or furthest from the target were available or both landmarks were available. The participant then used the arrow keys to place the target at its original location using the available cue(s). The current experiment was designed to test whether participants would be able to integrate the two landmark cues under simpler task demands than Experiments 1 and 2 and to extend previous work using a similar paradigm (e.g., Du et al., 2017) to a more ecologically valid context.

Methods

Participants

Participants ($N = 23$; age $M = 22.32$, $SD = 2.56$; 16 females) were recruited through SONA and completed the experiment online. Data for two additional participants were excluded due to server-side issues interfering with recording of trial data. A trial was considered an outlier if the response error fell above three times the interquartile range above the third quartile for a given cue condition. 0.02% of trials were cut using this criterion.

Materials and procedure

The virtual environment was rendered in Unity using the same computer as Experiment 2 and was published as a
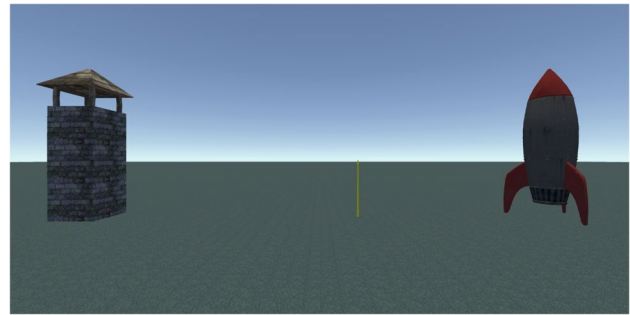

Fig. 7 Image of the environment as viewed by the participant in Experiment 3. During the learning phase, the participant began each trial 10 m away from and facing the midpoint between the two landmarks. During the test phase, the participant's perspective was position 7, 8, or 9 m away from and facing the midpoint between the two landmarks

WebGL application on a custom Apache2 webserver (http://apache.org). The environment consisted of an infinite ground plane and two landmarks, a tower and a rocket-ship, which were placed 12 m apart. A yellow post marking the target location appeared at a randomly chosen location between the two landmarks along a 6 m range, whose midpoint was the midpoint between the two landmarks (Fig. 7). Which side a landmark was on (tower to the left or rocket-ship to the left) was determined randomly for each trial. The perspective adopted by each participant was 1.5 m above the ground plane. The participant moved forward and backward using the up and down arrow keys and turned left and right using the left and right arrow keys, respectively.

After the participant input their demographic information, displayed text instructed the participant that they would complete a series of trials involving a learning and a test phase. The participant was instructed that they were to learn the location of the target relative to the landmarks and then walk to it to before the testing phase would begin. Then, during the testing phase, their goal was to place the target back at its original location, however, with the catch that only one of the landmarks would be available during testing on some trials, so it was important to learn the location of the target relative to both landmarks.

Each participant completed a practice block of three trials, one for each of three cue conditions: close cue, far cue, and both cues. During the learning phase, the perspective of the participant was positioned 10 m away from and facing the midpoint between the two landmarks with both landmarks visible on screen. Text displayed at the top of the screen instructed the participant to "walk" over to the yellow post until their perspective was at the yellow post's location. Once they were at the yellow post, they were able to press the spacebar to move onto the testing phase. The participant had as much time as needed to complete the learning phase. During the test phase, the participant's perspective was

positioned 7, 8, or 9 m (determined randomly) away from and facing the midpoint of the landmarks such that both landmarks would be available if present (see below), and they were no longer able to move or rotate using the arrow keys. The yellow post appeared at the midpoint between the landmarks and its height was increased by a multiple of 1.67 to eliminate possible cues stemming from the triangle formed by the tops of the yellow post and landmarks. During close-cue and far-cue trials, the landmark that was furthest from or closest to the target during learning was removed from the environment, respectively. During both-cue trials both landmarks remained visible. Instructions displayed at the top of the screen instructed the participant to place the yellow post back at its original location during learning using the left and right arrow keys, and then press the space-bar to confirm their response. Following the practice block, displayed text instructed the participant that they would now complete the main experimental trials. The participant completed ten blocks of three trials, one for each of the three cue conditions, in a random order. The debriefing was displayed at the end of the experiment.

Results and discussion

Response accuracy

Response accuracy was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,44) = 39.71$, GG epsilon $= 0.60$, $p < 0.001$, $\eta_p^2 = 0.64$ (BF < 0.01). Planned comparisons showed that participants were more accurate in the both-cue condition ($M = 0.43$, $SD = 0.19$) than in the close-cue ($M = 0.70$, $SD = 0.47$) and far-cue conditions ($M = 1.78$, $SD = 1.12$), $t(22) = 3.22$, $p = 0.004$, $d = 0.67$ (BF = 0.09) and $t(22) = 6.45$, $p < 0.001$, $d = 1.34$ (BF < 0.01), respectively. Participants were more accurate in the close-cue condition than in the far-cue condition, $t(22) = 6.71$, $p < 0.001$, $d = 1.40$ (BF < 0.01).

Response variability

Response variability was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,44) = 38.10$, GG epsilon $= 0.62$, $p < 0.001$, $\eta_p^2 = 0.63$ (BF < 0.01) (Fig. 8). Planned comparisons showed that response variability in the both-cue condition ($M = 0.52$, $SD = 0.25$) was less than response variability in both the close-cue ($M = 0.87$, $SD = 0.55$) and far-cue conditions ($M = 2.16$, $SD = 1.32$), $t(22) = 3.39$, $p = 0.003$, $d = 0.71$ (BF = 0.07) and $t(22) = 6.50$, $p < 0.001$, $d = 1.36$ (BF < 0.01), respectively. Response variability in the close-cue condition
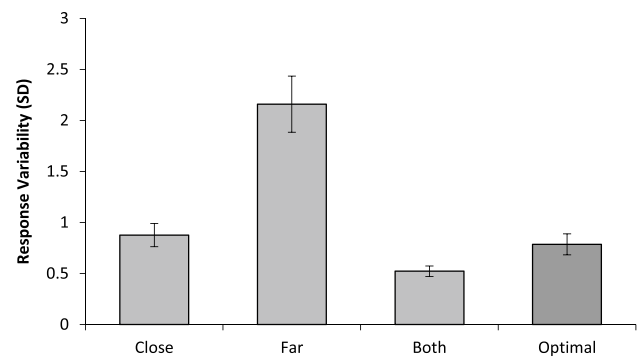

Fig. 8 Response variability (SD) as a function of cue condition in Experiment 3

was less than response variability in the far-cue condition, $t(22) = 6.22$, $p < 0.001$, $d = 1.30$ (BF < 0.01). Response variability in the both-cue condition was significantly less than optimal cue integration ($M = 0.79$, $SD = 0.49$), $t(22) = 2.28$, $p = 0.009$, $d = 0.59$ (BF = 0.54).

Surprisingly, participants' performance during both-cue trials conformed to the supra-optimal hypothesis, suggesting that participants combined the information from both landmarks with a latent cue that manifested as a result of the presence of both landmarks. Although there were only two landmarks in the present experiment, eliminating the possibility of a gestalt configural cue stemming from the landmarks themselves (Mou & Spetch, 2013), it is possible that participants were able to triangulate between their standing position and the two landmarks to self-localize and compute an egocentric bearing to the target location (Klatzky, 1998). During learning, the participant always faced the midpoint of the two landmarks at the beginning of each trial, and thus might have been able to encode the location of the goal relative to their initial heading and position (egocentric bearing). However, this heading might have been defined relative to the landmarks, and thus unavailable during trials in which only one landmark was available at testing.

Another possibility is that the sampling space of all possible target locations was available at testing during both-cue trials. That is, when both landmarks were available, the participant might have known that the target location must be encompassed by the space between the landmarks. Thus, when only one landmark was available during testing, the boundary specified by the removed landmark would no longer have been available, obfuscating the sampling space. By defining the space of possible targets during both-cue trials, participants might have been able to reduce the variability of their responses beyond the contributions of each landmark alone.

Fig. 9 Aerial perspective adopted by participant during the testing phase in Experiment 4. The participant was positioned 12 m above and facing the midpoint between the two landmarks

Fig. 10 Response variability (SD) as a function of cue condition in Experiment 4

Experiment 4

In Experiment 4, we tested whether the supra-optimal effect observed in Experiment 3 was due to participants encoding the target relative to the heading defined by the landmarks or the availability of the sampling space of possible targets when both landmarks were present at testing. We refer to the former as the *egocentric-bearing hypothesis* and the latter as the *boundary-cue hypothesis*. The learning phase was the same as Experiment 3, but during the testing phase, the participant's perspective was teleported aerially to be 12 m directly above and facing the midpoint between the two landmarks (Fig. 9). If participants encoded the location of the target relative to the heading defined by the landmarks, then performance during both-cue trials should be consistent with optimal integration as this cue should be disrupted between learning and testing phases. However, if the supra-optimal effect resulted from the available of the sampling space during both-cue trials, performance should again be supra-optimal.

Methods

Participants

Participants ($N = 22$; age $M = 22.85$, $SD = 2.83$; 15 females) were recruited through SONA and completed the experiment online. Data from two additional participants were excluded due to server-side issues interfering with recording of trial data. A trial was considered an outlier if the response error fell above three times the interquartile range above the third quartile for a given cue condition. 0.02% of trials were cut using this criterion.
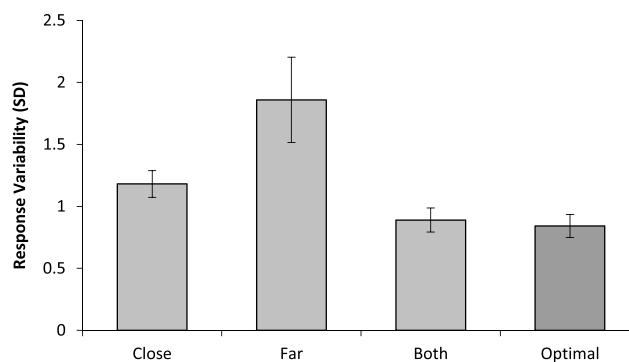
Materials and procedure

The materials and procedure were the same as Experiment 3 except that the perspective of the participant was teleported aerially to be 12 m directly above and facing the midpoint between the two landmarks during the testing phase. Although the viewing perspective during learning was constant across trials, a strategy of encoding the target relative to the edges of the screen would not be effective because the testing phase was completed from a ground perspective.

Results and discussion

Response accuracy

Response accuracy was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,42) = 5.68$, GG epsilon $= 0.51$, $p = 0.026$, $\eta_p^2 = 0.21$ (BF $= 0.16$). Planned comparisons showed that participants were more accurate in the both-cue condition ($M = 0.75$, $SD = 0.42$) than in the close-cue ($M = 0.97$, $SD = 0.39$) and far-cue conditions ($M = 2.03$, $SD = 2.34$), $t(21) = 3.16$, $p = 0.005$, $d = 0.67$ (BF $= 0.11$) and $t(21) = 2.49$, $p = 0.021$, $d = 0.53$ (BF $= 0.37$), respectively. Participants were more accurate in the close-cue condition than in the far-cue condition, $t(21) = 2.23$, $p = 0.037$, $d = 0.48$ (BF $= 0.58$).

Response variability

Response variability was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,42) = 6.32$, GG epsilon $= 0.56$, $p = 0.016$, $\eta_p^2 = 0.23$ (BF $= 0.16$) (Fig. 10). Planned comparisons showed that response variability in the both-cue condition ($M = 0.89$,

$SD = 0.46$) was less than response variability in both the close-cue ($M = 1.18$, $SD = 0.50$) and far-cue conditions ($M = 1.86$, $SD = 1.61$), $t(21) = 2.82$, $p = 0.010$, $d = 0.60$ (BF = 0.21) and $t(21) = 2.79$, $p = 0.011$, $d = 0.59$ (BF = 0.22), respectively. Response variability in the close-cue condition was less than response variability in the far-cue condition, $t(21) = 2.11$, $p = 0.047$, $d = 0.45$ (BF = 0.71). Response variability in the both-cue condition was consistent with optimal cue integration ($M = 0.79$, $SD = 0.49$), $t(21) = 0.44$, $p = 0.663$, $d = 0.09$ (BF = 4.11).

The results of Experiment 4 are consistent with the egocentric-bearing hypothesis. When participants' perspective was teleported above the landmarks and facing downward during the testing phase, performance in the both-cue condition was consistent with optimal cue integration. If the supra-optimal effect observed in Experiment 3 were due to the sampling space of possible target locations being defined by the presence of both landmarks at test, the supra-optimal effect should have also been observed here. However, another plausible explanation of why participants were not supra-optimal in their response variability during both-cue trials is that the perspective switch might have induced greater cognitive demands, particularly on working memory. The contrast in results between Experiment 3 and Experiments 1 and 2 suggest that task demands impact participants' ability to integrate landmark cues. If the lack of supra-optimal performance observed in Experiment 4 was due to the increased working memory demands during perspective switching, supra-optimal performance should be observed when the perspective at encoding and retrieval is aerial. However, it is also plausible that egocentric bearings are not computed from aerial perspectives as such perspectives are not often assumed in daily navigation. Thus, if a ground-level perspective is required for participants to compute an egocentric bearing between themselves and the target, optimal integration should be observed.

Experiment 5

In Experiment 5, we tested whether the supra-optimal effect observed in Experiment 3 is exclusive to taking a ground-level perspective at both learning and testing or if congruency in perspective across learning and test phases will invoke an additional latent cue, namely the triangulation of the landmarks with the participants position to form an egocentric bearing to the target, by having participants learn and test from an aerial perspective. We refer to the former as the *ground-level hypothesis* and the latter as the *congruency hypothesis*. If participants triangulate their position with the landmarks to self-localize from an aerial perspective, we should observe the supra-optimal effect here as well.

Methods

Participants

Participants ($N = 24$; age $M = 21.00$, $SD = 1.87$; 8 females) were recruited through SONA and completed the experiment online. Data from one additional participant were excluded due to server-side issues interfering with recording of trial data. A trial was considered an outlier if the response error fell above three times the interquartile range above the third quartile for a given cue condition. 0.01% of trials were cut using this criterion.

Materials and procedure

The materials and procedure were the same as in Experiments 3 and 4 with two exceptions: first, participants were positioned at a randomly chosen distance between 10 and 20 m directly above and facing the midpoint between the two landmarks during learning and test phases; second, participants were given 5 s to learn the location of the yellow post before transitioning to the testing phase. The duration of the learning phase was chosen to approximate the amount of time needed to travel to the yellow post in Experiments 3 and 4, thus maintaining consistency in the amount of learning time. A new viewing distance was selected for the learning and test phases to ensure that participants could not use distance of the target to the edges of the screen as a cue.

Results and discussion

Response accuracy

Response accuracy was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,46) = 17.22$, GG epsilon = 0.75, $p < 0.001$, $\eta_p^2 = 0.43$ (BF = 0.03). Planned comparisons showed that participants were more accurate in the both-cue condition ($M = 0.66$, $SD = 0.32$) than in the close-cue ($M = 0.78$, $SD = 0.25$) and far-cue conditions ($M = 1.06$, $SD = 0.48$), $t(23) = 2.35$, $p = 0.028$, $d = 0.48$ (BF = 0.49) and $t(23) = 6.00$, $p < 0.001$, $d = 1.22$ (BF < 0.01), respectively. Participants were more accurate in the close-cue condition than in the far-cue condition, $t(23) = 3.21$, $p = 0.004$, $d = 0.66$ (BF = 0.09).

Response variability

Response variability was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant,
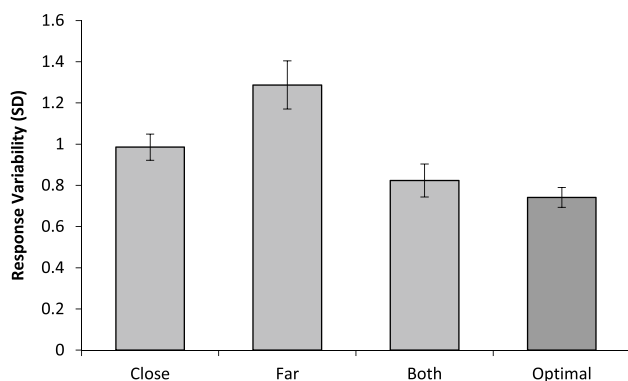
Fig. 11 Response variability (SD) as a function of cue condition in Experiment 5

$F(2,46) = 13.99$, GG epsilon $= 0.76$, $p < 0.001$, $\eta_p^2 = 0.38$ (BF $= 0.05$) (Fig. 11). Planned comparisons showed that response variability in the both-cue condition ($M = 0.82$, $SD = 0.39$) was less than response variability in both the close-cue ($M = 0.99$, $SD = 0.31$) and far-cue conditions ($M = 1.28$, $SD = 0.57$), $t(23) = 2.57$, $p = 0.017$, $d = 0.52$ (BF $= 0.32$) and $t(23) = 5.17$, $p < 0.001$, $d = 1.06$ (BF < 0.01), respectively. Response variability in the close-cue condition was less than response variability in the far-cue condition, $t(23) = 2.78$, $p = 0.015$, $d = 0.57$ (BF $= 0.22$). Response variability in the both-cue condition was near-optimal ($M = 0.74$, $SD = 0.24$), $t(23) = 1.40$, $p = 0.174$, $d = 0.29$ (BF $= 1.97$).

The results of Experiment 5 support the ground-level hypothesis. Although participants learned and were tested on the location of the yellow post from an aerial perspective, near-optimal cue integration was observed. If congruency of perspective across the learning and test phases was sufficient for triangulating between one's position and the landmarks during both-cue trials, supra-optimal performance should have been observed. The results of Experiments 3–5 suggest that supra-optimal cue performance is possible when the landmarks are presented at a ground-level perspective during learning, testing, or both. The results of these experiments, however, cannot answer directly whether congruency in perspective is required across learning and test phases for supra-optimal cue performance to occur. Experiment 4 demonstrated that a ground-level perspective during learning is not enough to produce supra-optimal cue performance, but we have not ruled out the possibility that a ground-level perspective at the test phase only is sufficient. It is possible that having the ground-level perspective at testing instantiates the latent cue during retrieval, but that this cue is necessary during encoding. It is also possible that a ground-level perspective is necessary at both encoding and retrieval to observe supra-optimal cue performance. We test these hypotheses in Experiment 6.

Experiment 6

Although Experiments 3–5 suggest that a ground-level perspective during retrieval (test phase) is necessary to observe supra-optimal cue performance, it is impossible to determine whether such a perspective is only required during retrieval or both encoding and retrieval. We refer to the former hypothesis as the *retrieval hypothesis* and the latter as the *encoding-plus-retrieval hypothesis*. We tested these hypotheses by having participants encode the location of the yellow post from an aerial perspective (learning phase), and then retrieve its location from a ground-level perspective (test phase). The retrieval hypothesis predicts that performance should be supra-optimal during both-cue trials, whereas the encoding-plus-retrieval hypothesis predicts optimal or sub-optimal performance.

Methods

Participants

Participants ($N = 22$; age $M = 22.82$, $SD = 4.49$; 5 females) were recruited through SONA and completed the experiment online. Data from one additional participant was excluded due to server-side issues interfering with recording of trial data. A trial was considered an outlier if the response error fell above three times the interquartile range above the third quartile for a given cue condition. Less than 0.01% of trials were cut using this criterion.

Materials and procedure

The materials and procedure were the same as Experiment 5 except that participants were tested from a ground-level perspective in the same manner as in Experiment 3. During testing, participants were positioned 1.5 m above the ground plane and facing the midpoint between the two landmarks from 7, 8, or 9 m.

Results and discussion

Response accuracy

Response accuracy was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,42) = 5.63$, GG epsilon $= 0.88$, $p = 0.010$, $\eta_p^2 = 0.21$ (BF $= 0.35$). Planned comparisons showed that participants were more accurate in the both-cue condition ($M = 0.93$, $SD = 0.31$) than in the far-cue condition ($M = 1.19$, $SD = 0.48$), $t(21) = 2.35$,
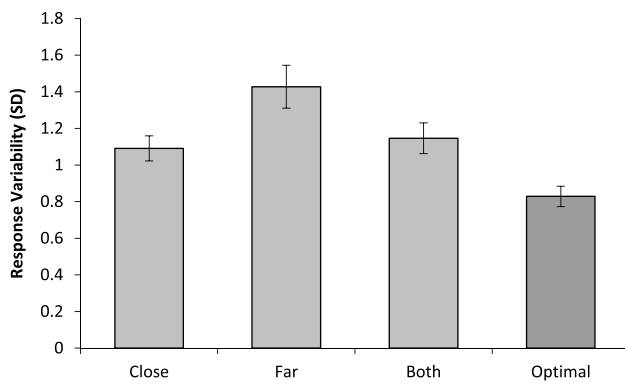
Fig. 12 Response variability (SD) as a function of cue condition in Experiment 6

$p = 0.028$, $d = 0.51$ (BF = 0.46), but not the close-cue condition ($M = 0.90$, $SD = 0.25$), $t(21) = 0.48$, $p = 0.634$, $d = 0.10$ (BF = 4.04). Participants were more accurate in the close-cue condition than in the far-cue condition, $t(21) = 2.99$, $p = 0.007$, $d = 0.64$ (BF = 0.15).

Response variability

Response variability was examined using a repeated-measures ANOVA, with cue condition as a within-subjects factor. The main-effect of cue condition was significant, $F(2,42) = 5.74$, GG epsilon = 0.95, $p = 0.007$, $\eta_p^2 = 0.22$ (BF = 0.46) (Fig. 12). Planned comparisons showed that response variability in the both-cue condition ($M = 1.15$, $SD = 0.39$) was less than response variability in the far-cue condition ($M = 1.43$, $SD = 0.31$), $t(21) = 2.39$, $p = 0.026$, $d = 0.51$ (BF = 0.45), but not the close-cue conditions ($M = 1.09$, $SD = 0.32$), $t(21) = 0.57$, $p = 0.572$, $d = 0.12$ (BF = 3.87). Response variability in the close-cue condition was less than response variability in the far-cue condition, $t(21) = 3.21$, $p = 0.004$, $d = 0.68$ (BF = 0.10). Response variability in the both-cue condition was suboptimal ($M = 0.82$, $SD = 0.26$), $t(21) = 3.55$, $p = 0.002$, $d = 0.76$ (BF = 0.05).

The results of Experiment 6 are concordant with the encoding-plus-retrieval hypothesis. When participants encoded the target from an aerial perspective and were tested from a ground-level perspective, they were suboptimal in

their response variability. In fact, participants demonstrated cue domination such that response variability in the both-cue condition was equal to the response variability of the most reliable (close) cue. Thus, these results also support the hierarchical hypothesis described earlier. Overall, the results of Experiments 3–6 suggest that supra-optimal cue performance with two visual landmark cues requires a ground-level perspective at both encoding and retrieval. We suspect that participants were triangulating between their position and the two landmarks during encoding, allowing them to encode the location of the target relative to their egocentric bearing (Klatzky, 1998). When the participant was then tested during both-cue trials, such a representation was recoverable. Otherwise, the representation was disrupted.

General discussion

Since Cheng et al. (2007) proposed that spatial cues are optimally integrated according to models of MLE, spatial cue integration during navigation has received considerable attention. Although a substantial number of studies have demonstrated that human navigators can optimally integrate visual and body-based cues during navigation, little attention has been paid to the problem of integrating sensory cues of the same type (e.g., multiple landmark cues). The few studies that have investigated this problem provide conflicting evidence, proving that the process of integrating multiple landmark cues continues to be a vexing problem. For example, Baguley et al. (2006) and Clark et al. (2013) showed little to no benefit of having access to more than one landmark when localizing a target, whereas Du et al. (2017) showed optimal integration of two landmarks when participants were given training with both landmarks present. Mou and Spetch (2013) showed further that the configural geometry of an array of objects can be integrated with the individual object-to-object vectors, leading to supra-optimal performance. In the current study, we examined adult humans' ability to integrate visual landmark cues across six experiments (Table 1).

Most notably, we showed supra-optimal performance during both-cue trials when participants encoded and retrieved the target location from a ground-level perspective and task demands were relatively simple (Experiment 3). On

Table 1 Summary of the six presented experiments displaying the type of virtual reality employed (immersive or desktop), the perspective assumed by the participant during encoding and retrieval, the spatial cues used, and the results

	Type of VR	Encoding	Retrieval	Cues	Results
Experiment 1	Immersive	Ground	Ground	4 landmarks	Hierarchical
Experiment 2	Immersive	Ground	Ground	4 landmarks	Hierarchical
Experiment 3	Desktop	Ground	Ground	2 landmarks	Supra-optimal
Experiment 4	Desktop	Ground	Aerial	2 landmarks	Optimal
Experiment 5	Desktop	Aerial	Aerial	2 landmarks	Optimal
Experiment 6	Desktop	Aerial	Ground	2 landmarks	Hierarchical

the other hand, when task demands were difficult (Experiments 1 and 2), or the perspective was aerial at encoding or retrieval (but not both; Experiments 4–6), we observed a range of optimal and suboptimal performance. We discuss the implications of these results on the integration of visual landmark cues in spatial memory in turn.

Spetch and colleagues (1997) observed that humans could use configural information from landmarks during goal-localization. Participants in their study consistently located the goal in the center of the configuration despite expansion of the configuration along one or both dimensions, contrasting with the results of pigeons performing the same task, who were biased toward individual landmarks during expansions. In Experiments 1 and 2, we tested participants' ability to integrate configural information from landmarks with information from individual landmarks during a goal-localization task. However, the results indicated that participants were unable to integrate even the landmark subsets. In Experiment 1, when landmarks were highly dissimilar, participants' performance during both-cue trials suggested that the most reliable subset of landmarks (the house and the rock) dominated. However, when probed about their strategy at the end of the experiment, some participants reported using a configural strategy (e.g., thinking about the space as a grid defined by the landmarks). In Experiment 2, we told participants to consider the configural nature of the landmarks repeatedly throughout the experiment as well as changed the landmarks to be more similar, yet still distinguishable, to further increase the relative saliency of the configuration. However, no cue integration was observed.

Although Spetch et al. (1997) observed the use of configural cues, their study was not designed or intended to test cue integration as in the current study and has many methodological differences. Most notably, the landmarks in their study were identical, making the configural cue the most salient cue for locating the target. Furthermore, the target was always located in the center of the configuration. However, another plausible explanation for the lack of cue integration in Experiments 1 and 2 is that the task was too demanding on working memory. From no target location were all the landmarks visible at one, meaning that the individual vector from one landmark to the target location would need to be stored in working memory while encoding the spatial relations with the other landmarks. It is possible that participants with superior spatial working memory might have been able to integrate the landmark subsets, and potentially the configuration as well, but future research is required to address this issue.

To test whether participants could integrate landmark cues during goal-localization under simpler task demands, in Experiment 3 participants learned the location of a target positioned between two landmarks, both of which were visible from a single location. Surprisingly, participants

demonstrated supra-optimal performance during both-cue trials. This finding is perplexing as the source of the latent cue is not obvious. Switching the perspective from ground-level to aerial (Experiment 4) resulted in optimal performance, ruling out the possibility that the bounds on the space of possible target locations induced by the presence of both cues served as the latent cue. One possibility is that a different type of configural cue was formed; specifically, one formed by the two landmarks and the virtual position of the participant. Triangulating between the landmarks and the standing position might have been sufficient to yield a reference frame for which the egocentric bearing from the participant to the target could be computed (Klatzky, 1998). However, when one of the landmarks was eliminated, such triangulation was no longer possible.

An alternative interpretation of the supra-optimal performance observed in Experiment 3 is that participants benefited from encoding specificity (Tulving & Thomas, 1973). That is, the similarity between encoding and retrieval during both-cue trials was not captured by the MLE model because the optimal predictions were made based on performance in single-cue trials in which the encoding and retrieval contexts were more dissimilar, leading to supra-optimal performance during both-cue trials. An advantage of employing conflict conditions is that they present both cues during encoding and retrieval, but the contexts differ (albeit subtly) as the cues are put in conflict during retrieval. Unfortunately, the current set of experiments cannot completely rule out effects of encoding specificity as we did not employ conflict trials. However, it should be noted that response variability in both-cue conditions in which both cues are consistent often mirror those of conflict trials (e.g., Chen et al., 2017; but see Zhang et al., 2020). Furthermore, results from a recent study in our lab has shown that the presence of both cues during encoding in single-cue trials did not impact performance relative to when only one cue was presented at both encoding and retrieval (Newman & McNamara, 2021). However, that study investigated the integration of visual and body-based cues, which are thought to be independent. Another interpretation of these results is that the egocentric-bearing cue dominated performance and was of such high reliability that use of this cue alone was enough to achieve supra-optimal performance. In other words, this latent cue was not integrated with the landmarks at all. Spetch et al.'s (1997) participants were able to use configural information alone to solve a goal-localization task when landmarks were indiscriminate, so such information does not necessarily need to be integrated with individual landmark cues to be utilized. Unfortunately, our experimental design does not differentiate between these two possibilities.[4]

---

[4] The authors thank two anonymous reviewers for these alternative interpretations of the data.

We also found that the perspective assumed during encoding and retrieval affected participants' ability to perform supra-optimally. Participants optimally integrated the landmarks when the encoding and retrieval perspectives were congruent but aerial (Experiment 5), suggesting that there is something unique about a ground-level perspective that elicits the latent cue. If the egocentric-bearing account of the results of Experiment 3 is accurate, the results of Experiment 5 could be explained by the assumption that egocentric bearings are not often formed or very useful from aerial perspectives in daily life. If pilots demonstrated supra-optimal performance on the same task, the egocentric-bearing account would be supported. However, a lack of supra-optimal performance would not necessarily rule out the egocentric-bearing account as other unknown contributing variables might inhibit the use of egocentric bearings from aerial perspectives. Another way to test the egocentric-bearing account would be to test how disrupting this cue influences performance when both encoding and retrieval occur from a ground-level perspective. For example, the egocentric bearing can be disrupted by changing the standing position of the participant, either by translation or rotation around the midpoint between the landmarks by 180°, between encoding and retrieval.[5] If the egocentric bearing does not act as a latent cue during both-cue trials, supra-optimal performance should be unaffected.

More evidence of the influence of perspective during encoding and retrieval on the integration of landmarks was observed in Experiment 6, wherein participants encoded the target from an aerial perspective and retrieved its location from a ground-level perspective. Contrary to Experiment 4, wherein the perspective switch between encoding and retrieval was reversed, participants were not optimal in their response variability, demonstrating dominance of the closest landmark during both-cue trials. Although this finding has little to say about the egocentric-bearing account of supra-optimal performance from a ground-level perspective, it does pose an interesting question as to why participants were able to integrate cues after having switched from a ground-level to an aerial perspective but not vice versa. Developmental work (e.g., Nardini et al., 2009) suggests that younger children have a difficult time switching between

from egocentric to allocentric representations of space during a search task. However, DeLoache (1989) showed that children of a similar age also had difficulty transferring allocentric spatial information from a scaled room model to a test room to locate a hidden target. Although older children were able to solve both tasks, suggesting bidirectional fluency in transfer between egocentric and allocentric representations, transferring from the latter to the former might be more difficult in some spatial tasks than in others. For example, navigators make more wayfinding errors with egocentrically aligned forward-up maps than static north-up maps (Münzer et al., 2020). Further research is required to address this issue.

Conclusion

In the current study, we showed that optimal cue integration of visual landmark cues is possible in naturalistic 3D environments. However, integrating multiple landmark cues may not be possible when the task is too demanding of working memory (Experiments 1 and 2). Furthermore, we showed that supra-optimal cue performance is possible when a target location is learned relative to two landmarks from a ground-level perspective and memory is tested from a ground-level perspective. We speculate that participants were able to triangulate between their standing position and the two landmarks to self-localize and establish and egocentric bearing to the target location. This process might not have been used when encoding and retrieval occurred from an aerial perspective, as navigation and computation of egocentric spatial parameters most often takes place from a ground-level perspective. Future research is necessary to uncover the nature of this latent cue and other factors surrounding its availability and utility during navigation with landmarks.

Appendix

We compute the standard deviation of response locations around the mean of those response locations. In this example, suppose that a participant made three responses, $R = [r1, r2, r3]$, on three trials with target locations, $T = [t1, t2, t3]$, where each response and target location have an x and y component. Table 2 lists example coordinate values for each response and target location. First, the response is centered on the target location by subtracting the target location vector from the response vector, yielding the adjusted response. Then, the mean response location is computed as the mean of the adjusted responses. In this example, the mean response location is, [1.67 −0.33]. For

---

[5] Although the standing position at the start of encoding and the standing position at retrieval were not identical in Experiment 3, undergoing a small translational shift forward or backward, such shifts might not have been large enough to fully disrupt the egocentric bearing. For example, large degrees of cue conflict are often required before the participant becomes consciously aware of the conflict (e.g., Zhao & Warren, 2015b). Thus, large translations in ground-level perspectives between encoding and retrieval might be necessary to fully disrupt an egocentric bearing.

each response, the Euclidean distance between the response and the mean response location is computed,

$$ED = \sqrt{\left(M_x - r_x\right)^2 + \left(M_y - r_y\right)^2}, \qquad (7)$$

where M_x is the x component of the mean response location. The standard deviation of responses can then be computed as,

$$\sigma_R = \sqrt{\frac{\sum ED^2}{n-1}}, \qquad (8)$$

where n is the number of observations; in this example, three. Note that in Experiments 3–6 responses are made along one dimension. Thus, the Euclidean distance is replaced by the distance between the unidimensional response and target location, which can be computed by simple subtraction.

Table 2 Example data for three trials

Trial	Response	Target	Adj. Response	Euclidean Dist
1	[5 6]	[6 2]	[−1 4]	5.09
2	[9 3]	[4 9]	[5 −6]	6.58
3	[7 11]	[6 10]	[1 1]	1.49

The adjusted response is computed as the target location vector subtracted from the response vector. The Euclidean distance between the adjusted response and the mean response location is then computed for each trial

Declarations

Conflict of interest None.

Ethics approval All experiments conducted as part of this study were approved by Vanderbilt University's internal review board.

Consent to participate All participants provided informed consent prior to participating in this study.

Consent for publication The authors consent to publication.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262.

Baguley, T., Lansdale, M. W., Lines, L. K., & Parkin, J. K. (2006). Two spatial memories are not better than one: Evidence of exclusivity in memory for object location. *Cognitive Psychology, 52*, 243–289. https://doi.org/10.1016/j.cogpsych.2005.08.001

Bates, S. L., & Wolbers, T. (2014). How cognitive aging affects multisensory integration of navigational cues. *Neurobiology of Aging, 35*(12), 2761–2769.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Optical Society of America, 20*(7), 1391–1397.

Biegler, R., & Morris, R. G. M. (1999). Blocking in the spatial domain with arrays of discrete landmarks. *Journal of Experimental Psychology, 25*(3), 334–351.

Boot, W. R., Kramer, A. F., Simons, D. J., Fabioni, M., & Gratton, G. (2008). The effect of video game playing on attention, memory, and executive control. *Acta Psychologica, 129*, 387–398.

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

Butler, J. S., Smith, S. T., Campos, J. L., & Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision, 10*(11), 1–13.

Chamizo, V. D. (2003). Acquisition of knowledge about spatial location: Assessing the generality of the mechanism of learning. *The Quarterly Journal of Experimental Psychology, 56B*(1), 102–113.

Chamizo, V. D., Sterio, D., & Mackintosh, N. J. (1985). Blocking and overshadowing between intra-maze and extra-maze cues: A test of the independence of locale and guidance learning. *The Quarterly Journal of Experimental Psychology, 37B*, 235–253.

Chen, X., McNamara, T. P., Kelly, J. W., & Wolbers, T. (2017). Cue combination in human spatial navigation. *Cognitive Psychology, 95*, 105–144.

Cheng, K., Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychological Bulletin, 133*(4), 625–637.

Clark, D. P. A., Dunn, A. K., & Baguley, T. (2013). Testing the exclusivity effect in location memory. *Memory, 21*, 512–523. https://doi.org/10.1080/09658211.2012.744421

DeLoache, J. S. (1989). Young children's understanding of the correspondence between a scale model and a larger space. *Cognitive Development, 4*, 121–139.

Du, Y., McMillan, N., Madan, C. R., Spetch, M. L., & Mou, W. (2017). Cue integration in spatial search for jointly learned landmarks but not for separately learned landmarks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1857–1871.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2010). Visual-vestibular cue integration for heading perception: Applications of optimal cue integration theory. *European Journal of

Neuroscience, 31, 1721–1729. https://doi.org/10.1111/j.1460-9568.2010.07207.x

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience, 15*(1), 146–154.

Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience, 29*(49), 15601–15612.

Friedmann, A., Ludvig, E. A., & Legge, E. L. G. (2013). Bayesian combination of two-dimensional location estimates. *Behavior Research, 45*, 98–107. https://doi.org/10.3758/s13428-012-0241-x

Frissen, I., Campos, J. L., Souman, J. L., & Ernst, M. O. (2011). Integration of vestibular and proprioceptive signals for spatial updating. *Experimental Brain Research, 212*(2), 163–176.

Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision, 9*(9), 1–20.

Hamilton, D. A., & Sutherland, R. J. (1999). Blocking in human place learning: Evidence from virtual navigation. *Psychobiology, 27*(4), 453–461.

Hardt, O., Hupbach, A., & Nadel, L. (2009). Factors moderating blocking in human place learning: The role of task instructions. *Learning and Behavior, 37*(1), 42–59. https://doi.org/10.3758/LB.37.1.42

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12), 967–992. https://doi.org/10.1167/4.12.1

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research, 39*, 3621–3629.

Jacobs, W. J., Laurance, H. E., & Thomas, K. G. F. (1997). Place learning in virtual space I: Acquisition, overshadowing, and transfer. *Learning and Motivation, 28*, 521–541.

Jacobs, W. J., Thomas, K. G. F., Laurance, H. E., & Nadel, L. (1998). Place learning in virtual space II: Topographical relations as one dimension of stimulus control. *Learning and Motivation, 29*, 288–308.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving, 7*, 2–9.

Kalia, A. A., Schrater, P. R., & Legge, G. E. (2013). Combining path integration and remembered landmarks when navigating without vision. *PLoS ONE, 8*(9), e72170. https://doi.org/10.1371/journal.pone.0072170

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial cognition—An interdisciplinary approach to representation and processing of spatial knowledge* (pp. 1–17). Berlin: Springer.

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology, 18*, 87–121.

McNamara, T. P., & Chen, X. (2020). Bayesian decision theory and navigation. https://doi.org/10.31234/osf.io/zduca

Morey, R., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs*. Retrieved from https://cran.r-project.org/web/packages/BayesFactor/index.html

Mou, W., & McNamara, T. P. (2002). Intrinsic frames of reference in spatial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 162–170.

Mou, W., & Spetch, M. L. (2013). Object location memory: Integration and competition between multiple context objects but not between observers' body and context objects. *Cognition, 126*(2), 181–197.

Münzer, S., Lörch, L., & Frankenstein, J. (2020). Wayfinding and acquisition of spatial knowledge with navigation assistance. *Journal of Experimental Psychology: Applied, 26*(1), 73–88.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology, 18*, 689–693.

Nardini, M., Thomas, R. L., Knowland, V. C. P., Braddick, O. J., & Atkinson, J. (2009). A viewpoint-independent process for spatial reorientation. *Cognition, 112*, 241–248.

Newman, P. M., & McNamara, T. P. (2021). A comparison of methods of assessing cue combination during navigation. *Behavior Research Methods, 53*(1), 390–398.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*, 2451–2468.

Petrini, K., Caradonna, A., Foster, C., Burgess, N., & Nardini, M. (2016). How vision and self-motion combine or compete during path reproduction changes with age. *Scientific Reports, 6*, 29163.

Philbeck, J. W., & O'Leary, S. (2005). Remembered landmarks enhance the precision of path integration. *Psicológica, 26*, 7–24.

Prados, J. (2011). Blocking and overshadowing in human geometry learning. *Journal of Experimental Psychology, 37*(1), 121–126. https://doi.org/10.1037/a0020715

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienne, Austria. https://www.R-project.org/

Ratliff, K. R., & Newcombe, N. S. (2008). Reorienting when cues conflict: Evidence for an adaptive combination view. *Psychological Science, 19*, 1301–1307.

Rodrigo, T., Arall, M., & Chamizo, V. D. (2005). Blocking and unblocking in a navigation task. *Psicológica, 26*, 229–241.

Rohde, M., van Dam, L. C. J., & Ernst, M. O. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research, 29*, 279–317.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237.

Sánchez-Moreno, J., Rodrigo, T., & Chamizo, V. D. (1999). Overshadowing in the spatial domain. *Animal Learning and Behavior, 27*(4), 391–398.

Shettleworth, S. J., & Sutton, J. E. (2005). Multiple systems for spatial learning: Dead reckoning and beacon homing in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 31*(2), 125–141.

Sjolund, L. A., Kelly, J. W., & McNamara, T. P. (2018). Optimal combination of environmental cues and path integration during navigation. *Memory & Cognition, 46*, 89–99. https://doi.org/10.3758/s13421-017-0747-7

Spetch, M. L., Cheng, K., & MacDonald, S. E. (1996). Learning the configuration of a landmark array: I. Touch-screen studies with pigeons and humans. *Journal of Comparative Psychology, 110*(1), 55–68.

Spetch, M. L., Cheng, K., MacDonald, S. E., Linkenhoker, B. A., Kelly, M. D., & Doerkson, S. R. (1997). Use of landmark configuration in pigeons and humans: II. Generality across search tasks. *Journal of Comparative Psychology, 111*(1), 14–24.

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology, 10*, 422–437.

Tcheang, L., Bulthoff, H. H., & Burgess, N. (2011). Visual influence on path integration in darkness indicates a multimodal representation of large-scale space. *Proceedings of the National Academy of Sciences of the United States of America, 108*(3), 1152–1157.

Tulving, E., & Thomas, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373.

Twyman, A. D., Holden, M. P., & Newcombe, N. S. (2018). First direct evidence of cue integration in reorientation: A new paradigm. *Cognitive Science, 42*, 923–936.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402.

Wang, L., & Mou, W. (2020). Effect of room size on geometry and features cue preference during reorientation: Modulating encoding strength or cue weighting. *Quarterly Journal of Experimental Psychology, 73*(2), 225–238.

Wang, L., Mou, W., & Dixon, P. (2018). Cue interaction between buildings and street configurations during reorientation in familiar and unfamiliar outdoor environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 631.

Xiao, C., Mou, W., & McNamara, T. P. (2009). Use of self-to-object and object-to-object spatial relations in locomotion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1137–1147.

Xu, Y., Regier, T., & Newcombe, N. S. (2017). An adaptive cue combination model of human spatial orientation. *Cognition, 163*, 56–66.

Zhang, L., Mou, W., Lei, X., & Du, Y. (2020). Cue combination used to update the navigator's self-localization, not the home location. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(12), 2314–2339.

Zhao, M., & Warren, W. H. (2015a). Environmental stability modulates the role of path integration in human navigation. *Cognition, 142*, 96–109.

Zhao, M., & Warren, W. H. (2015b). How you get there from here: Interaction of visual landmarks and path integration in human navigation. *Psychological Science, 26*(6), 915–924.