Cue Integration in Spatial Navigation – a Bayesian Approach

By

Xiaoli Chen

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

In

Psychology

December, 2014
Nashville, Tennessee

Approved:
Timothy P. McNamara, Ph.D.
Jonathan W. Kelly, Ph.D.
Gordon D. Logan, Ph.D.
Isabel Gauthier, Ph.D.

ACKNOWLEDGEMENTS

The completion of my dissertation reflects the efforts of many people. I am indebted to my advisor, Prof. Timothy McNamara, for his patient and inspiring instructions, his great availability, and recruiting me 5 years ago when all the other places I had applied for rejected me. I am grateful to my committee members, Isabel Gauthier, Gordon Logan, and Jonathan Kelly, for their helpful comments on my work. Their inputs were critical for the success of this project.

I would like to thank my labmate, Qiliang, who discussed literature with me during lab meeting and helped me run experiments. I would like to thank Prof. Bobby Bodenheimer and his graduate student Xianshi Xie, who provided excellent technical support. I need to thank our research assistants, Kitty and Thomas, for helping me run experiments. I also need to thank all the people who participated in my experiments for their time and cooperation.

Finally, no words can describe my gratitude to my parents for their unconditional love. I’m deeply indebted to my husband, Jubin Ma, who encouraged me to pursue education abroad and has been very supportive of my career.

TABLE OF CONTENTS

APPENDIX

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Imagine that you have buried a collection of valuables somewhere in your garden, a tax-free method to keep your treasure safe. Now you want to dig it out for sale in order to get through a financial crisis. It seems like an enjoyable treasure-hunting game, but you need to localize the exact location of your treasure before the digging so that you will not destroy the entire garden. How would you retrieve the treasure position? You can refer to various objects in the garden, such as a fence and plants. For example, you might have buried your valuables along the fence and 15 feet from the maple tree. You can also look at the mountain at distance, and try to fix your position with respect to it. If you buried the treasure on a dark and gusty night, you probably needed to remember how many steps to walk from a certain reference point, such as the fence opening; in this case you will need to use information generated from your own body movement. Your success in using those sources of information determines how successful you are in this treasure-hunting activity. One strategy is to rely on only one information source, which will probably be the most reliable cue in your estimation. Another strategy is to make use of multiple information sources and try to integrate them. Intuitively, the latter strategy is more advantageous. This is because every information source is not perfect and can be contaminated by noise. Usually there is a chance to decrease the errors by collecting different inputs. A more precise location representation is beneficial, which means successfully retrieving the treasure with a minimum damage to the garden in the current scenario.

This scenario represents a typical spatial localization problem, retrieving a target location that one has visited previously. Spatial localization tasks are ubiquitous in daily situations, such as retrieving one’s car in a bewildering parking building and navigating to the refrigerator in the kitchen from one’s bedroom in the early morning. Very often, multiple spatial cues are available for use. Basically, there are two critical problems to consider: what spatial cues are there and how to deal with multiple-cue scenarios. The following introduction will be organized in terms of these two questions. First, I will describe classifications of spatial cues. Next, I will introduce the Bayesian approach to cue interaction problem and explain why I used this approach in the current study. Finally, I will propose specific research questions within the framework of Bayesian theory.

Spatial Cues

Spatial cues can be divided into two categories. Internal self-motion cues (idiothetic cues) refer to bodily information generated by self-movement, such as vestibular cues and proprioceptive cues. Navigation using idiothetic cues alone is called dead-reckoning or path integration. Going back to the treasure-hunting scenario, remembering how many steps and the direction to walk from the fence opening utilizes self-motion cues. Because idiothetic cues are not directly informative about self-location, an integration operation needs to be performed continuously over time and distance. Studies have shown that rats are normally quite good at path integration. For example, after a rather random outbound journey, the rat was able to return directly to the path origin, as if it were continuously keeping in mind and updating a spatial vector pointing directly back to the path origin from its current position (Etienne & Jeffery, 2004). Humans, however, are relatively poorer at path integration than are rats (Loomis et al., 1993; Loomis, Klatzky, Golledge, & Philbeck, 1999).

External environmental cues (allothetic cues) refer to inputs from the outside world, such as visual and auditory cues. Visual cues have been studied extensively, possibly because humans rely heavily on the visual sensory system and visual cues are easy to manipulate in experiments. A classic categorization of visual cues is geometric cues vs. featural cues (Cheng & Newcombe, 2005; A. R. Lew, 2011; Twyman & Newcombe, 2010). Geometric cues refer to 3D extended surfaces, such as room shape and walls. Featural cues refer to spatially compact objects, such as landmarks and beacons, and aspects of objects (e.g., color). If we project the whole environment along the direction of gravity to the plane, landmark cues can be sketched as points whereas geometric cues are objects whose projections can be sketched as lines or shapes formed of lines. Geometric shape can also be conceptual, such as the configuration formed by discrete landmarks. Non-spatial properties of geometric cues are usually considered featural cues, such as the color and the texture of a wall. Allothetic cues can also be categorized as distal cues and proximal cues, based on their relative distances to a target location (Save & Poucet, 2000). Going back to the treasure-hunting game scenario, the fence can be classified as a geometric cue, the maple tree as a featural cue, and the far away mountain as a distal cue.

Bayesian Approach to Cue Interaction Problem

In the treasure-hunting scenario, you localize the buried collection of valuables by referring to the fence, the maple tree, the mountain, or your own physical movement. In order to achieve better performance, one heuristic is to combine various sources of spatial information instead of simply relying on one source. A second heuristic is to pay more attention to spatial cues considered as more reliable, so that limited cognitive resources are distributed efficiently. Bayesian principles on cue integration capture both of these heuristics.

Bayesian theory provides a systematic and quantitative method to investigate how different cues could be combined (Cheng, Shettleworth, Huttenlocher, & Rieser, 2007). Bayesian theory posits that the cue combination process is modulated by relative cue reliabilities. Cue reliability is usually inferred from subjects' responses and is inversely related to response variance, $r = 1/\sigma^2$. It is a measurement of how precise the location representation is associated with the cue. Cues are combined linearly with their weights equal to relative reliabilities. For example, for cue A, $r_A = 1/\sigma_A^2$, for cue B, $r_B = 1/\sigma_B^2$. Based on Bayesian principles, the weight for cue 1 is, $w_A = r_A / (r_A + r_B) = \sigma_B^2 / (\sigma_A^2 + \sigma_B^2)$, and the weight for cue B is, $w_B = r_B / (r_A + r_B) = \sigma_A^2 / (\sigma_A^2 + \sigma_B^2)$. The combined estimate of the underlying parameter is, $\mu_c = w_A * \mu_A + w_B * \mu_B$, where μ_A and μ_B are the single-cue estimates of the μ parameter for cue A alone and cue B alone respectively. Assuming that the single-cue estimates are unbiased and uncorrelated, when the two cues are congruent with each other, the combined estimate has a mean coincident with single-cue estimates. When they are set in conflict, the combined mean is a compromise between the two single-cue estimates and its distances to the two single-cue estimates are determined by the relative cue reliability. The combined estimate will be closer to the single-cue estimate of the more reliable cue. Thus, how the weight is distributed between A and B can be measured in terms of the relative distances between single-cue response distributions and combined-cue response distribution. This measure of relative distance should agree with cue relative reliability. Response variance of the combined estimate is calculated as, $\sigma_c^2 = \sigma_A^2 \sigma_B^2 / (\sigma_A^2 + \sigma_B^2)$. According to this equation, response variability should be reduced in double-cue conditions relative to single-cue conditions. Such a cue integration process is said to be optimal in the sense that the combined parameter estimate will have statistically minimum variability, or maximum precision.

Sometimes, instead of combining cues, navigators alternate between cues. Cue alternation refers to the use of only one cue for a certain proportion of trials and use of the other cue exclusively for the remaining trials. If cues are still weighted in a Bayesian manner, the alternation ratio should be equivalent to cue relative reliability. This means that the more reliable cue is used more often. However, response variance will not be reduced due to the lack of cue combination. Therefore, cue weights cannot distinguish Bayesian cue integration and Bayesian cue alternation. Whether or not there is response variability reduction can distinguish these two models apart.

Figure 1 illustrates Bayesian cue integration and Bayesian cue alternation. The two black curves represent response distributions associated with cue A and cue B respectively. In this example, cue A is more reliable than cue B -- its distribution is spatially sharper and corresponds to a more precise representation. The red curve represents the linear combination of cue A and cue B, with weights equal to cue relative reliability. The first thing to notice is that the combination curve is sharper than either of the two single-cue distributions, demonstrating the benefits of combining cues. Such benefit is at its maximum when the cue weights are equal to cue relative reliability. The second thing to notice is that the red curve is closer to the cue A distribution, meaning the more reliable cue gains greater weight. The blue curve represents Bayesian cue alternation. One important observation is that it has greater variance than the red combination curve. In addition there is no benefit of reducing variance relative to single cues as in the cue combination case. When the single cue distributions have the same centroids, the alternation distribution variance will lie somewhere between the single-cue variances. The variance of the alternation distribution will increase as the disparity between single-cue means increases. Another important observation is that the alternation distribution shares the same centroid as the combination distribution, illustrating the point that cue weights by themselves cannot distinguish these two multiple-cue strategies.

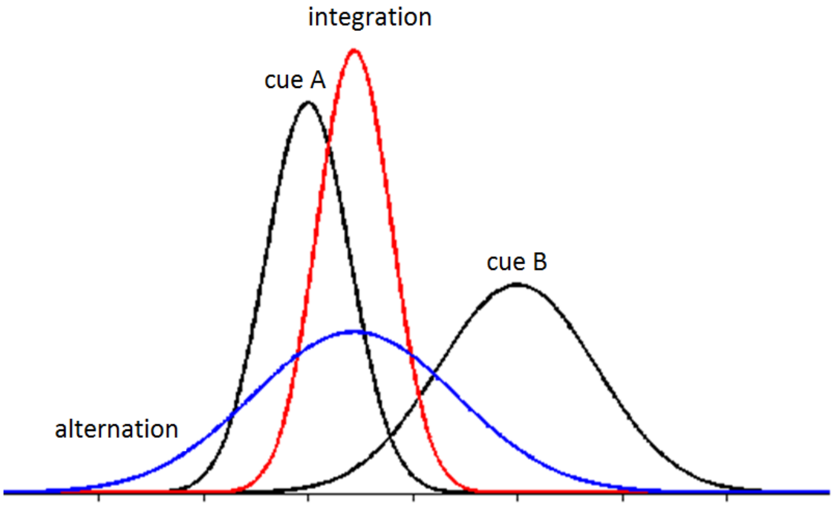


Figure 1. Bayesian framework of cue interaction. The two black curves represent single-cue distributions. The red curve represents the distribution of optimal cue integration. The blue curve represents the distribution of Bayesian cue alternation. The horizontal axis represents the continuous dimension of locations.

Under the banner of Bayesian perception, researchers in various study domains have demonstrated that people or animals are able to combine two different information sources optimally in a Bayesian manner (Butler, Smith, Campos, & Bülthoff, 2010; Ernst & Banks, 2002; Ernst & Bulthoff, 2004; Fetsch, Turner, DeAngelis, & Angelaki, 2009; Helbig & Ernst, 2007; Parise, Spence, & Ernst, 2012). Even when navigators do not combine cues, they appear to alternate between cues in a Bayesian manner (Nardini, Jones, Bedford, & Braddick, 2008). Such results imply that people or animals might be natural Bayesian observers and that the brain might be inherently organized with Bayesian principles (Fetsch, Pouget, DeAngelis, & Angelaki, 2012; Gu, Angelaki, & DeAngelis, 2008; Ma, Beck, Latham, & Pouget, 2006). However, the application of Bayesian principles to human spatial navigation is quite limited (Nardini et al., 2008; Zhao & Warren, 2010, November). The basic question still remains whether navigators performing a spatial task are able to integrate different sources of spatial information optimally, not to mention some further questions, such as how other cognitive processes might influence the implementation of Bayesian rules (Ernst, Banks, & Bulthoff, 2000; Talsma, Doty, & Woldorff, 2007). In addition, we speculate that the application of Bayesian theory to the domain of spatial navigation can be very helpful in reconciling some long-lasting debates in the spatial cognition literature and also provides new perspectives to navigational issues.

In order to reveal some critical advantages of the Bayesian approach, we will compare it to some other commonly used paradigms that have been employed in the spatial cognition domain. A frequently used paradigm is cue competition, to which the majority of the following section will be devoted. I will also discuss a few cue combination studies that have been conducted in the spatial navigation literature. These studies address the cue integration problem but lack the key features of the Bayesian integration paradigm.

Cue Competition and Integration Studies

Many studies have been conducted to investigate the interaction between different types of spatial cues. One extensively used paradigm is the cue competition or interference paradigm. There are two major paradigms, blocking designs and overshadowing designs. In a blocking design (Kamin, 1967), cue A is learned first. Next, both cue A and cue B are displayed. Finally, subjects are tested on their ability to use cue B alone to complete the task. Cue A is said to have blocked cue B if subjects' performance is impaired in the cue-B-only stage. In an overshadowing design (Pavlov, 1927), first, both cue A and cue B are presented. Then only cue B is presented. If performance is impaired in the cue-B-only stage, then cue A is said to have overshadowed cue B in the previous stage.

Cue competition in spatial memory and navigation tasks has been primarily investigated using visual cues. First, generally, there is interference between individual landmarks. For example, landmarks closer to the target were learned more than those farther away from the target (Chamizo, Rodrigo, Peris, & Grau, 2006; Goodyear & Kamil, 2004; Roberts & Pearce, 1999; Spetch, 1995). Second, interference between cues also happens to geometric cues vs. featural cues. Even though the relative competition capacity of these two cue types depends on various factors (A. R. Lew, 2011; Twyman & Newcombe, 2010), when cue interference was studied mutually between them, usually an asymmetric effect of blocking or overshadowing was found: the blocking/overshadowing effect of geometric cues on featural cues was greater than the blocking/overshadowing effect of featural cues on geometric cues (Wilson & Alexander, 2008). Sometimes the asymmetry could be so profound that there was substantial blocking/overshadowing effect of geometric cues on featural cues but no blocking/overshadowing effect of featural cues on geometric cues (Cheng, 1986; Doeller & Burgess, 2008). These results indicate that there is at least a general preference for geometric cues over featural cues, if not to the extent of an isolated geometric module in the brain (Cheng, 1986).

By comparison, relatively fewer studies have examined interference between visual cues and self-motion cues, and studies that have were conducted on animals. Whereas studies contrasting different visual cues indicate that interference usually occurs, studies on competition between path integration cues and visual cues suggest that these two types of spatial cues seem to be more independent of each other and that their interaction seems to be more complicated. Such discrepancy makes sense if we take into account that different visual cues come from a single sensory modality, whereas path integration cues and visual cues belong to different sensory modalities. For example, one recent animal study suggested that path integration might function as a back-up system and also as a reference system (Shettleworth & Sutton, 2005): when displaced landmarks were present, rats followed the displaced landmarks and ignored path integration; when landmarks were absent, rats could still retrieve the food correctly using path integration alone, suggesting path integration function as a back-up system. If landmarks were displaced a large amount, rats used path integration to detect the conflict and relied solely on path integration to retrieve the food location, suggesting that path integration functioned as a reference system.

In comparison to the literature on cue competition, fewer studies directly ask the question of how different cues might be integrated and what principles might govern the integration. This question is natural to ask since it is quite possible that multiple spatial cues are employed for spatial localization. In the case of path integration, several studies have shown that during locomotion, both path integration cues and external visual cues are used. In the study by Nico and his colleagues (Nico, Israel, & Berthoz, 2002), blindfolded subjects were transported passively in a robot chair and completed a triangle-completion task. Results showed that allowing subjects to see the world during the response contributed to a tendency of distance overestimation back to the path origin, compared to when they responded blindfolded. In studies by Kelly et al., geometric cues (different room shapes) and featural cues influenced path integration performance (Kelly, McNamara, Bodenheimer, Carr, & Rieser, 2008, 2009). And Kearn's study showed that allowing subjects to walk actively in the environment reduced response variability compared to a passive moving condition, indicating that the integration of self-motion cues from voluntary movement with visual cues resulted in more precise self-localization (Kearns, Warren, Duchon, & Tarr, 2002).

Even though the cue competition paradigm has been much more extensively used than the cue integration paradigm, these two paradigms might not be essentially different from each other. For one thing, cue integration might have occurred in cue competition studies. In cue-competition studies, usually performance was not completely reduced to a random level when only the blocked/overshadowed cue was displayed. In another words, during the previous double-cues stage, the blocked/overshadowed cue and the blocking/overshadowing cue might have been both utilized for location representation and combined in some way. When two cues were in small conflict and a continuous response was allowed, animals usually searched for a compromise between the different dictates of the cues, suggesting that both cues were used and integrated (Chittka & Geiger, 1995). Even complete blocking/overshadowing of cue A over cue B and no blocking/overshadowing of cue B over cue A can be interpreted as extremes in the cue-integration continuum; that is, one cue is assigned a weight of 0 while the other is assigned a weight of 1. For another thing, cue competition and integration processes might be influenced by the same set of factors. The ability of one cue to block or overshadow another cue probably depends on its intrinsic attributes (e.g., reliability, salience, stability) and prior experiences; in a similar way, the weight assigned to one cue relative to another cue in the two paradigms probably depends on the same set of factors. Compared to the overshadowing design or integration design, in the blocking design, the cue's competition capacity or its assigned weight is probably strengthened by prior experiences in addition to any of its intrinsic attributes.

However, the cue competition studies did not really answer the question of what determines a cue's competition ability against other cues. Researchers have examined various factors, such as cue type, cue distance, and navigation history. But they have not summarized the underlying determinants. Similarly, the cue integration studies did not address the determinants of how much consideration is given to different spatial cues, in addition to the fact that combining cues usually yielded better performance than using a single cue alone; and when observing that changing or adding one cue type influenced the performance using another cue, it is still unclear that what determined the extent of the influence. As discussed in the previous paragraph, cue competition paradigms and cue integration paradigms might share some fundamental features. A cue's competition ability as measured in cue competition tasks and a cue's weight as measured in cue integration tasks probably derive from a common concept. The Bayesian framework clearly proposes that cue reliability is one crucial determinant, meaning the higher the cue is in reliability, the greater the cue's ability to influence navigation decisions.

In addition, previous cue competition and integration studies do not pay attention to individual differences. Actually, individual differences have been frequently observed in the spatial navigation literature. For example, both animals and humans often show different preferences in using different types of visual cues in guiding their navigation (Cheng & Spetch, 1995; Kelly et al., 2009; Spetch & Mondloch, 1993). It is also well documented that people differ greatly in path integration abilities (Loomis et al., 1999). Even single cells within a spatially sensitive brain area responded differently to the same experimental manipulations (Knierim, 2002; Neunuebel, Yoganarasimha, Rao, & Knierim, 2013). However, the cue competition paradigm cannot offer an appropriate method to incorporate individual differences into the interpretation of the roles of different cues in spatial navigation, except acknowledging the existence of the phenomenon. This is largely because those studies usually limited their focus to objective properties of external visual cues but ignored navigators' subjective interpretation of those cues. On the contrary, within the Bayesian framework, cue reliability might reflect both cues' objective properties and navigators' subjective interpretations. Suppose one particular spatial cue corresponds to an objective level of cue reliability, subjective interpretation would scale the objective cue reliability up and down. If each individual sets his or her own cue combination weights, we would expect to observe a correlation between cue relative reliabilities and actual assigned weights to cues. That is, if Bayesian principles are applied at the individual level, we will be able to incorporate individual variances.

Current Project

The Bayesian approach has been employed extensively and has been proven useful in many fields of investigation, such as perception of shape from visual and haptic cues (Ernst & Banks, 2002), perception of slant from different visual cues (Nardini, Bedford, & Mareschal, 2010), and perception of location from visual and auditory cues (Parise et al., 2012). However, its application in spatial navigation is quite rare (Nardini et al., 2008; Zhao & Warren, 2010, November). This was our most important reason for conducting this project. We aimed to investigate human spatial navigation in the framework of Bayesian theory. There were two major steps. First, since very few studies have employed the Bayesian approach, we attempted to replicate some critical results of past studies (Nardini et al., 2008; Zhao & Warren, 2010, November). We also wanted to extend the experimental design adopted by past studies. Second, we planned to gain a better understanding of the Bayesian framework and expand its scope of application by asking new questions that have never been addressed in the spatial navigation literature. Completion of this project would not only answer the question of whether people are Bayesian navigators but also provide novel insights into important navigational problems in general.

Experiment 1A and Experiment 1B were designed to replicate and extend pioneering work in the field of human spatial navigation (Nardini et al., 2008; Zhao & Warren, 2010, November). We focused on the interaction between self-motion cues and visual cues. We adopted and modified the task developed by Nardini et al. (Nardini et al., 2008). Since self-motion cues were involved, we used immersive virtual reality technologies. One key question addressed by these experiments is whether people integrate visual cues and self-motion cues optimally in an immersive virtual environment. A novel aspect of our experiments is that we manipulated the reliability of visual cues and self-motion cues independently and within-subjects. This manipulation allowed us to assess whether human navigators could flexibly adjust their strategy in different experimental settings. To be a true Bayesian cue integrator, the navigator should be able to conditionalize the strategy to the current environmental parameters. For example, in the treasure-hunting scenario, you would probably rely more on external visual objects during daytime but would switch to your body-based sense of orientation and position if you dig at night; if you are near-sighted, you might want to use the distant mountain when wearing glasses but would abandon this cue if you have lost your glasses.

Experiments 2 to 4 were designed to answer new questions proposed within the Bayesian framework. Experiment 2 was devoted to advancing our understanding of cue reliability. Since cue reliability is at the core of Bayesian theory, one important question is whether it is necessary for people to acquire explicit knowledge about cue reliability and how they can be influenced by such knowledge. To answer this question, we provided distorted feedback to participants. The feedback was distorted in the sense that it never reflected the true level of performance. The purpose of the distorted feedback was to mislead people's beliefs about cue reliability in a top-down manner. We hypothesized that distorted feedback would affect the cue weighting process, in addition to the influence of cue reliability. In daily situations, we often receive feedback on our performance and we also evaluate the performance according to the feedback. For example, the treasure hunter might determine that the distant mountain was an uninformative cue to the buried position after digging in vain when using the mountain alone. Thus feedback might be an important source of information to help people develop accurate senses of a cue's usefulness, which is crucial for the implementation of Bayesian principles. Normally, feedback is directly associated with performance. In this experiment, by distorting the feedback, we could dissociate the two, which allows us to better understand the effect of feedback on the cue-weighting behaviors. As in Experiments 1AB, self-motion cues and visual cues were studied in an immersive virtual environment.

Experiment 3 expanded the application of Bayesian theory by incorporating another factor into the paradigm in addition to cue reliability. Based on the Bayesian framework of cue combination, cue reliability is the only determinant of how weights are assigned to different cues. Hence, a natural question is whether there are other factors besides cue reliability that would also affect the weights. We hypothesize that a given factor might have its effect via two different pathways (Figure 9). One is a direct pathway, wherein the factor affects the weights without affecting cue reliability; that is, cue reliability is short-circuited. The other is an indirect pathway, wherein the factor affects the weights by affecting cue reliability; in this case, cue reliability serves as a mediator. As an initial investigation, we studied cue instability, a factor that has been shown to affect navigational behaviors substantially (Biegler & Morris, 1993, 1996; Burgess, Spiers, & Paleologou, 2004; Knierim, Kudrimoti, & McNaughton, 1995; Lenck-Santini, Muller, Save, & Poucet, 2002). The Bayesian framework is suitable to distinguish these two pathways, because cue reliability can be measured and accounted for. The distinction between the two pathways with regard to cue instability will be discussed in greater depth in the presentation of Experiment 3. As in Experiments 1AB and 2, self-motion cues and visual cues were studied in an immersive virtual environment.

Experiments 1A, 1B, 2 and 3 examined the relationship between self-motion cues and visual cues. There are, of course, subcategories within each of these categories of cues. For example, visual cues can be classified as proximal vs. distal cues, geometric vs. featural cues, etc. Visual spatial cues have been examined intensively in spatial navigation literature. Therefore, it will be quite informative if we can apply the Bayesian cue combination paradigm to pairs of visual cues. Self-motion cues and visual cues belong to different sensory modalities, whereas different visual cues come from the same sensory modality. It is possible that navigators follow different rules when dealing with inter-modality vs. within-modality cue interaction. In Experiment 4, we modified the paradigm used in immersive virtual environments, so that we could examine the interaction between different visual cues using a 2D desktop virtual display. Desktop virtual reality is advantageous in that it requires less effort to collect data and thus large numbers of trials can be conducted. In addition, navigators are stationary, which allows brain imaging data to be collected. We contrasted proximal and distal landmarks as our initial foray into this domain.

CHAPTER II

EXPERIMENT 1A: MANIPULATING VISUAL RELIABILITY

In Experiment 1A, to vary reliability of visual cues, we pitted an environment rich in landmarks against an environment poor in landmarks. In Experiment 1B, to vary reliability of self-motion cues, we added different levels of body rotation into the task (Loomis et al., 1993). Participants experienced all of the noise levels of the manipulated cue. These two experiments represent the first time that the reliability of spatial cues has been manipulated using a cue integration paradigm, allowing us to answer the question of whether participants can conditionalize their strategy of dealing with multiple spatial cues across cue noise levels. Although cue reliability has been varied within-subjects (Ernst & Banks, 2002; Helbig & Ernst, 2007; Nardini et al., 2010), previous studies usually manipulated only one cue type and have not manipulated two cues independently; this allows us to compare how differently people sense cue reliability changes in different sensory modalities.

Method

Participants

Eighteen participants (10 males, 8 females) from the Nashville community participated in this experiment in exchange for course credit or monetary compensation. Participants ranged in age from 18 to 26 years, with a mean of 20.4. Participants had normal or corrected-to-normal vision.

Material

Virtual environments were displayed on an nVisor SX60 head-mounted display (HMD; NVIS, Reston, VA) with a 60º diagonal field of view. Stereoscopic images were presented at 1,280 × 1,024 pixel resolution, refreshed at 60 Hz. Graphics were rendered using Vizard software (WorldViz, Santa Barbara, CA) on a 3.0-GHz Pentium 4 processor with a GeForce 6800 GS graphics card. Head orientation was tracked by a 3-degrees-of-freedom orientation sensor (InertiaCube2; Intersense, Bedford, MA), and head position by a passive optical tracking system

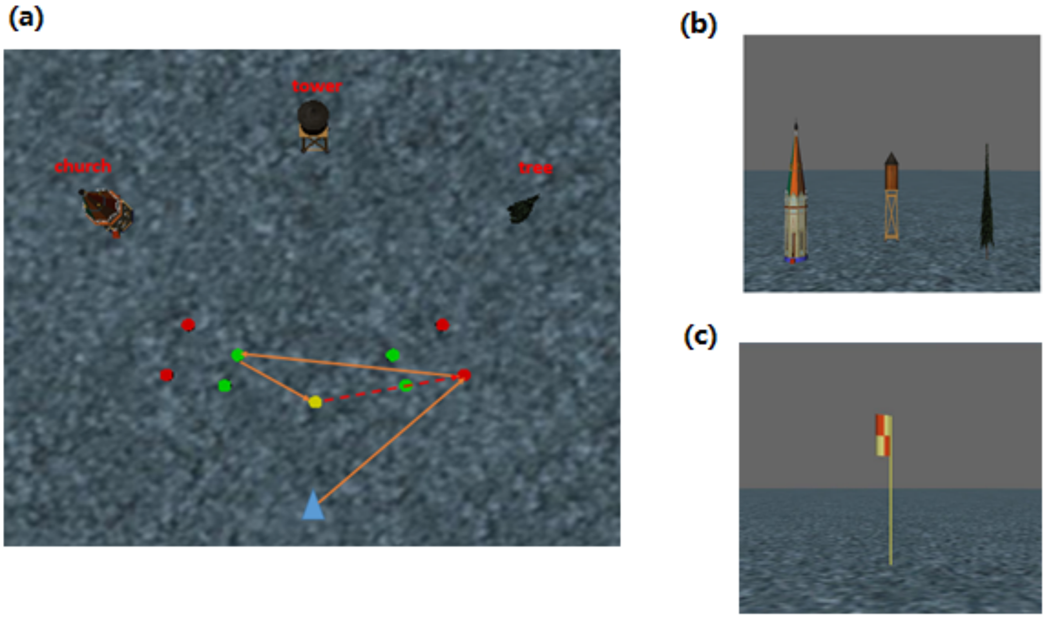


Figure 2. Experimental environments. (a) Bird's eye view of the virtual rich environment. Red dots represent locations of the first post; green dots represent locations of the second post; yellow dot represents location of the third post. Blue triangle represents the starting point of each trial. Orange arrows represent a typical outbound path and dashed red arrow represents the correct return path. (b) & (c) Typical views in rich and poor environments, respectively.

(PPTX4; WorldViz, Santa Barbara, CA). Drifting of the inertial sensor was corrected through optical tracking. Graphics were updated on the basis of sensed head movement, such that physical translations and rotations caused concomitant visual movement through the virtual world.

Each participant experienced both the rich environment and the poor environment (Figure 2b & c). The rich environment had three distinct landmarks (church, tower, & tree) positioned at different locations on a randomly textured ground plane. As shown in Figure 2a, on each trial, participants started by standing at a constant starting location, facing a constant direction (blue triangle). Next, they walked through a 3-post path, remembering the location of the first post. The posts appeared and disappeared in succession. At the end of the path, participants attempted to walk back to the location of the first post. There were four possible locations for the first post (Figure 2a, red dots) and four for the second post (green dots). The third post's location remained constant across the experiment (yellow dot). The distance from the two side landmarks to the third post was 3.4m and the distance from the middle landmark to the third post was 4.5m. The distances between the first posts and the third post were 2.15m and the distances between the second posts and the third post were 1.31m. The landmark configuration and post arrangement were symmetrical along the facing direction of the starting point. In the experiment, the location of the first red post was jittered by randomly sampling the jitter's x and y coordinates from a Gaussian distribution with mean = 0 and standard deviation = .25m. Then, an angle θ was randomly sampled from a uniform distribution between 0° and 360°. The final coordinate values of the jitter were calculated as, $x' = x * \cos(\theta)$ and $y' = y * \sin(\theta)$. Both x' and y' were truncated to .4m. In the poor environment, there was only a flag serving as the landmark, which was placed at the same location as the middle landmark in the rich environment.

Procedure

The task was to walk through a path consisting of 3 posts and then to attempt to walk back to the first post's location. Two sources of positional information were manipulated, self-motion cues and visual cues (e.g., landmarks). Participants could use both cues on some trials but only one of the two on others. The conditions were distinguished by the events that occurred at the end of the path: in the vision condition, participants were disoriented once they reached the last post (yellow dot in Figure 2a), so that during the response stage, self-motion information was lost and only landmarks remained; in the self-motion condition, the visual world was rendered invisible when participants reached the last post, so that they could only use self-motion cues; in the combination condition, participants were not disoriented and the world remained visible, so both landmarks and self-motion cues were available; and in the conflict condition, the landmark configuration was rotated clockwise by 15° so that the correct location defined by landmarks was different from the one defined by self-motion cues. The two single-cue conditions were used to estimate the cue reliabilities of visual cues and self-motion cues separately. The two double-cue conditions allowed us to examine how the two cues interacted when in agreement and when in disagreement.

In the vision condition, participants were disoriented before they attempted to make responses. The disorientation procedure was conducted as follows. The participant was guided to sit on a swivel chair once they reached the third post. Then the experimenter spun the chair for 20s while the participant counted down from a number between 100-200 in steps of 3. To match the vision condition, participants performed backward counting while standing still in the other three conditions. There were 10 blocks of trials, and each block contained 8 trials. Half of the trials were in the rich environment, and half were in the poor environment. The 10 blocks were evenly divided into 2 sessions on 2 consecutive days. The two environments alternated by segments of 4 trials, and for each 4-trial segment, each of the 4 conditions was experienced once. Hence, for each block, the first 4 trials and the last 4 trials were experienced in two environments, respectively. Half of the participants experienced the rich environment first in each block, and the other half experienced the poor environment first.

Results

Data Analysis

The calculation of response variability and relative reliability proceeded as follows. As shown in Figure 2a, the four target locations were positioned symmetrically around the midline, with two on each side. First, for each of the four target locations, responses were transformed into a spatial coordinate system with the target location as the origin and the correct walking direction as y-axis. Second, responses for the two target locations on the same side were pooled together. Outliers were defined as responses whose distance from the response centroid exceeded the 3rd quartile by 3 times the interquartile range, considering all four target locations together. Third, before pooling responses from the two sides, we corrected the bias in each side's response distribution by centering each distribution over its centroid. This was done to eliminate disparity in the centroids of the distributions on the two sides, which would increase the variability of the pooled distribution artificially. Fourth, after pooling together the two sides' response distributions, we computed the response standard deviation as,

$$\sigma = \sqrt{\sum d^2 / (n-1)}$$

where d is the distance of the response from the pooled centroid and n is the number of responses. Relative reliability of visual cues to self-motion cues was calculated as the inverse of relative variability of the two single-cue conditions, $rr = \sigma^2_{motion}/(\sigma^2_{motion} + \sigma^2_{vision})$.

The calculation of relative proximity needs to take into account any intrinsic bias in the response distribution; therefore we computed the relative proximity for each side first, then took the mean as the final estimate. For each side, we pooled together the responses on the two target locations, and then calculated the distribution centroid, μ. Euclidean distances between single-cue distributions and double-cue distributions were computed (e.g., $d_{motion\text{-}comb} = | \mu_{motion} - \mu_{comb} |$). Relative proximity to the visually defined location was calculated as the inverse of the relative distance,

$$rp_{comb} = d_{motion\text{-}comb}/(d_{motion\text{-}comb} + d_{vision\text{-}comb}), \text{ and}$$

$$rp_{conf} = d_{motion\text{-}conf}/(d_{motion\text{-}conf} + d_{vision\text{-}conf}).$$

The relative proximities computed for each side were then averaged across sides. This index reflects the actual weight assigned to visual cues relative to self-motion cues. According to Bayesian theory, response relative proximity to the visually defined location should be equal to cue relative reliability of visual cues.

Predictions for variability of double-cue conditions were calculated from variability of single-cue conditions. We tested two models, which were distinguished based on whether there was cue integration or cue alternation. The Bayesian integration model predicts that participants would integrate cues optimally,

$$\sigma^2_{comb} = \sigma^2_{conf} = \sigma^2_{vision} * \sigma^2_{motion}/(\sigma^2_{vision} + \sigma^2_{motion}), \text{ and}$$

$$rp_{comb} = rp_{conf} = rr.$$

The Bayesian alternation model predicts that people would alternate between landmarks and self-motion cues with an alternation ratio equivalent to the optimal weight, such that

$$\sigma^2_{comb} = (1\text{-}rr)^* (\mu^2_{motion} + \sigma^2_{motion}) + rr * (\mu^2_{vision} + \sigma^2_{vision}) - ((1\text{-}rr) * \mu_{motion} + rr * \mu_{vision})^2.$$

The same equation applies to the conflict condition. It is worth noting that the Bayesian alternation model shares the same predictions with the Bayesian integration model in terms of response relative proximity. Response relative proximity would correspond to the actual weights assigned to visual cues in the linear combination in the Bayesian integration model, and would correspond to the percentage of trials in which only visual cues were used in the Bayesian alternation model. The two models differ in predicting response variability in double-cue conditions. The integration model predicts reduced variability relative to single-cue conditions, whereas the alternation model predicts no variability reduction.

Responses were analyzed both in individual environments and across environments. For the latter, responses were pooled across the two environments based on which of the four conditions the response was in. We defined individuals as outliers when the averaged response variability across all conditions and both environments was 3 interquartile ranges above the 3rd quartile or below the 1st quartile. In correlational analyses, bivariate outliers were defined as having unstandardized residuals greater than the 3rd quartile by 3 times the interquartile range in the linear regression analysis.

Response Variability

The manipulation check shows that we successfully manipulated the reliability of visual cues without changing the reliability of self-motion cues. As shown in Figure 3a, vision variability was significantly lower in the rich environment than in the poor environment ($t(17) = 5.68$, $p = .000$), whereas there was no difference in self-motion variability ($t(17) = .09$, $p = .926$).

The second question is whether there was variability reduction in double-cue conditions compared to single-cue conditions, a hallmark of cue integration (Figure 3a). We also compared double-cue conditions to the Bayesian integration model. For the second analysis, we calculated Bayes factors (BFs) in addition to traditional inferential tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009). The customary cut points for Bayes factors are 3 and 1/3. A Bayes factor greater than 3 means that the null hypothesis is more than 3 times as likely as the alternative hypothesis; in the present comparisons, this means that the model fits the data. Conversely, a Bayes factor smaller than 1/3 means the model fails to fit the data. Table 1 includes the statistics on the comparisons between double-cue conditions and models for all experiments.

As shown in Figure 3a, in the rich environment, the combination condition showed lower variability than vision and self-motion conditions, although only the combination-motion comparison was significant (t(17) = 1.37, p = .188; t(17) = 8.15, p = .000). The combination condition was consistent with the Bayesian integration model (t(17) = 0.51, p = .614, BF = 4.92). The conflict condition was also consistent with the Bayesian integration model (t(17) = 1.27, p = .220), although the Bayes factor was slightly short of the cut point (BF = 2.65).

In the poor environment, the combination condition showed lower variability than vision and self-motion conditions, but neither comparison was significant (vision, t(17) = 1.72, p = .103; self-motion, t(17) = 1.00, p = .332). The combination condition seemed to differ from the Bayesian integration model (t(17) = 2.26, p = .037, BF = 0.65). The conflict condition showed reduced variability relative to both single-cue conditions (vision, t(17) = 3.72, p = .002; self-motion, t(17) = 4.98, p = .000). The conflict condition was consistent with the Bayesian integration model (t(17) = 0.33, p = .748, BF = 5.31).

When responses were pooled across the rich and the poor environments, both double-cue conditions showed significant variability reduction relative to the two single-cue conditions (ps < .05). The t-test suggested that the combination condition did not differ from the integration model (t(17) = 1.48, p = .157), although the Bayes factor fell short of the cut point (BF =2.07). The conflict condition was consistent with the Bayesian integration model (t(17) = 0.34, p = .735, BF = 5.28).

We also compared all the double-cue conditions to the Bayesian alternation model, which does not predict variability reduction. All double-cue conditions were different from this model, except the combination condition in the poor environment, which seemed to be a

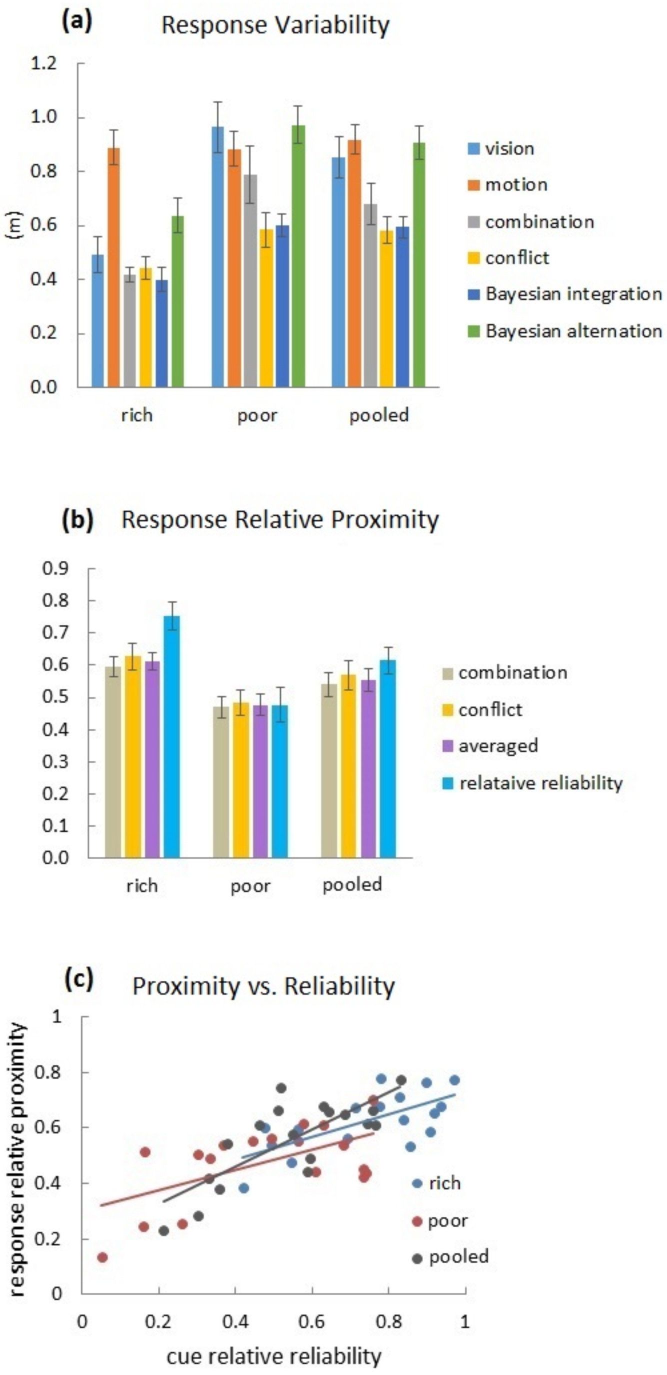


Figure 3. Results of Experiments 1A. (a) Response variability and model predictions. The alternation model's prediction represents the average of the predictions for the combination and the conflict conditions. (b) Response relative proximity and cue relative reliability; response relative proximity is displayed for the combination condition and the conflict condition respectively, and when the two conditions were averaged. (c) Correlations across participants between averaged response relative proximity and cue relative reliability. Error bars represent ± SE of the mean.

compromise between cue integration and alternation (combination vs. integration model, t(17) = 2.26, p = .037, BF = 0.65; combination vs. alternation model, t(17) = 1.94, p = .069, BF = 1.09).

Response Relative Proximity

We calculated response relative proximity in combination and conflict conditions separately. Because results showed that response relative proximity was very consistent between the two double-cue conditions, we took the mean of the two conditions to get a more reliable estimate of the actual weight assigned to the cues. We then compared the average response relative proximity to relative cue reliability calculated from single-cue conditions (Figure 3b). Response relative proximity measures the actual weight assigned to the visual cues relative to self-motion cues, and cue relative reliability represents the Bayesian weighting-strategy. The two indices should correspond with each other if the cues were weighted in a Bayesian manner.

In the rich environment, response relative proximity was significantly lower than cue relative reliability in both double-cue conditions and when the two conditions were averaged (t(17) = 3.55, p = .002). The response relative proximity and cue relative reliability were consistent with each other in the poor environment (t(17) = 0.00, p = .998) and when responses were pooled across environments (t(17) = 0.42, p = .678). As shown in Figure 3c and Table 2, positive correlations existed between response relative proximity and cue relative reliability, across participants, in both individual environments (bivariate outliers deleted) and in the pooled analysis. These results indicate that in general participants weighted cues based on their relative quality.

Summary and Discussion

Participants' performance tended to deviate from the Bayesian integration model in individual environments. First, variability reduction was not consistently observed in double-cue conditions relative to single-cue conditions, and response variability in double-cue conditions was not consistently fit by the Bayesian integration model. Second, response relative proximity was lower than cue relative reliability in the rich environment. When responses were pooled across environments, participants' behaviors were optimal or near-optimal, both in variability reduction and response relative proximity. Across all levels of analysis, positive correlations existed across subjects between relative proximity and relative reliability, implying participants weighted visual cues and self-motion cues based on their relative quality.

CHAPTER III

EXPERIMENT 1B: VARYING SELF-MOTION CUES

Experiment 1B was parallel to Experiment 1A. Here, the reliability of self-motion cues was varied within-subjects.

Method

Participants

Eighteen people (9 males, 9 females) from Vanderbilt University or the Nashville community participated in this experiment. They ranged in age from 18 to 32 years, with a mean of 23.7. Participants had normal or corrected-to-normal vision.

Material and Procedure

Participants performed the same task as in Experiment 1. Only the poor environment was used (Figure 2c). The reliability of self-motion cues was manipulated within-subjects by having participants either stand still or perform a body rotation while they were doing the backward counting. For body rotation, participants were instructed to rotate either to the left or to the right randomly until their facing direction was aligned with a designated marker. The body rotation was 270° on average, with jitter sampled from a Gaussian distribution with mean of 0° and standard deviation of 10°. Jitter was truncated to 10°. The layout of trials was very similar to Experiment 1, except there were only 7 types: Vision only, self-motion only with or without rotation, combination with or without rotation, conflict with or without rotation. There were 10 blocks of trials, each with 7 trials corresponding to the 7 trial types. The 10 blocks were evenly divided over two days. In each block, participants experienced no rotation trials first, then rotation trials, and the vision-only trial with disorientation was inserted into a random position within the block.

Results

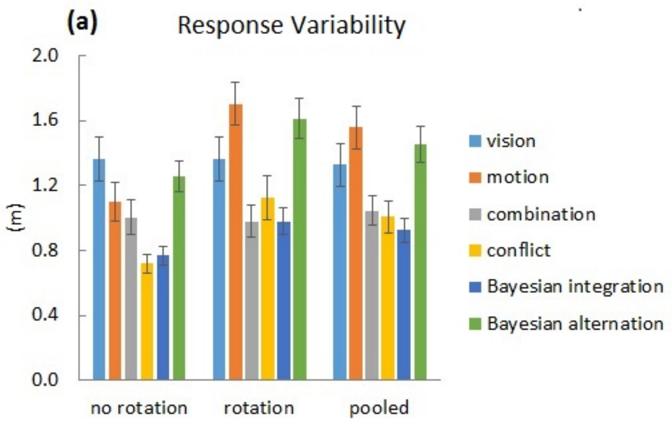
Response Variability

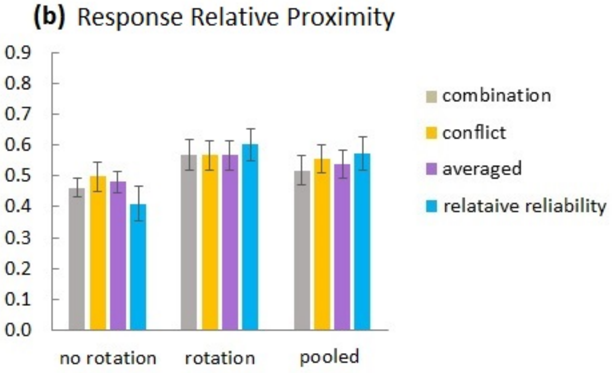
The manipulation check indicated that adding body rotation of 270° at the end of the outbound path successfully increased response variability (Figure 4a). Response variability in the self-motion condition was significantly greater when body rotation was added than when no body rotation was added ($t(17) = 5.88$, $p = .000$).

When there was no body rotation, the combination condition showed significantly smaller variability than the vision condition ($t(17) = 2.56$, $p = .020$) but not the self-motion condition ($t(17) = 0.99$, $p = .337$). The combination condition differed from the Bayesian integration model ($t(17) = 2.87$, $p = .011$, BF = .22). The conflict condition showed variability significantly lower than the vision condition ($t(17) = 4.65$, $p = .000$) and the self-motion condition ($t(17) = 2.99$, $p = .008$). The conflict condition was consistent with the integration model ($t(17) = 0.66$, $p = .515$, BF = 4.53).

When body rotation was introduced, the combination condition showed significantly lower variability than both single-cue conditions (vision, $t(17) = 3.20$, $p = .005$; self-motion, $t(17) = 5.70$, $p = .000$). The combination condition was consistent with the integration model ($t(17) = 0.01$, $p = .989$, BF = 5.59). The conflict condition showed significant variability reduction relative to the self-motion condition ($t(17) = 3.49$, $p = .002$) but not to the vision condition ($t(17) = 1.68$, $p = .111$). The conflict condition was consistent with the integration model, although the Bayes factor was short of the cut point ($t(17) = 1.39$, $p = .184$, BF = 2.32).

When responses were pooled across rotation levels, both double-cue conditions showed significantly lower variability than single-cue conditions ($ps < .04$). The combination condition was marginally greater than predicted by the Bayesian integration model and Bayes factor analysis suggested that the evidence was ambiguous on whether this model fit the data ($t(17) = 1.92$, $p = .072$, BF = 1.12); these results imply near-optimal mode of response. The





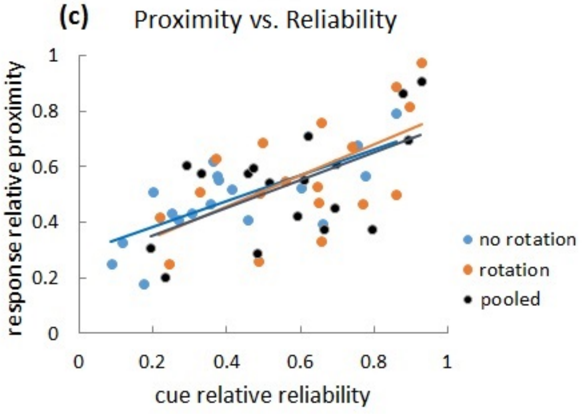


Figure 4. Results of Experiment 1B. (a) Response variability and model predictions. The alternation model's prediction represents the average of the predictions for the combination and the conflict conditions. (b) Response relative proximity and cue relative reliability; response relative proximity is displayed for the combination condition and the conflict condition respectively, and when the two conditions were averaged. (c) Correlations across participants between averaged response relative proximity and cue relative reliability. Error bars represent ± SE of the mean.

conflict condition was consistent with the Bayesian integration model (t(17) = 1.05, p = .309, BF = 3.35).

All of the double-cue conditions differed from the alternation model, except the combination condition in the no rotation condition, which seemed to be a compromise between cue integration and alternation (combination vs. integration model, t(17) = 2.87, p = .011, BF = .22; combination vs. alternation model, t(17) = 2.54, p = .021, BF = 0.40).

Response Relative Proximity

As shown in Figure 4b, when there was no rotation, cue relative reliability did not differ from response relative proximity in the combination condition (t(17) = 1.24, p =.233). Differences between relative reliability and relative proximity were significant in the conflict condition (t(17) = 2.08, p = .053) and marginally significant when the two double-cue conditions were averaged (t(17) = 1.84, p = .083). The two indices were consistent with each other when there was rotation and when responses were pooled across rotation levels (ps > .300). At all levels of analysis, positive correlations existed between relative proximity and relative reliability, as shown in Figure 4c and Table 2.

Summary and Discussion

The current experiment showed a pattern of results similar to that of Experiment 1A. In the current experiment, when responses were pooled across environments, participants' behaviors were optimal or near-optimal in Bayesian terms, with significant variability reduction in double-cue conditions consistent with Bayesian predictions and relative proximity consistent with relative reliability. At individual rotation levels, participants' behaviors tended to deviate from the Bayesian integration model. First, double-cue conditions did not consistently show variability reduction and were not consistently fit by the Bayesian integration model. Second, response relative proximity in the no rotation condition tended to be higher than the Bayesian weight (i.e., cue relative reliability).

It is worth noting that the Bayesian weight when there was no rotation was lower than the Bayesian weight when there was rotation (in Figure 4b, the no-rotation blue bar is lower than the rotation blue bar) and the actual weights assigned to visual cues in the no-rotation condition were higher than the Bayesian weights (e.g., purple bar vs. blue bar). This result suggests that the actual weights assigned to visual cues were increased in the no-rotation condition to be closer to the actual visual weights in the rotation condition. The same tendency was observed in Experiment 1 as well, where the actual weights assigned to visual cues were apparently decreased in the rich environment to be closer to the actual visual weights in the poor environment (in Figure 3b, rich condition blue bar is higher than poor condition blue bar, and purple bar is lower than blue bar in rich condition). When we analyzed this variable across the two experiments (cue relative reliability – response relative proximity in rich condition in Exp. 1A; response relative proximity – cue relative reliability in no rotation condition in Exp 1B), results showed that this bias was significant (p = .001) and no difference was found between experiments (p = .20). This finding indicates that when subjects experienced cues changing in reliability, the actual weights were distorted at the richer level of the manipulated cue in such a direction that they would be closer to the weights at the poorer level of the same cue.

The most obvious compromise in response variability between the integration model and the alternation model occurred in the poor environment in Experiment 1 and in the no-rotation condition in Experiment 2. It is worth noting that these two conditions shared exactly the same experimental settings: The flag was the sole landmark and no extra body rotation was added. A subsequent experiment (Experiment 2) showed that when subjects only experienced this condition without another reliability level, variability reduction was consistently observed in both double-cue conditions and this reduction was consistent with the Bayesian integration model. This finding implies that the compromise observed in the present experiments resulted from interference caused by people alternating between different reliability levels of spatial cues; it also indicates that the better fit of the Bayesian model to the pooled data was not caused solely by the doubling of the number of trials.

We also carefully compared these two conditions in other aspects. We found that response variability was significantly higher in the no-rotation condition than in the poor environment (p=.020, note that scales of ordinates differ in Figure 3a & Figure 4a). In addition, as stated previously, relative proximity and relative reliability were consistent with each other in the poor environment but they tended to be different in the no-rotation condition. Finally, in the no-rotation condition, additional analyses showed that response variability in the conflict condition was not correlated across subjects with the predictions of either the integration model or the alternation model, the only double-cue condition devoid of such relationships in both experiments. These discrepancies provide additional support to the claim that even though these two conditions shared the same experimental settings, subjects behaved differently, depending on the other reliability level they experienced simultaneously.

In both Experiments 1AB, the combination condition showed a greater interference effect than the conflict condition, in that at individual cue reliability levels, the combination condition sometimes exhibited a compromise between the cue integration model and the cue alternation model whereas the conflict condition was usually consistent with the cue integration model. This pattern of results is counterintuitive. In fact, participants might have treated the two double-cue conditions equivalently. The disparity we created between visual and self-motion cues was very small in the conflict condition (≈0.56 m), so participants probably were unaware of the conflict. On each individual trial, the estimate from visual information and the estimate from self-motion information were bound to differ, forming cue conflict even in the combination condition. However, it remains unclear to us and needs future investigation why the conflict condition did not suffer from as much interference as the combination condition. One possible reason is that the underlying single-cue distributions might have changed in the conflict condition. This hypothesis will challenge the validity of the Bayesian framework, which assumes that single-cue distributions measured in single-cue conditions remain unchanged in double-cue conditions. Another possible reason for greater interference in the combination condition is related to the relatively limited physical space in the lab. When visual cues were rotated, the visually defined target location might become closer to physical obstacles, such as walls and desks. This would limit the response space, which in turn could have reduced response variability.

Summary and Discussion of Experiments 1A and 1B

In both experiments, we observed similar patterns of results. When responses were pooled across noise levels of the manipulated cue, people behaved in a Bayesian optimal way. Response variability reduction was evident and generally consistent with Bayesian predictions. Relative proximity was also predicted by Bayesian principles. This pattern of results is consistent with findings on cue integration of other stimulus features (Butler et al., 2010; Ernst & Banks, 2002; Fetsch et al., 2009; Helbig & Ernst, 2007; Parise et al., 2012) and pioneering work in the domain of human spatial navigation (Nardini et al., 2008). However, people's behaviors tended to deviate from Bayesian principles at individual noise levels, in that relative proximity differed from cue relative reliability and response variability was not always reduced by the amount predicted by Bayesian theory. Interference between noise levels was also evident in the apparent tendency of participants to adjust their responses at one noise level away from Bayesian predictions to be closer to that at the other noise level. These results showed that people could be non-optimal under certain circumstances.

The interference effect between noise levels might have been caused by properties shared by them. For example, the same set of physical locations was used for both noise levels, the virtual visual environments are either very similar or the same, and the navigation space was always mapped to the same real-world environment, the lab space. It is possible that Bayesian coding principles were restricted to or targeted at the target locations or the surrounding spatial context. Future research might attempt to determine which factors shared by noise levels would lead to interference. This research might shed light on how Bayesian coding principles are organized and at which level they occur in the spatial navigation system.

It was surprising that people deviated from Bayesian integration theory at individual noise levels but managed to behave in a Bayesian way when responses were pooled. A mathematical analysis suggested that when both single-cue distributions are not biased, if people behave as Bayesian integrators at individual noise levels, they would become supra-optimal at the pooled level (Appendix 1). Further simulations showed that people can compromise at individual noise levels in order to be just Bayesian optimal at the pooled level (Appendix 1). It would be worth investigating in future simulation work the forms that the interference must take in order for behavior to be optimal at the pooled level. Future simulation work should also take into account intrinsic biases in response distributions and try to incorporate the interference effect we observed in response relative proximity.

We found positive correlations between response relative proximity and cue relative reliability in both experiments and in all double-cue conditions. This demonstrates that people weighted self-motion cues and visual cues based on their relative reliability. Since we also found similar results in the subsequent 3 experiments, we will discuss implications in the general discussion.

CHAPTER IV

EXPERIMENT 2: DISTORTED FEEDBACK

In Experiments 1A and 1B, we found that cue reliability was a critical factor determining how cues would be weighed. Cue reliability was positively correlated with cue weights. It seems that people actually evaluated how useful the cues were; they then weighted the cues accordingly. But is explicit evaluation necessary to successfully implement the weighting-by-reliability strategy? Do people need to calculate variances of sensory estimators associated with different spatial cues? This question has barely been asked or investigated in the cue integration literature. There has been one hypothesis proposed by Ernst and Banks (Ernst & Banks, 2002). They speculated that there is no need to explicitly extract cue reliability from the cue, since neural firing and inter-neuron transmission process can accomplish it unconsciously. This hypothesis makes sense if we consider that people usually can quickly acquire an impression of how good a cue is upon seeing it. However, this does not necessarily mean that the cue weighting process would not be affected by top-down processes.

It is plausible that top-down influences exist on the cue combination process. In the cue combination literature, some studies have shown that active attention can affect how observers combine different cues. For example, Mozolic et al. found that instructing participants to pay attention to only one sensory channel abolished the performance enhancements caused by multi-sensory display; performance enhancements in multi-sensory condition needed divided attention to both channels (Mozolic, Hugenschmidt, Peiffer, & Laurienti, 2008). Berger and Bulthoff demonstrated that when participants were aware of cue conflict, active attention to visual inputs against auditory inputs increased the weights assigned to visual cues (Berger & Bülthoff, 2009). With ERP techniques, Talsma et al. showed that divided attention to both visual and auditory aspects of an object was needed for early integration in the brain: the P50 component showed superadditive integration effect in the divided-attention condition but not in the unisensory-attention condition (Talsma et al., 2007).

The current experiment was designed to test this conjecture on top-down influences. We investigated the top-down influences of subjective evaluation of cue reliability. We speculate that one convenient way to evaluate cue quality is to judge by behavioral consequences. If people obtain a high level of performance by using a certain cue, the cue will be labeled as useful and gain substantial reliance accordingly. To test this hypothesis, we displayed feedback to participants on their performance on a trial-by-trial basis. The feedback was distorted in the sense that it never reflected the true performance level. We hoped that offering explicit feedback on performance would affect people's beliefs about cue reliability, and their reliance on the cue.

We hypothesized that distorted feedback would change the cue weights determined by cue reliability alone; that is, the actual weights would be different from Bayesian weights. We hypothesized that ultimately, cue weights would be determined by subjective experience of cue reliability, which would be affected by the feedback information. We displayed two different types of feedback. Positive feedback indicated to participants they were closer to the target location than they actually were, so that subjective cue reliability would be greater than cue reliability that we actually measured. If participants assigned weights to cues based on subjective cue reliability, they would rely on the cue associated with positive feedback more than predicted by the measured cue reliability. On the contrary, negative feedback indicated to participants that they were farther away from the target location than they actually were, so they would rely on the corresponding cue less than predicted by the measured cue reliability.

Figure 5 depicts our hypothesis. Suppose visual cues and self-motion cues are equivalent in cue reliability. When there is no feedback and participants weight cues according to the Bayesian rules, the Bayesian weight and the actual weight are consistent and lie in the middle between the two single-cue distributions (Figure 5a). If positive feedback is associated with visual cues, participants' subjective evaluation of the reliability of visual cues will be higher than the actual reliability. Participants then weight cues based on the subjective evaluation of cue usefulness and thus assign greater weight to visual cues than predicted by Bayesian principles. This hypothesis predicts disparity between the actual weight and the Bayesian weight, but the actual cue reliability will not be changed by the feedback (Figure 5b).

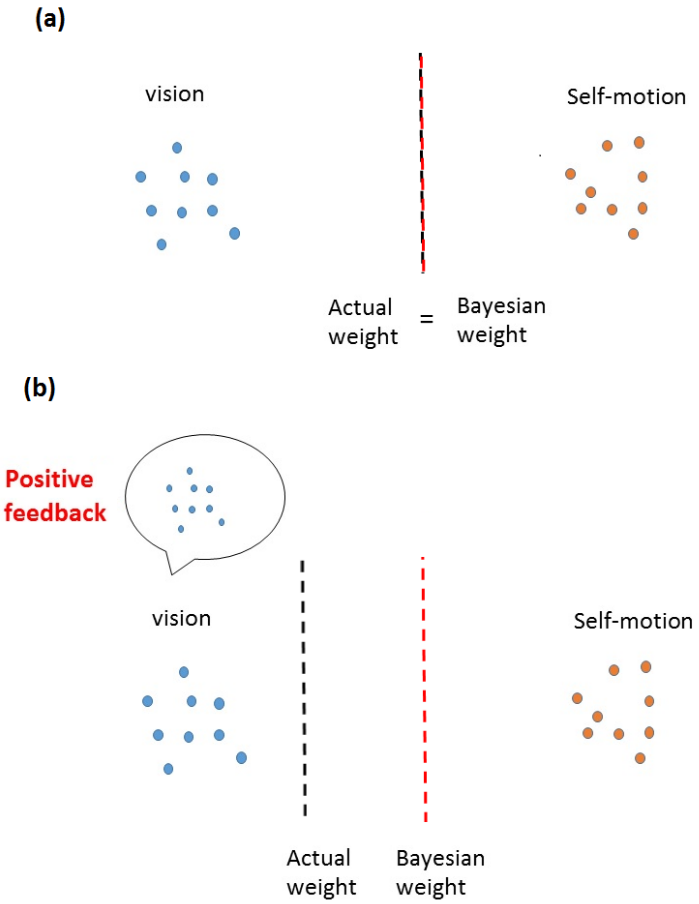


Figure 5. Illustration of predictions of Experiment 2. (a) Bayesian cue-weighting strategy without feedback. The dotted black and red lines represent the Bayesian weight and the actual weight, respectively. Visual and self-motion cues are equivalent in reliability, so the weights are in the middle between the two. (b) Cue-weighting strategy with positive feedback to visual cues. Because visual cues are considered as more reliable than they actually are (as in the cloud), the actual weight becomes closer to visual cues and also deviates from the Bayesian weight calculated from actual reliabilities. Cue reliability remains unchanged in the presence of feedback.

In the current experiment, the distorted feedback was given regarding the distance error, which is the absolute distance of the response location to the target location. The distance error actually contains two error components, variability and bias. Ideally, we wanted to mislead participants' beliefs on variability. With such a small number of trials, it is hard to estimate bias at the very beginning of the procedure or iteratively throughout the procedure and then isolate variability from that. However, considering that response bias is a constant error, the misleading effect of feedback on distance error should be transferred to a proportional misleading effect on response variability within the current paradigm.

Method

Participants

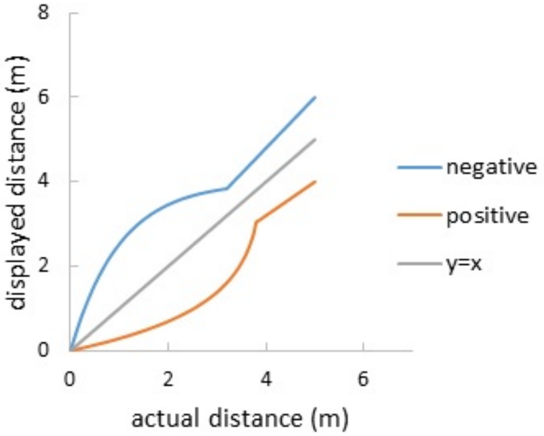
Thirty-six people (18 male) from Vanderbilt University or Nashville community participated in this experiment. They all had normal or corrected-to-normal vision. They ranged in age from 18 to 29 years old, with a mean of 20.38.

Material and Procedure

The task and procedure were very similar to the poor condition in Experiment 1A and the no-rotation condition in Experiment 1B -- the environment contained a single landmark and no body rotation was introduced at the end of the outbound path. The only difference was that in the current experiment, we provided feedback to participants after they had attempted to walk back to the target location. The feedback was used to tell participants how far way their response location was from the correct target location. The feedback information was distorted for the purpose of misleading participants' beliefs about the relative quality of visual cues and self-motion cues. There were two different types of feedback, positive and negative. Positive feedback told participants that they were closer to the target location than they actually were. Negative feedback gave the opposite information, and thus participants would consider the spatial cue to be less reliable than its true level. Figure 6 shows the functions of the positive feedback and negative feedback. The positive feedback function is always below the diagonal line, meaning that the distance error displayed to participants is always smaller than the actual distance error, whereas the negative feedback function is always above the diagonal, meaning that the distance error displayed to participants is larger than the actual distance error.

Participants were randomly divided into two groups, vision positive group (9 male and 9 female) and self-motion positive group (10 male and 8 female). Gender was counterbalanced between groups. In the vision positive group, participants received positive feedback in the vision only condition and negative feedback in the self-motion only condition. In the self-motion positive group, participants received negative feedback in the vision only condition and positive feedback in the self-motion only condition. No feedback was provided in double-cue conditions. We hoped that the vision positive group would over-appreciate visual cues and under-appreciate self-motion cues. The opposite pattern was expected in the self-motion positive group.

There were 40 trials in total, 10 trials in each of the 4 conditions. The trials were divided into 5 blocks in a one-day session. In each block, the 4 single-cue trials were experienced ahead of the 4 double-cue trials. Conditions were randomized within the single-cue part and the double-cue part. We hoped that the distortion effect accumulated in the first 4 trials would be strong enough to influence participants' behaviors in the subsequent trials. The feedback manipulation was implemented in each block, ensuring that the distortion effect, if there was any, would remain at a relatively constant level across the whole experiment.



<u>Figure 6.</u> Feedback distortion functions used in Experiment 2. Displayed distance error is plotted against actual distance error. Gray line represents authentic feedback, y=x. The blue line represents negative feedback,

$$\begin{cases} y = 1.2x & , x > 3.2 \\ y = 4 \times (1 - e^{-x}) & , x \leq 3.2 \end{cases}$$

The red line represents positive feedback,

$$\begin{cases} y = 0.8x & , x > 3.8 \\ y = -\ln(\frac{1-x}{4}) & , x \leq 3.8 \end{cases}$$

Results

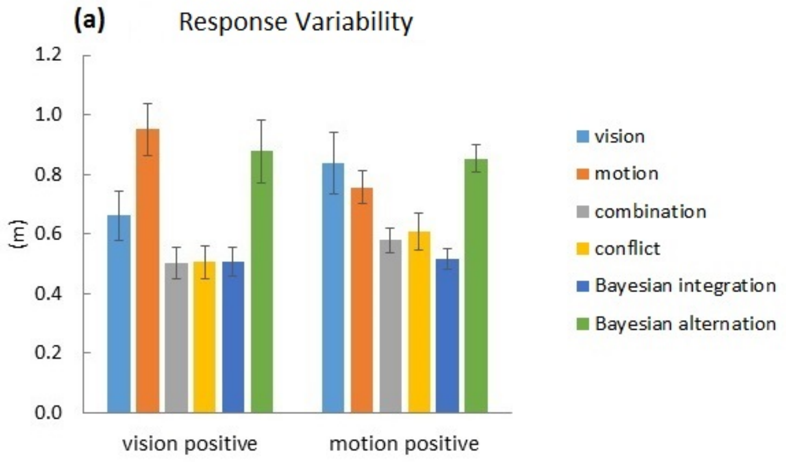
Data Analysis

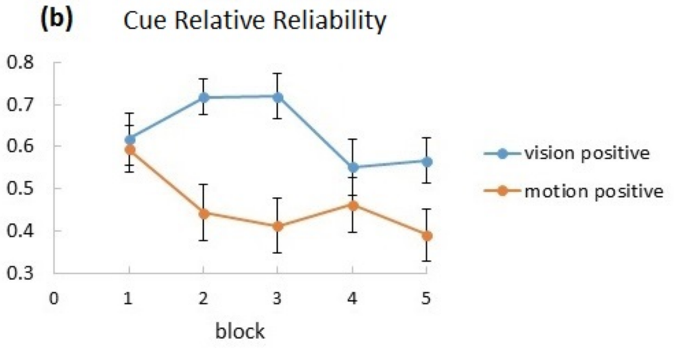
Data analysis followed the same procedure as in Experiments 1A and 1B, except that the current experiment only contained a single level of cue reliability. Hence, there was no pooled analysis across noise levels.

Response Variability

Overall, both groups showed variability reduction (Figure 7a). In the vision positive group, the combination condition was marginally smaller than the vision condition (t(17) = 2.052, p = .056) and significantly smaller than the self-motion condition (t(17) = 6.695, p= .000). The conflict condition was significantly smaller than both single-cue conditions (vision, t(17) = 2.484, p = 0.024; self-motion, t(17) = 5.376, p = .000). In the self-motion positive group, the combination condition was significantly smaller than both single-cue conditions (vision, t(17) = 3.177, p = 0.006; self-motion, t(17) = 3.062, p = 0.007), and so was the conflict condition (vision, t(17) = 3.483, p = 0.003; self-motion, t(17) = 2.380, p = 0.029).

With respect to the model predictions, in the vision positive group, both double-cue conditions were consistent with the Bayesian integration model (combination condition, t(17) = 0.099, p = 0.922, BF = 5.560; conflict condition, t(17)= 0.038, p = 0.970, BF = 5.583) and different from the Bayesian alternation model (combination condition, t(17) = 3.544, p = 0.002, BF = 0.061; conflict condition, t(17) = 4.757, p = .000, BF = 0.006). In the self-motion positive group, however, the two double-cue conditions differed from both models, showing a compromise (combination vs. integration, t(17)= 2.334, p= 0.032, BF= 0.576; conflict vs. integration, t(17) = 2.553, p = 0.021, BF = 0.394; combination vs. alternation, t(17) = 6.766, p=0.000, BF= 0.000; conflict vs. alternation, t(17)= 5.792, p=.000, BF=0.001).





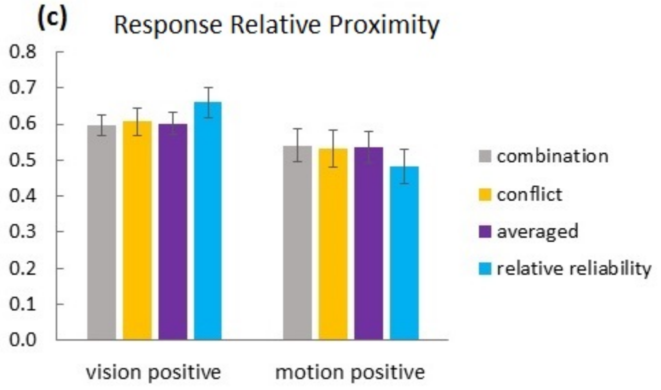


Figure 7. Results of Experiment 2. (a) Response variability and model predictions. The alternation model's prediction represents the average of the predictions for the combination and the conflict conditions. (b) Cue relative reliability as a function of block. (c) Response relative proximity and cue relative reliability; response relative proximity is displayed for the combination condition and the conflict condition respectively, and when the two conditions were averaged. Error bars represent ± SE of the mean.

Surprisingly, cue relative reliability of was substantially higher in the vision positive group than the self-motion positive group (t(34) = 2.738, p = 0.010), indicating that the vision positive group made better use of visual cues relative to self-motion cues compared to the other group A closer look revealed that the vision positive group tended to have lower visual variability and higher self-motion variability than the motion positive group, although neither difference was statistically reliable (Figure 7a, light-blue bars and orange bars; visual variability, t(34) = 1.313, p = 0.198; self-motion variability, t(34) = 1.878, p = 0.069). To make sure that these two groups were equivalent initially, we analyzed cue relative reliability over time. To calculate cue relative reliability in individual blocks, we first estimated the centroid of all responses across all blocks, and then for each block calculated averaged distance of the responses to the centroid. Cue relative reliability in each block is then equal to the inverse ratio of squared averaged distance of the vision and the self-motion conditions, as in the calculation of overall cue relative reliability. Figure 7b shows cue relative reliability in each of the 5 blocks. Cue relative reliability was nearly equal between the two groups in the first block. They began to diverge from each other by the second block. Such results suggest that the two groups were well matched in navigational preferences at the beginning of the experiment. The difference in cue relative reliability was caused by the different types of feedback that they received.

Response Relative Proximity

As shown in Figure 7c, in the vision positive group, response relative proximity in the combination condition tended to be lower than cue relative reliability, and this difference was marginally significant (t(17)= 1.807, p= 0.089). In the conflict condition, the two indices were not different from each other (t(17)= 1.305, p= 0.209). When the two double-cue conditions were averaged, response relative proximity was marginally lower than cue relative reliability (t(17)= 1.694, p= 0.109). In the self-motion positive group, response relative proximity and cue relative reliability were not different from each other in any of the double-cue conditions (combination condition, t(17)= 1.234, p= 0.234; conflict condition, t(17)= 0.785, p= 0.443) or when the conditions were averaged (t(17)= 1.050, p= 0.308).

Table 2 shows Pearson correlations between response relative proximity and cue relative reliability. In the vision positive group, the two indices were positively correlated with each other in both double-cue conditions and when the conditions were averaged. In the self-motion positive group, the two indices were positively correlated in the combination condition, but not in the conflict condition. When the two conditions were averaged, the correlation was marginally significant.

Summary and Discussion

In the current experiment, we displayed distorted feedback to the participants, with the aim of misleading them about relative cue quality between visual and self-motion cues. In the vision positive group, we expected participants to assign greater weight to visual cues and smaller weight to self-motion cues than the Bayesian weights. In the self-motion positive group, we also expected discrepancies between actual weights and Bayesian weights, but in the opposite direction. However, our hypothesis was disconfirmed. Response relative proximity was not different from cue relative reliability in most cases. Furthermore, even though the results failed to reach statistical significance, response relative proximity of visual cues tended to be smaller than the cue relative reliability of visual cues in the vision positive group, meaning that participants assigned smaller weights to visual cues than Bayesian predictions. In the self-motion positive group, response relative proximity tended to be larger than cue relative reliability, meaning that participants tended to weigh visual cues more than expected by Bayesian theory. Thus the results showed the opposite pattern of our initial expectations. If there was any top-down influence of the feedback on the cue-weighting process, the influence should operate in the opposite direction. This was also inconsistent with Bayesian weighting-by-reliability hypothesis, if we consider the reliability composed of the bottom-up stimulus-driven component and the top-down feedback-driven component.

Do our results suggest that the cue-weighting process is purely bottom-up? Response relative proximity was consistent with cue relative reliability most of the time. This pattern of results would lead to the conclusion that participants weighted spatial cues based on the actual cue reliability. In this sense, the cue weighting process was relatively bottom-up in nature, consistent with Ernst and Banks' speculation. However, as stated previously, there was also a tendency in the vision positive group for the actual weights to be smaller than the Bayesian weights; in the self-motion positive group, there was a tendency for the actual weights to be larger than the Bayesian weights. Considering the two groups together, it seemed that even though the actual weights assigned to visual cues tended to be greater in the vision positive group than the self-motion positive group (which was not statistically significant, t(34) = 1.234, p = 0.226), the actual weights did not really catch up with the recently changed level of cue relative reliability caused by the distorted feedback. We refer to this phenomenon as the hysteresis effect. When analyzing the hysteresis effect across the two groups (= rr - rp in the vision positive group; = rp - rr, in the motion positive group), this effect was marginally significant (t(35) = 1.8561, p = .072) and there was no interaction between groups (t(34) = 0.083, p = .935).

Figure 8 summarizes the results pictorially. When we displayed positive feedback to visual cues and negative feedback to self-motion cues, the actual visual reliability increased and the actual self-motion reliability decreased. Accordingly, the Bayesian weight calculated with actual reliabilities moved towards the visual distribution. The actual weight also moved closer to the visual distribution, but did not really catch up the Bayesian weight's changes, creating the hysteresis effect.

Why did this hysteresis effect occur and what are the underlying mechanisms? The hysteresis effect makes sense if we consider the fact that cue relative reliability had been modified very recently during a relatively short period of time. It is possible that without the feedback, cue relative reliability would reflect people's long-ingrained habits of using spatial cues, which probably have formed a strong prior in the navigation system. This strong prior might have drawn cue weights towards levels as if no changes had ever occurred to cue relative reliability in the experiment. Therefore, we need to combine the bottom-up cue-weighting hypothesis and the hysteresis hypothesis to account for our findings.

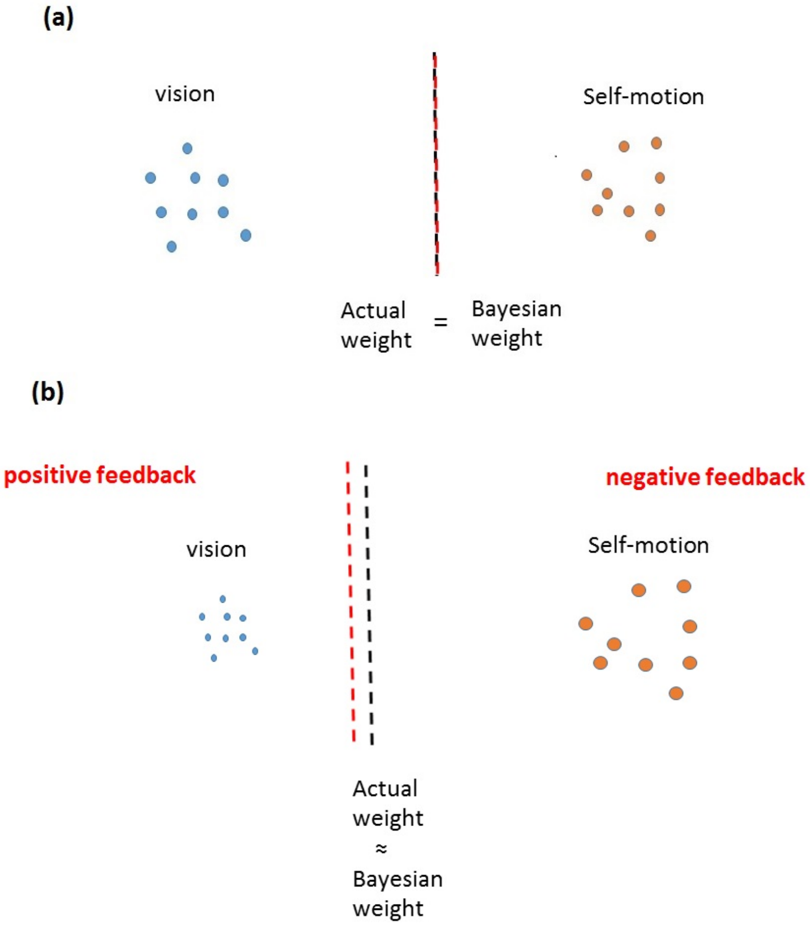


Figure 8. Illustration of results of Experiment 2. (a) As in Figure 5a, Bayesian cue-weighting strategy without feedback. Red and black dotted lines represent Bayesian weight and actual weight respectively. (b) Illustration of findings. When visual cues receive positive feedback and self-motion cues receive negative feedback, the actual visual reliability increases and the actual self-motion reliability decreases. The Bayesian weight becomes closer to visual cues. The actual weight also becomes closer to visual cues, but lags behind to some extent relative to the change of the Bayesian weight.

The hysteresis effect observed in the current experiment raises an interesting question, namely, when cue relative reliability changes, will the cue-weighting process change along with it in a prompt or delayed manner? The findings from the current experiment seem to support the latter hypothesis. Cue relative reliability might be changed by other factors besides distorted feedback. For example, extensive training on a certain spatial cue might be able to improve response precision associated with the cue. Therefore, by incorporating training into the cue integration paradigm, one could investigate: 1, whether extensive training can improve reliability of a spatial cue; and 2, if so, whether the weight it gains would be adjusted accordingly in a timely manner.

Neither our initial hypothesis nor the hypothesis proposed by Ernst and Banks (Ernst & Banks, 2002) predicts the difference in cue relative reliability between the groups. When receiving positive feedback on visual cues and negative feedback on self-motion cues, participants actually performed more precisely with visual cues and less precisely with self-motion cues, compared to participants who received the opposite pattern of feedback. The results implied that participants made better use of a spatial cue if the cue had been shown to be reliable, whereas they did not take full advantage of the cue if it had been shown to be unreliable. To determine why distorted feedback changed how well a spatial cue could be exploited requires further investigation. In all of the other experiments in this project, we consistently observed a wide range of cue relative reliability across participants given exactly the same experimental settings. The findings from the current experiment thus lends further support to the notion that cue reliability is highly subjective and flexible, depending on various subjective factors in addition to cues' intrinsic physical properties.

In terms of response precision, the vision positive group achieved optimal cue integration. However, the self-motion positive group deviated from optimal integration, manifesting a compromise between the integration model and the alternation model. We noted that the vision positive groups always showed positive correlations between actual cue weights assigned and Bayesian weights, whereas these relationships were not significant in the conflict condition in the self-motion positive group. We also noted that the self-motion positive group was relatively more precise with self-motion cues than the vision positive group because of the distorted feedback. The path integration system based on self-motion cues has been shown to be automatic to some degree and has been speculated to be a novelty detection system (Etienne & Jeffery, 2004; Shettleworth & Sutton, 2005). Researchers in other domains also observed that internally-generated haptic cues were used as a reference to distinguish two different visual cues set in conflict, such that the visual cue consistent with haptic cues gained greater weights than predicted by cue reliability (Ernst et al., 2000). This effect may occur because internally-generated signals are generally more trustworthy than externally-given signals, since naturally the former are seldom unexpectedly changed by external forces whereas the latter can be altered easily. It is thus possible that the improved path integration system in the self-motion positive group might have helped participants detect the cue conflict in the conflict condition. Awareness of cue conflict then might have impaired participants' ability to efficiently integrate spatial cues (Berger & Bülthoff, 2009). It will be interesting for future studies to examine how awareness of cue conflict affects cue-weighting behaviors.

CHAPTER V

EXPERIMENT 3: CUE INSTABILITY

The preceding experiments investigated how self-motion cues and visual cues interacted. In the current experiment, we attempted to incorporate cue instability into the paradigm and examined how this factor would affect the cue weighting process.

Cue instability seems to play an important role in how a navigator uses a cue. One general finding is that stable cues are usually weighted more than unstable cues given the same cue properties. For example, it is widely assumed that geometric shapes usually dominate featural cues partly due to the fact that the former are perceived to be more stable in the natural environment across time than the latter (Gallistel, 1990). It is easy to see that a room's shape is usually fixed, whereas a tree is bound to change in shape and color at different times or even disappear. In addition to this assumption, there is a good body of empirical evidence supporting the role of cue stability in the process of spatial navigation.

Burgess and his colleagues attempted to disassociate allocentric and egocentric representations in human spatial memory (Burgess et al., 2004). The allocentric representations were defined in terms of an external cue card, whose stability relative to the to-be-memorized object array on a table was manipulated. Their results showed that participants used the cue card as an allocentric reference point, and their reliance on the cue card was greater when the cue card had always been consistent with the object array in the previous session. This implies that the observer weighted the cue card more when it had been perceived as stable.

With rats, Biegler and Morris demonstrated that landmark stability was required for spatial learning but was not necessary for simple associative learning (Biegler & Morris, 1993). The rat was trained to learn a food location associated with one of the two landmarks. The food location was fixed relative to the relevant landmark in terms of direction and distance. The other landmark was not associated with food. The rat was only able to learn the spatial relation of the food location to the correct landmark when the landmarks were both fixed in the arena. When landmarks moved around randomly, the rat was able to run to the correct landmark, but did not search at the correct location, demonstrating preservation of associative learning but absence of spatial learning. Their results were replicated and extended in a series of subsequent studies (Biegler, 1996; Biegler & Morris, 1996). In another study conducted with rats by Lenck-Santini and his colleagues, a cue card attached to the wall was rotated either when the rat was hidden in its home cage or when the rat was in the process of navigation (Lenck-Santini et al., 2002). The former condition prevented the rat from seeing the cue card move (hidden condition), whereas the latter condition allowed the rat to see the moving cue card (visible condition). Thus, the cue card was more likely to be perceived as unstable in the visible condition than in the hidden condition. Single-cell recording showed that a much larger proportion of place cells followed the cue card in the hidden condition than in the visible condition when the cue card was rotated. And accordingly, the rat searched the card-defined reward location less often in the visible condition. These results implied that the rat relied on the cue card less if it was clearly perceived to be unstable.

Given that cue stability affects how spatial cues are used, how do navigators know whether the cue is stable or unstable in the first place? One source of reference is other visual cues present in the same environment, such as extra-maze landmarks. Another source of reference is internal sense of direction, which corresponds to the path integration system. In a study with rats conducted by Knierim and his colleagues, as a prior training, half of the rats were disoriented every time before they entered the activity arena, so that a steady relationship between internal direction sense and the external visual cues could not be established (Knierim et al., 1995). The other half of the rats remained oriented before entering the arena. In a subsequent test session, the disoriented rats followed the rotation of the visual cue more often than their oriented peers, both evidenced in the firing activity of place cells and head direction cells. Such results suggest rats relied more on visual cues that they had learned to be stable previously. In conjunction with above-mentioned studies, it seems that not only an unstable cue decreased the navigator's reliance on it, but also a cue has to be learned as stable to exert substantial influence on navigational decisions.

We now know that cue stability affect how the navigator weighs the cue, but in what way is this influence accomplished? This is a question never addressed in previous studies. In the framework of Bayesian theory, there are two possible ways. As illustrated in Figure 9, in the indirect pathway cue stability changes cue reliability, and cue reliability in turn controls the cue-weighting process. In the direct pathway, cue stability directly adjusts the cue weights beyond the influence of cue reliability without changing cue reliability. The direct pathway corresponds to the prior information of Bayesian framework and is referred to as the heuristic route in some studies (Byrne & Crawford, 2010). In Experiment 2, we discussed some articles on the influence of active attention on the cue weighting process. Some studies found that paying attention to one sensory channel increased the channel’s weight (Berger & Bülthoff, 2009). These results can be achieved via the two different pathways. However no studies have been conducted to differentiate them regarding the role of active attention.

Most of the above-mentioned studies found that rats usually relied little on unstable cues (Burgess et al., 2004; Knierim et al., 1995; Lenck-Santini et al., 2002). Why are unstable cues relatively ignored? We speculate that there are two different possibilities. One hypothesis is that unstable cues were downplayed because they impaired performance itself; that is the influence was exerted via the indirect pathway, with cue reliability as the mediator. This hypothesis is supported by some above-mentioned studies which showed that cue instability adversely affected rats’ performance (Biegler, 1996; Biegler & Morris, 1993, 1996). Unstable cues might impair performance because they impose extra burden on the memory system, especially the working memory system. In order to exploit unstable spatial cues, the navigator probably needs to constantly update environmental representations and to coordinate the spatial correspondence between the cue’s information and other sources of spatial information, such as path integration and other visual inputs. All of these factors could increase task difficulty. Another hypothesis is that unstable cues might be simply considered less useful regardless how precise they may be. This possibility corresponds to the direct pathway. Usually cues that have moved no longer point to the reward location, which is usually assumed by navigators to be stable relative to the environment. Under such circumstances, if searches still followed the moved cue, they would be centered at a reward-absent location, even though the precision of searches might remain unchanged. In addition, navigators might already have unsuccessful experiences with unstable cues in daily situations and simply decide not to rely on them too much at the moment; this illustrates the influence of prior experiences. What makes the situation more complicated is that these two possibilities might be intermingled. For

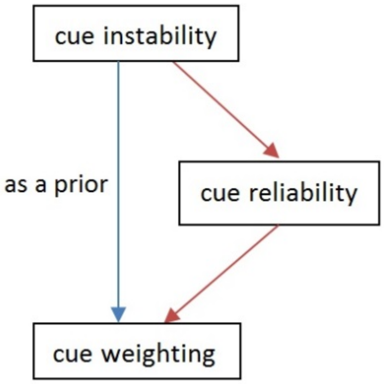


Figure 9. Two possible pathways for cue instability to affect cue weights. The red arrows show the indirect pathway, in which cue reliability serves as a mediator. The blue arrow shows the direct pathway, in which cue instability directly affects cue weighting as a prior.

example, if a cue is predicted by the navigator to change location in future trials, the navigator might not take full advantage of the cue, which would lead to decreased performance and cue reliability. This point has been suggested in our Experiment 2, in which we found that giving feedback to navigators that indicated that a cue is useful actually encouraged them to make better use of the cue.

In the domain of spatial navigation, these two forms of influence have never been differentiated; moreover, the effect of cue instability has never been examined in the cue integration paradigm. The current experiment was designed to test the influence of landmark stability using the cue integration paradigm. We wanted to know whether cue stability would affect the assignment of cue weights and whether it would do so directly or indirectly. To manipulate landmark stability, half of the participants experienced landmarks changing in locations in different layouts (unstable group), whereas the other half experienced landmarks stable across the whole experiment (stable group). As an initial prediction, we expected that ultimately the unstable group would assign smaller weights to visual landmarks than the stable group, via either the direct pathway or the indirect pathway. We still predicted that people would integrate visual and self-motion cues, so variability reduction would be reduced in double-cue conditions relative to single-cue conditions.

Method

Participants

Forty-two people from Vanderbilt University and the Nashville community participated in this experiment. They all had normal or corrected-to-normal vision. They ranged in age from 18 to 28 years old, with a mean of 18.9.

Material and Procedure

In this experiment, the same 3 landmarks were displayed throughout the whole procedure. These landmarks were the same as those used in the rich condition of Experiment 1A. Participants were randomly assigned to the unstable landmarks group or to the stable landmarks group. In the unstable group, each participant experienced 10 different landmark configurations and the order of the 3 landmarks in the configuration was determined randomly on each trial. Figure 10 shows the 10 different configurations and the 4 target post locations. The same 4 conditions were included, vision, self-motion, combination, and conflict conditions. Each condition contained 10 trials, and each trial used one of the 10 landmark configurations. Hence, the 4 conditions were matched in terms of landmark configurations. In the stable group, we selected one of the 10 landmark configurations for each subject and the same configuration was presented throughout the experiment.

We also varied the distance of the landmarks from the locations of the posts between participants, because we had observed that landmarks too close to the target locations might result in a ceiling effect in the vision condition. Figure 10 depicts the closer landmarks. For the farther landmarks, landmarks were placed twice as far along the radial direction. The patterns of results were quite similar for closer and farther landmarks, so we combined closer and farther landmarks together, and then compared stable and unstable groups. Each of these groups included 21 participants.

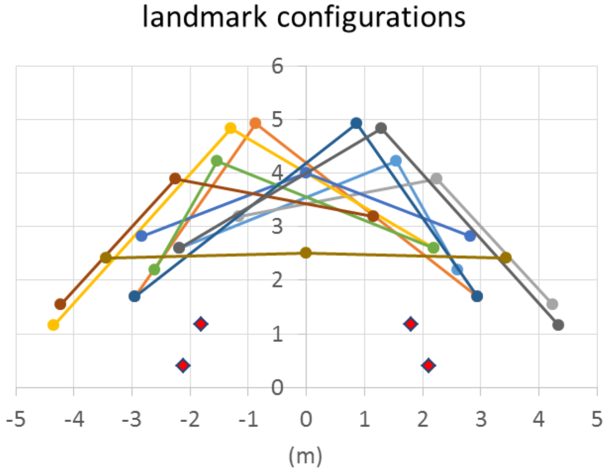


Figure 10. Landmark configurations used in Experiment 3. The three dots inter-connected with lines represent 3 landmarks in a landmark configuration. Two configurations were symmetrical around the midline (light blue lines and yellow brown lines). The other eight configurations were asymmetrical, but each was a mirror image of another configuration. Red diamonds represent target locations without jitter. Each square of the grid represents 1m x 1m.

Results

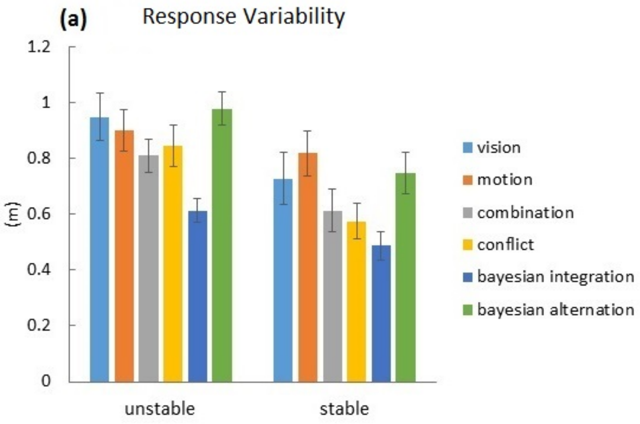
Response Variability

As shown in Figure 11a, visual inspection of the bar graphs indicates that there was variability reduction in double-cue conditions compared to single-cue conditions. However, in the unstable group, the variability reduction did not reach significance in any of the comparisons (vision vs. combination, t(20) = 1.715, p = 0.102; motion vs. combination, t(20) = 1.305, p = 0.207; vision vs. conflict, t(20) = 1.238, p = 0.230; motion vs. conflict, t(20) = 0.639, p = 0.530). In the stable group, double-cue conditions showed significant variability reduction relative to the self-motion condition (motion vs. combination, t(20) = 2.908, p = 0.009; motion vs. conflict, t (20) = 2.401, p = 0.026), but not relative to the vision condition (vision vs. combination, t (20) = 1.458, p = 0.160; vision vs. conflict, t (20) = 1.659, p = 0.113). Overall, the trends in the stable group were consistent with the results observed in previous experiments. The stable group showed marginally lower response variability in the vision condition than the unstable group (t(40) = 1.751, p = .088). When averaging response variability across all four conditions, the stable group showed significantly lower variability than the unstable group (t(40) = 2.418, p = .020).

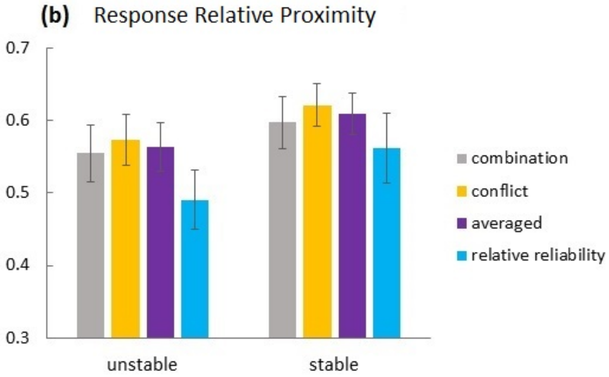
As shown in Table 1, for the unstable group, both double-cue conditions were different from the Bayesian integration model (combination vs. Bayesian integration, t(20) = 3.805, p = 0.001, BF = 0.032; conflict vs. Bayesian integration, t(20) = 3.472, p = 0.002, BF = 0.064); the Bayesian alternation model also did not fit the data (combination vs. Bayesian alternation, t(20) = 2.413, p = 0.026, BF = 0.509; conflict vs. Bayesian alternation, t(20) = 1.794, p = 0.088, BF = 1.417). For the stable group, the combination condition was different from the Bayesian integration model (t(20) = 3.063, p = 0.006, BF = 0.147), whereas the conflict condition seemed to be explained by it (t(20) = 1.328, p = 0.199, BF = 2.647); both double-cue conditions were different from the Bayesian alternation model (combination vs. Bayesian alternation, t(20) = 2.663, p = 0.015, BF = 0.320; conflict vs. Bayesian alternation, t(20) = -2.516, p = 0.021, BF = 0.422). The evidence for cue integration was greater in the stable group than in the unstable group. However, overall cue integration was weaker in this experiment compared to the preceding experiments.

Response Relative Proximity

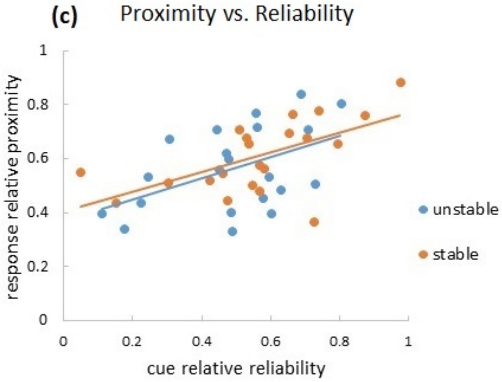
As shown in Figure 11b, for both groups, there was a tendency for response relative proximity to be larger than cue relative reliability. In the unstable group, this tendency was significant for the conflict condition (t(20) = 2.243, p = .036), marginally significant for the average (t(20) = 1.880, p = .075), and non-significant in the combination condition (t(20) = 1.367, p = 0.187). In the stable group, response relative proximity and cue relative reliability were not different (combination condition, t(20) = 0.875, p = 0.392; conflict condition, t(20) = 1.406, p = 0.175; averaged, t(20) = 1.235, p = 0.231). However, when we analyzed the two groups together, this tendency was significant (F(1,40) = 4.863, p = .033) and there was no difference between the groups (F(1,40) = .220, p = .642). Considering that participants in both groups experienced the same set of irregular landmark configurations, this pattern of results suggests that when the landmark configuration was irregular participants overweighted visual landmarks



(a) Response Variability



(b) Response Relative Proximity



(c) Proximity vs. Reliability

Figure 11. Results of Experiment 3. (a) Response variability and model predictions. The alternation model's prediction represents the average of the predictions for the combination and the conflict conditions. (b) Response relative proximity and cue relative reliability; response relative proximity is displayed for the combination condition and the conflict condition respectively, and when the two conditions were averaged. (c) Correlations across participants between averaged response relative proximity and cue relative reliability. Error bars represent ± SE of the mean.

relative to Bayesian predictions, regardless of whether the configuration was stable or unstable. When comparing the two groups, there was a tendency for the unstable group to assign smaller weights to landmarks and to have a lower level of visual reliability, but this tendency was not significant (F(1,40) = 1.533, p = .223).

For both groups, positive correlations were found between relative proximity and relative reliability, demonstrating that cues were weighted based on their relative reliability (Table 2). The scatter-plot shows relationships between relative reliability and relative proximity averaged across combination and conflict conditions (Figure 11c).

Summary and Discussion

In contrast to Experiments 1-3, variability reduction in double-cue conditions was weak in the current experiment, and it was even weaker in the unstable group than in the stable group. Response variability was not significantly lower in double-cue conditions than in single-cue conditions, although such trends existed in both groups. This pattern of findings was not predicted by our initial hypothesis. One question is whether this pattern of results is a sign of single cue dominance. Relative proximity in the combination condition was well between 0 and 1 (which would indicate complete reliance on self-motion cues and visual cues, respectively), countering a single-cue-dominance hypothesis. Another hypothesis is that participants adopted a mixed strategy of cue integration and cue alternation, considering that double-cue variability seemed to be a compromise between the cue integration model and the cue alternation model. A third possibility is that participants tried to combine landmarks and self-motion cues, but this integration was obscured by our experimental design. Different landmark layouts might have introduced additional variability to the data. For the stable group, even though each participant experienced a stable landmark layout, the 10 different landmark configurations seen by different participants might vary in visual cue reliability. For the unstable group, each participant experienced 10 different landmark layouts. Most of the landmark configurations were not symmetrical. It is possible that the target locations on different sides are associated with different levels of visual cue reliability and bias, which means that the ultimate response distribution in the vision condition for a particular participant might be a composite of different distributions with different variances and centroids.

As to the actual assigned weights, participants seemed to weight the visual landmarks more than Bayesian predictions in both groups, since response relative proximity was larger than cue relative reliability (Figure 11b, in both groups, the blue bars, representing cue relative reliability, are lower than all the other three bars, representing the actual visual weights). This result counters the Bayesian hypothesis, since the actual weights were not consistent with cue relative reliability. In addition, this evidence also counters the heuristic hypothesis, which proposes that unstable landmarks would be weighed less than predicted by cue relative reliability. If layout asymmetry affects cue-weighting, it probably would weaken landmarks' ability to gain navigators' consideration, because irregular object layouts or irregular geometric space are usually harder to learn than regular ones (Adina R. Lew, Gibbons, Murphy, & Bremner, 2010). One possibility is that irregular and unstable landmark configurations can impair spatial learning. Two different forms of learning have been frequently contrasted in spatial navigation literature, associative learning and spatial learning. In associative learning, only identities of reference points are learned, for example which landmark is associated with reward and which is not. Spatial learning, on the other hand, involves memorizing spatial relations relative to reference points, like distance and direction, in addition to recognizing reference points. In a series of studies with rats conducted by Biegler and his colleagues (Biegler, 1996; Biegler & Morris, 1993, 1996), unstable landmarks impaired spatial learning and searches tended to cluster around the reward-associated landmark. Such clustering means that the centroid of the response distribution was closer to the landmark, a pattern of results found in the current experiment.

To summarize the results on variability and proximity, it seemed that the two groups did not differ in terms of variability reduction or response relative proximity. This suggests that landmark stability did not really affect the ways participants handled the cue integration problem. However, there was a tendency for the unstable group to have lower visual reliability than the stable group (t(40) = 1.751, p = 0.088, Cohen's d = 0.540, power = 0.401). We also notice a tendency for the unstable group to assign smaller weights to visual cues than the stable group (t(40) = 1.025, p = 0.312, Cohen's d = 0.316, power = 0.170). Our interpretation of these results is that the differences caused by landmark stability might have been incorporated into the component of visual cue reliability, but did not affect the cue integration process This interpretation is consistent with the indirect pathway hypothesis on the role of cue stability, in that cue instability decreased cue reliability, which in turn reduced the weights. However, those patterns of results did not reach statistical significance and power was low. These limitations existed because the effect sizes were small to medium and between-subjects comparisons with relatively small sample sizes were used. Future investigations should attempt to increase the power by increasing cue instability in the unstable group, using a within-subjects design, and testing more participants.

The cue instability examined in the current experiment can be conceptualized as explicit instability, for the reason that the instability was present in current sources of information and obvious to the observer. The contrasting type of cue instability is implicit instability. Implicit instability is not contained in current sources of information; instead it is suggested by the cue's identity or navigators' history of using the cue. For example, geometric cues are usually assumed to be more stable than featural cues, even when both cues remain stable throughout the experiment. In this sense, in the navigators' eyes, geometric cues are implicitly assumed to be more stable than featural cues. As another example, if we replace trees and buildings used in our experiments with animal models to serve as landmarks, the latter would probably be assumed to be less stable because animals are known to be able to move voluntarily whereas trees and buildings cannot. In addition to pre-existing implicit stability, it might be possible to manipulate implicit stability by exposing navigators before the test task to cues which move explicitly. During the test, the cues remain stable. It will be interesting to examine whether these two forms of cue instability influence cue weighting process differently in future studies. Our speculation is that implicit instability can also influence how cues are weighed, but probably via the direct pathway.

CHAPTER VI

EXPERIMENT 4: PROXIMAL AND DISTAL LANDMARKS

The previous experiments demonstrated that overall participants weighted spatial cues based on cue relative reliability, although there were sometimes deviations from Bayesian predictions at the group level. However, the previous experiments focused only on self-motion cues and visual landmark cues. As described in the general introduction, spatial cues are primarily divided into two categories, internally-generated self-motion cues and externally-given visual cues. The latter visual cues can be further divided into proximal vs. distal cues and geometric vs. featural cues. It is therefore important to extend the cue integration paradigm to investigate the interaction issue for different visual cues. Different visual cues probably belong to the same sensory modality or similar sensory modalities, whereas self-motion cues and visual cues belong to distinct modalities. Investigating interactions between visual cues, and comparing the results to those from the preceding experiments would reveal whether navigators handle within-modality and inter-modality cue interactions differently.

It has been demonstrated in other domains that within- and inter-modality cue interactions followed different principles. For example, Hills et al. demonstrated that in shape perception, visual and haptic cues interacted with each other differently from disparity and texture cues, in that the latter induced mandatory fusion of cues whereas the former did not (Hillis, Ernst, Banks, & Landy, 2002). Mou and Spetch tested people in a location-change detection task when they remained stationary (Mou & Spetch, 2013). They found that people could integrate in a Bayesian optimal way egocentric information defined by the observer's body and allocentric information defined by context objects, and that little or no competition existed between these two sources of information. On the contrary, the integration of different context objects yielded better performance than the optimal level, and competition existed between them. Thus, it is possible that different mechanisms also exit between within-modality and inter-modality cue interaction in the domain of spatial navigation.

The current experiment represents our pioneering efforts to investigate this issue. We focused on proximal and distal landmarks. These two types of spatial cues have been frequently compared in the animal literature. Cue distance has been shown to be a critical factor in determining cues' weights. For example, animal behavioral studies have frequently shown that the competition capacity of a landmark increases with decreasing distance from the target location (Spetch, 1995). Single-cell recordings of hippocampal cells have found that some cells were controlled by proximal cues and some followed distal cues, when these two cues were dissociated by arena rotation (Knierim, 2002). There is also evidence that proximal cues and distal cues relied on different brain areas in the navigation process (Save & Poucet, 2000). For simplicity and to be parallel to most behavioral and physiological studies on animals, here we defined proximal and distal landmarks as objects differing substantially in distance so that they can be conceptualized to reside in different spatial contexts. One common example is landmarks inside vs. outside of the running arena. The arena can be considered as the immediate or local environment, whereas space beyond the arena can be considered as the distant or global environment.

Comparisons of this kind between proximal and distal landmarks are relatively rare in human studies. In addition, it has never been examined in a cue integration paradigm. Most importantly, previous studies did not examine how the weights are assigned to proximal and distal cues, besides the common finding that both cue types contributed to location searches. The current experiment modified the spatial task used in previous experiments and applied it on a 2D desktop virtual environment.

Method

Participants

Forty people (19 male) from Vanderbilt University or the Nashville community participated in this experiment for course credits or monetary compensation. They all had normal or corrected-to-normal vision. They ranged in age from 18 to 30 years old, with a mean of 21.08.

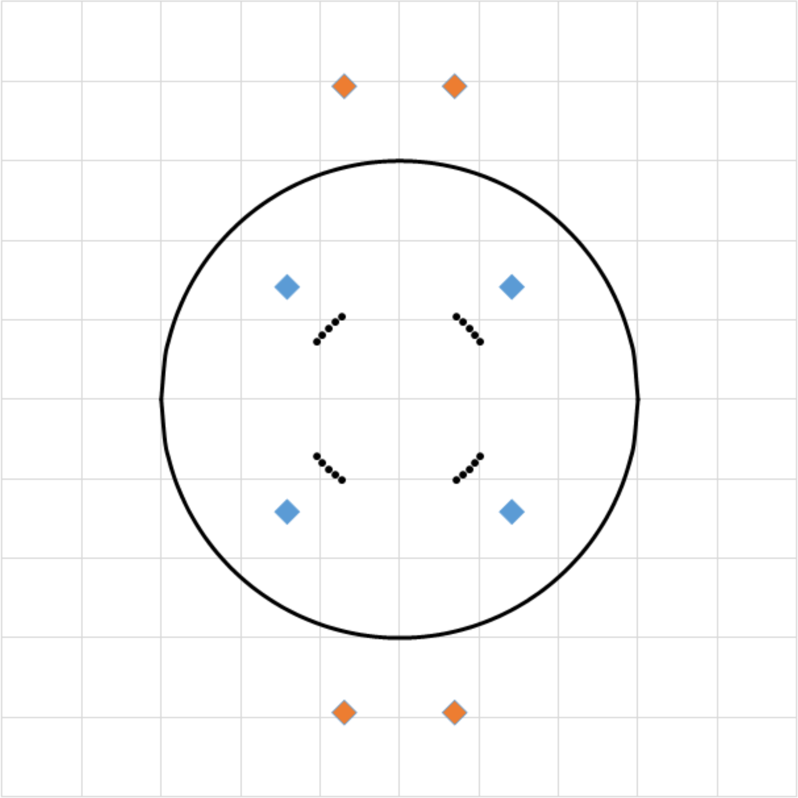


Figure 12. Bird's eye view of the environment layout in Experiment 4. Dark circle represents the enclosure. Black dots represent target locations. Blue diamonds represent proximal landmarks inside the enclosure, and orange diamonds represent distal landmarks outside of the enclosure. Proximal landmarks were positioned 4 m from the enclosure center. Distal landmarks were actually 30 m away from the environment center, but in the graph they were placed at a distance of 8 m for easy illustration. Each square of the grid represents 2m x 2m.

Material and Procedure

In this study, we examined how people integrate proximal and distal visual landmarks. The environment was displayed on a desktop computer, and participants used a joystick to navigate inside an enclosure, which was 12m wide in diameter. The environment is depicted in Figure 12. Four distinct landmarks were positioned inside the enclosure as proximal cues; and the other four distinct landmarks were positioned outside the enclosure as distal cues. Proximal landmarks and distal landmarks were positioned 4m and 30m from the enclosure center, respectively.

During the encoding phase, all proximal and distal cues were present. Participants navigated to a red post at one of the target locations. The post disappeared once participants reached its location, at which point participants could not translate their perspective but could rotate it to look around and get views of the surrounding. Once participants indicated that they had memorized the target location, their perspective was transported to the enclosure center in darkness and the facing direction was randomly determined. After 2 seconds of a blank screen, the environment re-appeared and participants attempted to navigate back to the target location as they remembered it.

During this retrieval stage, only proximal cues reappeared in the proximal condition, only distal cues re-appeared in the distal condition, and both proximal and distal cues re-appeared in the combination condition. In the conflict condition, the locations of proximal and distal cues were shifted to create a disparity in the correct location defined by the two different cue types. Considering the 4-fold symmetry of the environment, the disparity was first created in the first quadrant (upper right in Figure 12) and then was matched to the other 3 quadrants. For participants in group 1 (n=21, 11 male), for target locations in the first quadrant (upper right) and the third quadrant (lower left), the four distal cues as a whole were shifted counterclockwise by 7.5^o and the four proximal cues as whole were shifted clockwise by 7.5^o. The rotation directions for proximal and distal landmarks were switched for the second quadrant (upper left) and the fourth quadrant (lower right). In this way, proximal and distal cues were pulled closer by 15^o in total. For participants in group 2 (n=19, 8 male), the shift directions of proximal and distal cues were reversed relative to group 1, so that proximal and distal cues were pulled further apart by 15^o in each quadrant. The enclosure was displayed at both encoding and retrieval stages. The enclosure was included because we wanted to make the environment similar to those typically used in animal studies.

There were 5 target locations in each quadrant, resulting in 20 target locations for 20 trials in each condition. No jitter was added to the target locations. Data analysis followed the same procedure used in previous experiments, except that there were now four quadrants to consider instead of two. Responses were first transformed to a spatial coordinate with the target location as the origin and the direction from environment center to the target location as the y-axis. Responses for the 5 target locations within the same quadrant were pooled together. Outliers were defined as responses whose distance from the response centroid exceeded the 3rd quartile by 3 times the interquartile range, considering all 4 quadrants together. Before pooling responses from the 4 quadrants, we corrected the bias in each quadrant's response distribution by centering each distribution over its centroid. After that, we computed the response variability. Cue relative reliability was computed as the inverse ratio of response variances of proximal and distal landmarks. This index assessed the relative reliability of proximal landmarks. Response relative proximity was computed for each quadrant, and was then averaged across the four quadrants to get a final estimation of the actual weights. Response relative proximity assessed the actual weights assigned to proximal landmarks relative to distal landmarks.

We defined individuals as outliers when the averaged response variability across all conditions and both environments was 3 interquartile ranges above the 3rd quartile or below the 1st quartile. In correlational analyses, bivariate outliers were defined as having unstandardized residuals greater than the 3rd quartile by 3 times the interquartile range in the linear regression analysis. By these standards, no participants were identified as outliers.

Results

A subgroup of participants (n=14, 8 participants in group 1 and 6 participants in group 2) experienced proximal landmarks which were slightly larger and closer to target locations whereas other participants (n=26, 13 participants in each group) experienced smaller and more distant proximal landmarks (still inside the enclosure). Patterns of results did not differ between these participants, so all participants were analyzed together.

Response Variability

Results were analyzed for group 1 and group 2 both separately and jointly. Visual inspection of Figure 13a reveals that there was little evidence of variability reduction. Double-cue conditions either showed similar variability as the more reliable proximal condition or close to the less reliable distal condition. In addition, overall double-cue conditions deviated from

Bayesian integration predictions, either indicating a compromise between the integration model and the alternation model or being closer to the alternation model. Proximal landmarks showed lower response variability than distal landmarks. These observations were confirmed by statistical tests.

In group 1, the combination condition had lower variability than the distal condition (t(20) = 5.078, p = .000), but did not differ from the proximal condition (t(20) = 0.464, p = .648). The combination condition differed from predictions of both the integration model (t(20) = 2.703, p = 0.014, BF = 0.297) and the alternation model (t(20) = 4.227, p = 0.000, BF = 0.013), showing a compromise between these two models. The conflict condition showed similar patterns. The conflict condition was lower than the distal condition (t(20) = 3.556, p = 0.002), but was not different from the proximal condition (t(20) = 0.907, p = 0.375). The conflict condition differed from both the integration model (t(20) = 3.431, p = 0.003, BF = 0.069) and the alternation model (t(20) = 2.473, p = 0.022, BF = 0.456), again a compromise between the two models.

In group 2, results were similar, but with even weaker evidence of variability reduction. The combination condition was lower than the distal condition (t(18) = 4.328, p = 0.000), but did not differ from the proximal condition (t(18) = 1.068 , p = 0.300). The combination condition differed from both the integration model (t(18) = 3.514, p = 0.002, BF = 0.063) and the alternation model (t(18) = 2.549, p = 0.020, BF = 0.396), but with a greater likelihood to be consistent with the alternation model. The conflict condition was marginally lower than the distal condition (t(18) = 2.035, p = 0.057) and higher than the proximal condition(t(18) = 2.480, p = 0.023). The conflict condition differed from the integration model (t(18) = 5.981, p = 0.000, BF = 0.000), but was consistent with the alternation model (t(18) = 0.651, p = 0.523, BF = 4.678).

When the two groups were analyzed together, results showed similar patterns. That is, first, there was weak evidence of variability reduction; second, the double-cue conditions seemed to be a compromise between the integration model and the alternation model, sometimes consistent with the latter.

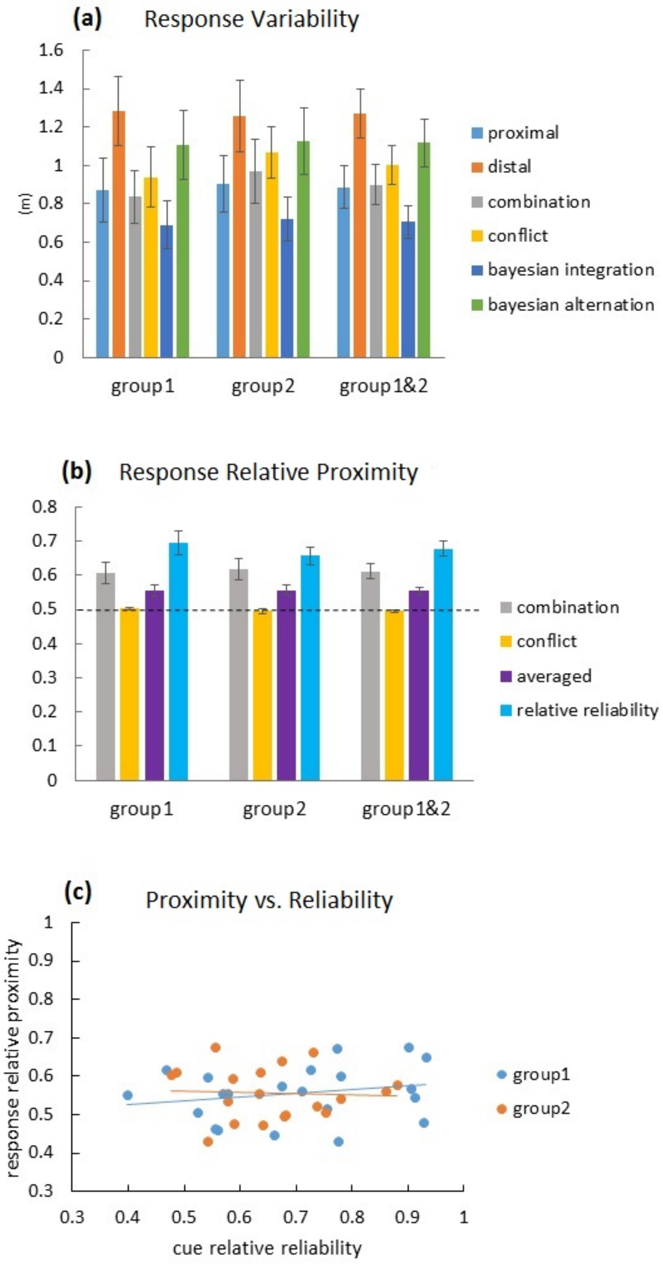


Figure 13. Results of Experiment 4. (a) Response variability and model predictions. The alternation model's prediction represents the average of the predictions for the combination and the conflict conditions. For group 1, proximal and distal cues were pushed closer by 15º in the conflict condition; for group 2, proximal and distal cues were pulled further away by 15º in the conflict condition. (b) Response relative proximity and cue relative reliability; response relative proximity is displayed for the combination condition and the conflict condition respectively, and when the two conditions were averaged. (c) Correlations across participants between averaged response relative proximity and cue relative reliability.

Response Relative Proximity

As shown in Figure 13b, the actual assigned weights to the proximal landmarks were lower than the Bayesian prediction. It is especially the case in the conflict condition, where participants very consistently assigned equivalent weights to proximal and distal landmarks (rp = .5), regardless of the fact that the former was substantially more reliable than the latter.

These observations were confirmed by statistical tests. In group 1, response relative proximity was lower than cue relative reliability in the combination condition (t(20) = 2.109, p = 0.048) and in the conflict condition (t(20) = 5.414, p = 0.000) and when the two double-cue conditions were averaged (t(20) = 3.974, p = .001). In group 2, the two indices were consistent with each other in the combination condition (t(18) = 0.945, p = 0.357), but response relative proximity was lower than the reliability index in the conflict condition (t(18) = 6.011, p = 0.000) and when the two double-cue conditions were averaged (t(18) = 3.248, p = 0.004). When we compare the two double-cue conditions, participants assigned smaller weights to proximal landmarks in the conflict condition than the combination condition in both groups (group 1, t(20) = 3.136, p = 0.005; group 2, t(18) = 3.883, p = 0.001). This means that the disparity between response relative proximity and cue relative reliability was larger in the conflict condition than in the combination condition.

As shown in Table 2, in contrast to the previous experiments, correlations between response relative proximity and cue relative reliability were not significant in any conditions (Group 1: combination, r=0.210, p = 0.360; conflict, r = 0.022, p = 0.924; averaged, r = 0.217, p = 0.344. Group 2: combination, r = - 0.044, p = 0.858; conflict, r = - 0.036, p = 0.883; averaged, r = -0.051, p = 0.836). In the scatter-plot (Figure 13c), the response relative proximity was rather constrained and tended to cluster around the level of 0.5, whereas cue relative reliability spans a large range from 0.3 to 1.0.

Summary and Discussion

The obtained results indicate that participants in the current experiment did not integrate proximal and distal cues in a Bayesian way. First, participants did not seem to take advantage of double-cue displays relative to single-cue displays, in that no variability reduction was observed in double-cue conditions relative to single-cue conditions and in some cases double-cue conditions were consistent with the cue alternation model. Second, participants did not weight proximal cues and distal cues based on cue relative reliability. The actual assigned weights differed from cue relative reliability in most conditions, and no significant correlations existed between them.

One possible explanation for these discrepancies from Bayesian principles is that cue interaction between different types of visual cues does not follow Bayesian principles, as opposed to cue interaction between internal self-motion cues and external visual landmarks. This hypothesis suggests that there might be fundamentally different rules governing within-modality vs. inter-modality cue interaction in spatial navigation.

Another possible explanation is that the current paradigm was not appropriate for examining the spatial cue interaction issue on a 2D desktop virtual environment. Participants frequently reported after the experiment that they used a visual-matching strategy. They often used the screen frame as a reference. They attempted to fit proximal and/or distal landmarks into the screen and tried to match the landmarks' positions to remembered positions in relation to the screen frame. The reaction time was longer in double-cue conditions than in single-cue conditions (21.50s vs. 17.32s, $p = 0.000$), suggesting that participants spent more time on fitting two landmarks into the screen than one landmark. This hypothesis implies that a visual-matching strategy probably does not result in variability-reducing behaviors in multiple-cue occasions.

A third hypothesis is also related to the 2D desktop display. The stable screen frame might have helped participants to notice the conflict between landmarks in the conflict condition. They might have noticed that they could not match the current snapshot with the remembered snapshot in the conflict condition. This increased awareness of cue conflict might thus have discouraged participants from integrating cues (as in Experiment 2, the motion positive group). This hypothesis implies that consciousness of cue inconsistency can disrupt cue integration behaviors, because it means that one of the cues should not be trusted.

Future studies need to examine which of these three factors contributes to or whether all contribute to disrupting cue integration behaviors.

We also found that participants tended to assign lower weights to proximal cues than predicted by Bayesian theory, especially in conflict conditions. One possible explanation of this result is that prior information played a role in influencing the weighting process in addition to the effect of cue reliability. Previous studies have suggested that navigators tended to rely more on distal cues than proximal cues, probably because the former cues are often assumed to be more stable in position, and also usually maintain relatively fixed spatial relations to the moving navigator throughout the navigation process (Cressant, Muller, & Poucet, 1997; A. R. Lew, 2011; Save & Poucet, 2000). This hypothesis is consistent with the direct pathway as illustrated in Experiment 3 (Figure 9). Some properties of the distal landmarks serve as the prior information, which is probably the implicit stability of the distal landmarks. Another possible explanation is that when employing the visual-matching strategy, participants attempted to assign equal weights to both cue types, even though the proximal landmarks were substantially more reliable than the distal landmarks. The actual weights were very near .50 in the conflict condition and this pattern was quite consistent across participants (see the very small error bars in Figure 13b). Therefore, it is quite possible that weighting behaviors in the combination condition might have been affected by the strategy used in the conflict condition.

To summarize, overall, participants did not follow the Bayesian principles when dealing with two different visual cues. However, we cannot draw firm conclusions without excluding the confounding effect of the visual-matching strategy widely adopted by participants. The current paradigm might need to be modified to discourage or even eliminate the visual-matching strategy. Such a modification could be crucial if we want to extend the topic to investigate any other pairs of visual cues on a desktop virtual environment. One possible solution is to conduct the same experiment using the immersive virtual display. This might be able to decrease participants' reliance on the visual-matching strategy, since the HMD is much less stable than a desktop screen frame.

CHAPTER VII

GENERAL DISCUSSION

Summary

In the current project, we applied Bayesian principles to investigate cue interaction in human spatial navigation. In Experiments 1A and 1B, we extended previous pioneering work by manipulating cue reliability within-subjects. We found that overall participants' behaviors were consistent with Bayesian integration, replicating the critical findings in previous work. However, such optimal cue integration often came at a cost of participants being sub-optimal at individual reliability levels. There was interference between reliability levels and its pattern suggested that participants blended reliability levels they experienced simultaneously.

In Experiment 2, we aimed to gain a deeper understanding of the concept of cue reliability. We asked whether the cue-weighting process is purely bottom-up, such that cue weights are determined solely by cues' objective reliability, or whether the process can be influenced in a top-down manner by participants' subjective evaluation of cue quality. Participants' subjective impressions of cue reliability were manipulated by distorted feedback. Our results leaned towards the bottom-up hypothesis. Misleading participants to believe that a cue was more reliable that it actually was did not result in greater weights assigned to the cue than predicted by measured cue reliability. The results showed tendencies opposite to the top-down hypothesis; for instance, the actual weights tended to be lower than the Bayesian weights when participants were misled to believe that the cue was more reliable than it actually was. We also discovered an interesting phenomenon in which distorted feedback changed cue reliability measured objectively. Positive feedback enhanced cue reliability, whereas negative feedback impaired it. A hysteresis hypothesis might be able to explain the discrepancies between actual weights and cue reliability predictions. The hysteresis hypothesis states that the assignment of weights to cues lagged behind changes in cue reliability, and proposes that the recently altered cue reliability needs to be interpreted in the context of a prior reliability component.

In Experiment 3, we tried to incorporate the influence of visual cue stability into the cue integration paradigm. We wanted to know whether cue stability would affect cue weights and in what form such influence would be exerted. Even though the trends were not significant, our results seemed to support the indirect influence pathway; that is, cue stability changed cue reliability, and cue reliability influences cue weights subsequently. The current study might not have enough statistical power to detect these patterns. We need to increase the effect size as well as the sample size in future studies.

In all of the first 3 experiments, we examined self-motion cues and visual landmarks. A universal finding was that participants weighted spatial cues in terms of cue relative reliability, as evidenced in positive correlations between response relative proximity and cue relative reliability, even though sometimes the two indices did not accord with each other at the group level.

In Experiment 4, we modified the cue integration paradigm in immersive virtual environment and adapted it to a 2D desktop virtual display. This paradigm could be used to study any pairs of visual cues. We examined the interaction between proximal landmarks and distal landmarks as an initial investigation. Quite different from interactions between self-motion cues and visual cues, interactions between these two types of visual cues did not follow Bayesian principles. Cue reliability did not seem to be a factor substantially influencing cue weights. This might be caused by the snapshot-matching strategy used frequently by participants. Future studies need to develop solutions to prevent such a strategy to further examine differences between cue interaction within the visual system and between the visual system and the path integration system.

The application of Bayesian principles to cue integration problem in spatial navigation is unusual. To our knowledge, there are only two closely related studies. The first study was conducted by Nardini et al. (Nardini et al., 2008). They compared adults with children using the cue integration paradigm. The study was run in a real environment. Results showed that adults were able to combine self-motion cues and visual landmarks optimally in a Bayesian way, which is consistent with our findings that people were optimal across reliability levels in Experiments 1 and 2. On the contrary, children did not integrate but alternated between the cues. Our project replicated Nardini et al.'s main findings on adults in immersive virtual environments. This consistency of findings shows that immersive virtual reality technologies are a valid tool to study human spatial navigation.

Our project, however, differed from Nardini's study in experimental design and data analysis. We included 10 trials in each condition instead of just 4. In addition, we applied Bayesian calculations to each individual participant. Instead of assuming that all persons would adopt a common strategy, we allowed as well as accounted for individual differences in the data analysis. Using small numbers of trials and holding incorrect assumptions on individual differences can be problematic. Appendix 2 includes simulation work showing that to implement Bayesian principles, it is critical to embrace correct assumptions about individual differences and to have a sufficient number of trials in the task. Another critical difference between our project and Nardini et al.'s study is the way we created cue conflict. When setting visual cues and self-motion cues in conflict, we always rotated the visual world clockwise, whereas Nardini et al. used both clockwise and counterclockwise rotations for each participant. Their design may be problematic if the two single-cue distributions were inherently biased and did not share the same centroid. In this case, rotations of the same amount but to opposite directions will create different degrees of conflict between the cues, making it invalid to pool all the conflict trials together. We believe that the visual rotation direction should be counterbalanced between subjects not within subjects.

Another study is a recent project reported by Zhao and Warren (Zhao & Warren, 2010, November). Overall, our findings are consistent with their observations, in that we both discovered that participants could optimally or nearly optimally integrate visual landmarks and self-motion cues. However, in their conflict condition, Zhao and Warren observed a dissociation between response relative proximity and response variability. When the conflict was not greater than 90^{o}, participants followed either landmarks or self-motion cues based on response relative proximity, but still showed optimal reduction in response variability. We only tested a 15^{o} conflict, and found that response relative proximity seemed to be a compromise between landmarks and self-motion cues, and this compromise was predicted by cue relative reliability. We speculate that this discrepancy may be explained by different experimental procedures. In

Zhao and Warren's study, both spatial cues were stable and participants practiced the task for a very long time (5 days). Participants also only experienced the conflict condition in the conflict sessions. Extensive familiarity with the experimental procedure and the blocked conditions might have encouraged the tendency to rely on one spatial cue only during the initial stage of response. It might also have allowed people to notice the landmark shift and even estimate the amount of shift. Participants then could have mentally corrected the landmark shift and been able to optimally integrate corrected landmarks with the self-motion system in a later stage of response. Familiarity with the experimental set-up was also evident when the authors found that the walked distance was not informative among different conditions, since the participants might have learned the constant distance to walk from the path end to the target location.

The current project not only largely replicated the results obtained by past pioneering work but also extended this work by asking and answering new questions. Together, our results provide new insights into spatial cue interaction in human navigation process and also impose new questions to look into in future studies.

The Composite Nature of Cue Reliability

In the current project, we have shown that cue reliability plays an important role in influencing the cue weighting process. In our first three experiments, cue weights were found to be positively correlated with cue relative reliability across participants. But what does cue reliability really mean? It is critical to gain a thorough understanding of cue reliability in order to interpret mechanisms underlying the navigation process.

Cue reliability is at its essence cue quality, which is determined by a cue's physical properties. Methodologically, we cannot measure cue reliability directly; the only way to know about cue reliability quantitatively is to measure the precision of the navigator' responses given the cue. For this reason, researchers tend to equate response precision with cue reliability. However, this impression is misleading. Whereas cue reliability is equivalent to cue quality, response precision measures how well the navigator can perform with the cue. Navigational performance is jointly determined by two factors, cue quality and efficiency of cue exploitation. Exploitation efficiency manifests itself when different individuals show different levels of response precision given the same environmental cue. Therefore, response precision is determined by more than just cue quality. If cue quality determines an objective level of cue reliability, then cue exploitation can be conceptualized to scale it up or down depending on how much the navigator can get out of the cue.

Our first three experiments found that cue reliability can be highly flexible and subjective, not only in that a wide range of values existed across different individuals, but also in that it seemed that we could alter cue reliability within-subjects through experimental manipulations. Based on these results, we speculate that cue reliability contains at least 3 different components. As shown in Figure 14, measured response precision (equivalent to measured cue reliability in the language used in the literature) can be affected by cues' physical properties (e.g., size, distance, richness), individual utilization of cues, and instantaneous changes caused by experimental manipulations. The physical properties of cues determine their quality as spatial information sources. Individual utilization corresponds to navigators' long-term habits and preferences in using different spatial cues. This component should be very stable across time and situations and reflects one's pre-existing prior on cue usage. The third component is alterations of cue usage caused by experimental manipulations. In Experiment 3, we found that positive and negative distorted feedback affected cue reliability differently, corresponding to the third component of cue reliability. It is quite possible that different participants reacted differently to the distorted feedback. Therefore, we speculate that the second and the third components contribute to individual differences measured in cue reliability.

What are the relationships between different reliability components and cue weights? As illustrated in our first two experiments, the first component and the second component should directly correspond to cue weights, if navigators follow Bayesian principles. However, the third component of instantaneous changes in cue reliability is not necessarily transferred to changes in cue weights immediately and probably will show a hysteresis effect due to the pre-existing prior information in the second component. Experiment 3 illustrates such relationship.

As to the third component, instantaneous changes, it was surprising that distorted feedback actually changed measured cue reliability. And, as to the relationship between cue reliability and cue weighting, a modest magnitude of hysteresis effect was observed. These patterns of results are novel and intriguing. Future studies are needed to replicate and confirm our findings. Based on our results, several interesting questions arise concerning this component. For example, one can ask whether such instantaneous changes will be sustained after distorted feedback ends and whether these changes can be incorporated into the pre-existing prior reliability component as the manipulation continues. A sign of such incorporation might be regained consistency between measured cue reliability and actual cue weights. A third interesting question is whether other experimental manipulations can induce changes to cue reliability, such as extensive training, and whether the hysteresis effect will occur similarly.

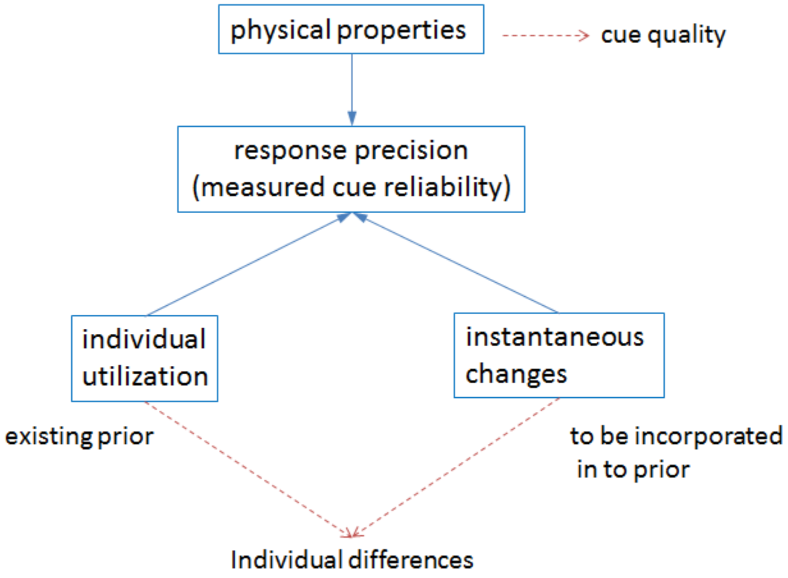


Figure 14. Composite nature of cue reliability. Response precision is equivalent to measure cued cue reliability, which consists of three elements, physical properties, individual utilization by different navigators, and instantaneous changes caused by experimental manipulations. Physical properties of a cue correspond to its quality. Individual utilization and instantaneous changes in cue usage together contribute to individual differences in measure cue reliability given the same experimental settings. Individual utilization represents navigators' habit of cue usage, which forms a strong prior. Instantaneous changes in measured cue reliability are temporal and need to be integrated into the individual utilization prior.

Cue Salience and Cue Reliability

The cue salience effect is a common phenomenon observed in the spatial navigation literature. Cue size, distance, and richness are three main factors affecting cue salience. A rather consistent finding is that the bigger, the closer, and the richer the spatial cue is, the greater weight the cue gains. More salient cues have greater competition capacity. For example, past studies have found that more salient cues usually win in the competition against less salient cues (Chamizo et al., 2006; Goodyear & Kamil, 2004; Gouteux, Thinus-Blanc, & Vauclair, 2001; Learmonth, Nadel, & Newcombe, 2002; Spetch, 1995). More salient cues led to better spatial learning (Chamizo et al., 2006). The cue salience effect also occurs in the adaptation paradigm, in which visual information is shifted intentionally by experimenters to conflict with path integration cues in a continuous manner. In the adaptation paradigm, navigators would adapt to the distorted visual feedback information over time. Through adaptation, navigators would gradually learn new relationships between visual information and self-motion information. For example, Bruggeman and his colleagues found that environments richer in visual cues resulted in quicker and greater adaptation (Bruggeman, Zosh, & Warren, 2007). In our first experiment, cue reliability was manipulated by providing more landmarks. This manipulation actually generated two effects: visual richness of the surrounding environment was enhanced, and it was more likely that there would be landmarks close to target locations. It remains relatively unclear, however, whether altering reliability of self-motion cues will lead to similar results. Path integration needs to be implemented continuously, so its error accumulates over time and distance. If errors are taken into consideration when cues are combined, self-motion cues would be assigned smaller weights relative to fixed visual cues as trajectory length increased. Experiment 1B is one example.

We speculate that cue salience corresponds to the first component of cue reliability (Figure 14); that is, cue reliability is influenced by physical properties of the cue. Size, distance, and richness are three main factors of physical properties. Bigger, closer, and richer visual cues have better quality and will result in higher levels of cue reliability. Appendix 3 includes a simple quantitative analysis, which suggests positive correlations between cue salience and cue reliability.

Weighting Cues by Cue Relative Reliability

In our first three experiments, which were conducted in immersive virtual reality environments, positive correlations were always found between cue reliability and cue weights, no matter whether they were equal at the group level. This Bayesian weighting-by-reliability strategy observed in our experiments suggests that it is as important to consider quality of cues as to consider different cue types. Cue competition studies usually focused on contrasts between different types of spatial cues. However, when observing that one particular spatial cue type dominates another spatial cue type, we cannot simply conclude that any cues belonging to the former category would win against any cues belonging to the latter category. The contrast could have been confounded by cue reliability. As stated in the previous discussion section, cue reliability can change if the cue's physical properties are varied, and cue reliability has been shown to be a critical factor determining cue weights in our experiments.

A typical example is the debate on whether an encapsulated module exclusively devoted to geometric cues exists in the brain (Cheng, 1986). The observation that geometric cues block/overshadow featural cues does not necessarily mean that such a module exists in the brain. As to the issue of whether there exists an encapsulated geometric module in spatial navigation, some studies found absolute control by geometric cues in guiding spatial behaviors, and some found that featural cues also influenced spatial behaviors but geometric-cues seemed to be more influential. However, these results were confounded by cue reliability and are not definitive evidence for an encapsulated module. It is quite possible that geometric cues gained greater control simply because they were more reliable in general, just as close landmarks often dominate far landmarks. Even though there might be other factors playing into the advantages associated with geometric cues, cue reliability certainly is an important one. Thus, it is critical to differentiate whether geometric cues and featural cues belong to different modules or they simply differ in the single dimension of cue reliability.

Many studies have shown that geometric cues, such as room shape, had absolute control or relatively larger control over featural cues, even though the former were ambiguous and led to incorrect searches, whereas the latter were unambiguous and would have led to correct searches (Cheng, 1986). This seems to be contradictory evidence against the cue reliability hypothesis. However, cue ambiguity is not equivalent to cue salience. For example, room shape can be more reliable than landmarks even if the former is more ambiguous than the latter. Suppose that the room shape is rectangular and there are no other cues. If rats learn the room shape, they would inevitably go to the correct corner and the rotationally-equivalent corner equally, resulting in a bimodal distribution of searches, if responses are recorded in a continuous manner. Instead of simply calculating response variability from the bimodal distribution to know about cue reliability of the room, we should combine the two sub-distributions of the bimodal distribution by fusing geometrically equivalent locations of the two corners. Such calculation of cue reliability of the room shape eliminates the influence of cue ambiguity. However, it would be interesting to examine how cue ambiguity and cue reliability affect navigational behaviors differently and interactively.

In addition, the Bayesian paradigm may be able to serve as a useful framework to study other factors influencing the cue interaction process after cue reliability has been accounted for. For example, in Experiment 3 we investigated the effect of explicit landmark stability in addition to the influence of cue reliability. One initial speculation was that landmark stability would affect cue weights beyond the influence of cue reliability as a prior. Results contradicted this speculation. Instead, landmark instability seemed to be incorporated into cue reliability, which in turn affected how people weighed the cues; that is, landmark instability affected cue weights with cue reliability as the mediator. In future studies, it will be worthwhile to compare explicit cue instability, which was examined in our experiment, with implicit cue instability. Another factor that might serve as a prior is pre-experimental training. Its effect might be mixed. Pre-existing interaction history with a cue might increase people's performance associated with the cue or simply prime people to use the cue regardless of whether their performance increases or remains constant.

Finally, it is possible that there are different rules underlying within-modality and inter-modality cue interaction (Hillis et al., 2002; Mou & Spetch, 2013). One obvious difference is that inter-modality cue integration usually needs to perform coordinate transformations, since different reference frames are probably employed by different sensory modalities (Deneve & Pouget, 2004). As described in Experiment 4, past studies found within-modality and inter-modality cue interaction followed different rules within the Bayesian framework. Another difference is that within-modality cue interaction seems to be more mandatory than inter-modality cue interaction, in that observers feel more obliged to fuse estimates associated with different cues into a coherent representation and probably lose access to individual estimates that feed into the coherent representation (Hillis et al., 2002). Finally, cues from the same modality might not result in optimal cue integration behaviors (Mou & Spetch, 2013). One possible reason for these differences is that correlations usually exist between cues within a modality and correlations correspond to redundant information, which would reduce the efficiency of combining cues (Oruç, Maloney, & Landy, 2003). Another possible reason is that interference is generally stronger between cues from a common modality so that navigators would be blind to some useful components of the blocked/overshadowed cue (Doeller & Burgess, 2008). A third possible reason is that, cues within a modality are more likely to create a holistic perception or representation which is greater than simple linear sum of its parts, for example when objects are arranged to form a regular structure (Garner & Felfoldy, 1970; Mou & Spetch, 2013).

Regression to the Mean

In Experiments 1AB, Experiment 2, and Experiment 3, positive correlations were always found between the actual weights and Bayesian predictions. However, the slope of the regression line (Figure 3c, 4c, 11c) was always less than 1, a very consistent pattern of results across experiments (Table 3 includes statistics). The means of the two indices were either equivalent or close. This represents a regression-to-the-mean phenomenon, which suggests the cue-weighting process might not be a perfect Bayesian relationship. For participants with similar visual reliability and self-motion reliability (points in the intermediate range in the scatter-plot), the actual weights tended to match the Bayesian weights; for those with substantially higher visual reliability than self-motion reliability (points to the right), the actual weights tended to be lower than the Bayesian predictions; for those with substantially lower visual reliability than self-motion reliability (points to the left), the actual weights tended to be higher than the Bayesian predictions.

In a typical Bayesian framework, as used in other domains of study, before the double-cue conditions, observers have already experienced a larger number of single-cue trials and probably have constructed correct representations of how noisy the cues are. Therefore, we speculate that the regression-to-the-mean phenomenon observed in our experiments might be due to the fact that participants were still acquiring information about how useful the cues actually were. At last one theory proposes that cue reliability is inherently encoded in the sensory population activity itself, so the acquisition of cue reliability representations is nearly automatic and does not require learning (Ma et al., 2006). However, considering that the environment in our experiments was rendered in a virtual reality setup, with which participants had little experience with, it is possible that forming accurate representations of cue reliability takes time. In our study, only 10 trials were included in each condition, and conditions were interleaved, which means participants needed to perform cue integration while still developing a sense of the quality of cues, especially in the early phase of the experiment. Given their unfamiliarity with the immersive virtual reality set-ups, participants probably assumed that visual cues and self-motions cues were similar in reliability at the very beginning of the experiment. As the experiment proceeded, their representations of cue usefulness started to approximate the true levels. For participants in the middle range of the scatter-plot, the initial reliability representations were close to the true levels, so the actual weights were close to the Bayesian weights. For participants at the two ends, the initial representations were biased, but they did not have enough trials to fully correct the representations, so the actual weights deviated from the Bayesian weights and were biased towards the initial assumptions. This hypothesis predicts that, with more trials, we would see a better fit between the actual weights and the Bayesian predictions across participants.

This hypothesis might also be able to explain the interference effects observed in Experiments 1AB. Participants experienced interference from the other reliability level and seemed to blend different reliability levels of the spatial cue. At the beginning of the experiment, they probably assumed equal reliability across all levels. As participants' interaction with the environment increased, the subjective evaluation of different reliability levels started to diverge. But participants probably did not have enough experiences to form accurate representations of different visual environments and different body rotation conditions. This hypothesis predicts that as the length of experiment increases, the interference effects between different reliability levels would gradually vanish. Further studies are needed to test this hypothesis.

Limitations

This project has several limitations. First, the experiment examining cue instability might not have had sufficient power to detect the underlying effects. To improve the experiment, cue instability should be manipulated more substantially and probably within-subjects instead of between-subjects. In addition the sample size should be increased. Second, when examining interactions between different visual cues in desktop virtual reality, we found that the visual-matching strategy was frequently adopted by participants. To discourage participants from adopting the visual-matching strategy, an immersive virtual environment might be used to compare different visual cues. One might need to develop a different paradigm if desktop virtual reality is required, for example in the case of collecting brain-imaging data. Third, participants' responses were 2-dimensional but we transformed the 2-dimensional data to 1-dimension by calculating the distance of the response point to the response centroid. For future studies with a larger number of trials, it may be possible to perform a more comprehensive analysis by calculating variances along the two dimensions separately and the covariance between them (Friedman, Ludvig, Legge, & Vuong, 2013).

CHAPTER VIII

CONCLUSIONS AND FUTURE DIRECTIONS

The project was designed to investigate how spatial cues are integrated in the process of human spatial navigation. To achieve this goal, we applied the Bayesian framework to our experiments. First, we replicated the main findings of some pioneering work in the domain of human spatial navigation, and showed that human navigators were able to integrate visual and self-motion information optimally in a Bayesian manner. We also extended previous studies to show that human navigators experienced interference when alternating between different reliability levels of the spatial information. Second, we achieved a better understanding of the concept of cue reliability by showing that cue reliability was highly flexible and could be changed by biased feedback. This study also sought to expand the Bayesian framework to incorporate the influence of cue instability in addition to cue reliability, and examined whether cue instability has its influence on cue-weighting directly or indirectly. When different spatial modalities were involved, human navigators consistently weighted spatial cues based on cue reliability. Finally, we modified the paradigm for 2D desktop display and found that interactions between different visual cues were different from interactions between visual cues and self-motion cues.

Our results provide insights into the question of how spatial cues interact with one another during the process of spatial navigation. First, we demonstrated that, as in other domains of investigation, human navigators were able to integrate spatial cues from different modalities optimally in a Bayesian manner. Second, the very consistent pattern of positive correlations between cue reliability and cue weights highlights the need to consider cue quality in addition to cue type when one examines the various roles of different spatial cues in the navigation process. Third, the wide range of cue relative reliability across subjects and the fact that measured cue reliability was affected systematically by experimental manipulations imply the flexible and strategic nature of cue reliability. These findings also suggest that measured cue reliability is multifaceted. Fourth, we propose that the Bayesian framework can be very useful in studying how factors affect the cue-weighting process; they can affect cue weights directly without changing cue reliability or indirectly by changing cue reliability.

This project also opens the door for investigating several new questions within the Bayesian framework. First, it remains unclear why interference occurred between different reliability levels in Experiments 1AB. One possible explanation attributes the interference to shared spatial context between reliability levels. The spatial context can consist of multiple elements. In Experiments 1AB, the elements might be the target locations, the visual environment, the ground texture, and the physical space the virtual environment is mapped to. Another possible explanation is that participants did not have sufficient experience with the experimental environment, which led to inaccurate and blended representations of different reliability levels. This explanation assumes that cue reliability needs to be learned, at least for virtual reality technologies, with which people usually have little prior experience. Future studies thus can manipulate commonalities between reliability levels or the duration of the experiment to distinguish these two possibilities.

A second new line of research could explore navigators' metacognitive abilities. Considering that measured cue reliability reflects navigators' subjective evaluations, it would be worthwhile to ask participants to report their confidence levels in their performance. It has been shown that people differ in their abilities of introspection (Barttfeld et al., 2013; Fleming, Weil, Nagy, Dolan, & Rees, 2010). It is possible that those who have accurate knowledge about their own performance would more strictly follow the Bayesian-weighting-by-reliability rule.

Third, when participants in our experiments attempted to integrate spatial cues, the cues were present at the same time. It would be interesting to examine whether people are able to integrate two spatial cues when one of them is not displayed simultaneously with the other one but instead is maintained in working memory. Sequential display of the cues could be used to study this question.

A fourth line of research might investigate whether explicit and implicit cue instability affect the cue-weighting process differently; the latter was not tested in our study. Implicit cue instability might be inherent in objects, such as animal models vs. plant models. Implicit instability might also be experimentally manipulated. For example, navigators can be exposed to unstable landmarks for a sufficient amount of time before the real test, when landmarks will remain stable. We hypothesize that implicit cue instability might affect cue weights directly without changing cue reliability, whereas explicit cue instability might achieve its effect indirectly by changing cue reliability.

Finally, when examining different visual cues in a reorientation paradigm on a 2D desktop virtual environment (as in Experiment 4), it is possible to eliminate the visual-matching strategy by dissociating representation of orientation from representation of location. For example, for orientation reproduction, people could be required to reproduce the remembered orientation at a different location, whereas for location reproduction, people could be required to navigate to the correct location from a fixed orientation.

In summary, the results of this dissertation advance the scientific understanding of how spatial cues interact during the process of spatial navigation. Our findings help to answer key questions about the nature of human spatial navigation and also raise important new questions for future research.

REFERENCES

Barttfeld, P., Wicker, B., McAleer, P., Belin, P., Cojan, Y., Graziano, M., . . . Sigman, M. (2013). Distinct patterns of functional brain connectivity correlate with objective performance and subjective beliefs. *Proceedings of the National Academy of Sciences, 110*(28), 11577-11582.

Berger, D. R., & Bülthoff, H. H. (2009). The role of attention on the integration of visual and inertial cues. *Experimental Brain Research, 198*(2-3), 287-300.

Biegler, R. (1996). Landmark stability: Further studies pointing to a role in spatial learning. *The Quarterly Journal of Experimental Psychology: Section B, 49*(4), 307-345.

Biegler, R., & Morris, R. G. M. (1993). Landmark stability is a prerequisite for spatial but not discrimination learning. *Nature, 361*, 631-633.

Biegler, R., & Morris, R. G. M. (1996). Landmark stability: studies exploring whether the perceived stability of the environment influences spatial representation. *Journal of Experimental Biology, 199*(1), 187-193.

Bruggeman, H., Zosh, W., & Warren, W. H. (2007). Optic flow drives human visuo-locomotor adaptation. *Current Biology, 17*(23), 2035-2040.

Burgess, N., Spiers, H. J., & Paleologou, E. (2004). Orientational manoeuvres in the dark: dissociating allocentric and egocentric influences on spatial memory. *Cognition, 94*(2), 149-166. doi: http://dx.doi.org/10.1016/j.cognition.2004.01.001

Butler, J. S., Smith, S. T., Campos, J. L., & Bülthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *J Vis, 10*(11), 23.

Byrne, P. A., & Crawford, J. D. (2010). Cue reliability and a landmark stability heuristic determine relative weighting between egocentric and allocentric visual information in memory-guided reach. *J Neurophysiol, 103*(6), 3054-3069. doi: 10.1152/jn.01008.2009

Castro, C. A., & Rudy, J. W. (1987). Early-life malnutrition selectively retards the development of distal-cue but not proximal-cue navigation. *Developmental Psychobiology, 20*(5), 521-537. doi: 10.1002/dev.420200506

Chamizo, V., Rodrigo, T., Peris, J., & Grau, M. (2006). The influence of landmark salience in a navigation task: An additive effect between its components. *Journal of Experimental Psychology: Animal Behavior Processes, 32*(3), 339.

Cheng. (1986). A purely geometric module in the rat's spatial representation. *Cognition, 23*(2), 149-178.

Cheng, & Newcombe, N. S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychonomic Bulletin & Review, 12*(1), 1-23.

Cheng, Shettleworth, S. J., Huttenlocher, J., & Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychol Bull, 133*(4), 625-637. doi: 10.1037/0033-2909.133.4.625

Cheng, & Spetch, M. L. (1995). Stimulus control in the use of landmarks by pigeons in a touch-screen task. *J Exp Anal Behav, 63*(2), 187-201.

Chittka, L., & Geiger, K. (1995). HONEYBEE LONG-DISTANCE ORIENTATION IN A CONTROLLED ENVIRONMENT. *Ethology, 99*(2), 117-126.

Cressant, A., Muller, R. U., & Poucet, B. (1997). Failure of centrally placed objects to control the firing fields of hippocampal place cells. *Journal of Neuroscience, 17*(7), 2531-2542.

Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology-Paris, 98*(1–3), 249-258. doi: http://dx.doi.org/10.1016/j.jphysparis.2004.03.011

Doeller, C. F., & Burgess, N. (2008). Distinct error-correcting and incidental learning of location relative to landmarks and boundaries. *Proc Natl Acad Sci U S A, 105*(15), 5909-5914. doi: 10.1073/pnas.0711433105

Doeller, C. F., King, J. A., & Burgess, N. (2008). Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences, 105*(15), 5915-5920.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429-433. doi: 10.1038/415429a

Ernst, M. O., Banks, M. S., & Bulthoff, H. H. (2000). Touch can change visual slant perception. *Nature neuroscience, 3*(1), 69-73. doi: 10.1038/71140

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*(4), 162-169. doi: 10.1016/j.tics.2004.02.002

Etienne, A. S., & Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus, 14*(2), 180-192.

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). . *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.

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science, 329*(5998), 1541-1543.

Friedman, A., Ludvig, E., Legge, E. G., & Vuong, Q. (2013). Bayesian combination of two-dimensional location estimates. *Behavior Research Methods, 45*(1), 98-107. doi: 10.3758/s13428-012-0241-x

Gallistel, C. R. (1990). *The organization of learning*: The MIT Press.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cogn Psychol, 1*(3), 225-241. doi: http://dx.doi.org/10.1016/0010-0285(70)90016-2

Goodyear, A. J., & Kamil, A. C. (2004). Clark's nutcrackers (Nucifraga columbiana) and the effects of goal--landmark distance on overshadowing. *J Comp Psychol, 118*(3), 258-264. doi: 10.1037/0735-7036.118.3.258

Gouteux, S., Thinus-Blanc, C., & Vauclair, J. (2001). Rhesus monkeys use geometric and nongeometric information during a reorientation task. *Journal of Experimental Psychology: General, 130*(3), 505.

Gu, Y., Angelaki, D. E., & DeAngelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature neuroscience, 11*(10), 1201-1210.

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179*(4), 595-606.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science, 298*(5598), 1627-1630. doi: 10.1126/science.1075396

Kamin, L. J. (1967). Predictability, surprise, attention, and conditioning.

Kearns, M. J., Warren, W. H., Duchon, A. P., & Tarr, M. J. (2002). Path integration from optic flow and body senses in a homing task. *Perception, 31*(3), 349-374. doi: 10.1068/p3311

Kelly, J. W., McNamara, T. P., Bodenheimer, B., Carr, T. H., & Rieser, J. J. (2008). The shape of human navigation: How environmental geometry is used in maintenance of spatial orientation. *Cognition, 109*(2), 281-286. doi: 10.1016/j.cognition.2008.09.001

Kelly, J. W., McNamara, T. P., Bodenheimer, B., Carr, T. H., & Rieser, J. J. (2009). Individual differences in using geometric and featural cues to maintain spatial orientation: Cue quantity and cue ambiguity are more important than cue type. *Psychon Bull Rev, 16*(1), 176-181. doi: 10.3758/PBR.16.1.176

Knierim, J. J. (2002). Dynamic interactions between local surface cues, distal landmarks, and intrinsic circuitry in hippocampal place cells. *The Journal of neuroscience, 22*(14), 6254-6264.

Knierim, J. J., Kudrimoti, H. S., & McNaughton, B. L. (1995). Place cells, head direction cells, and the learning of landmark stability. *The Journal of neuroscience, 15*(3), 1648-1659.

Learmonth, A. E., Nadel, L., & Newcombe, N. S. (2002). Children's use of landmarks: Implications for modularity theory. *Psychol Sci, 13*(4), 337-341.

Lenck-Santini, P. P., Muller, R. U., Save, E., & Poucet, B. (2002). Relationships between place cell firing fields and navigational decisions by rats. *Journal of Neuroscience, 22*(20), 9035-9047.

Lew, A. R. (2011). Looking beyond the boundaries: Time to put landmarks back on the cognitive map? *Psychol Bull, 137*(3), 484-507. doi: 10.1037/a0022315

Lew, A. R., Gibbons, B., Murphy, C., & Bremner, J. G. (2010). Use of geometry for spatial reorientation in children applies only to symmetric spaces. *Dev Sci, 13*(3), 490-498. doi: 10.1111/j.1467-7687.2009.00904.x

Loomis, J. M., Klatzky, R. L., Golledge, R. G., Cicinelli, J. G., Pellegrino, J. W., & Fry, P. A. (1993). Nonvisual navigation by blind and sighted: assessment of path integration ability. *Journal of Experimental Psychology: General, 122*(1), 73.

Loomis, J. M., Klatzky, R. L., Golledge, R. G., & Philbeck, J. W. (1999). Human navigation by path integration. *Wayfinding behavior: Cognitive mapping and other spatial processes*, 125-151.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature neuroscience, 9*(11), 1432-1438.

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.

Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research, 184*(1), 39-52.

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proc Natl Acad Sci U S A, 107*(39), 17041-17046. doi: 10.1073/pnas.1001699107

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Curr Biol, 18*(9), 689-693. doi: 10.1016/j.cub.2008.04.021

Neunuebel, J. P., Yoganarasimha, D., Rao, G., & Knierim, J. J. (2013). Conflicts between Local and Global Spatial Frameworks Dissociate Neural Representations of the Lateral and Medial Entorhinal Cortex (May, pg 9246, 2013). *Journal of Neuroscience, 33*(32), 13249-13249. doi: 10.1523/jneurosci.2897-13.2013

Nico, D., Israel, I., & Berthoz, A. (2002). Interaction of visual and idiothetic information in a path completion task. *Experimental Brain Research, 146*(3), 379-382. doi: 10.1007/s00221-002-1184-8

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Res, 43*(23), 2451-2468.

Parise, C. V., Spence, C., & Ernst, M. O. (2012). When correlation implies causation in multisensory integration. *Current Biology, 22*(1), 46-49.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. Translated and edited by GV Anrep*: London: Oxford University Press.

Roberts, A. D., & Pearce, J. M. (1999). Blocking in the Morris swimming pool. *J Exp Psychol Anim Behav Process, 25*(2), 225-235.

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.

Save, E., & Poucet, B. (2000). Involvement of the hippocampus and associative parietal cortex in the use of proximal and distal landmarks for navigation. *Behavioural Brain Research, 109*(2), 195-206. doi: 10.1016/s0166-4328(99)00173-4

Shettleworth, S. J., & Sutton, J. E. (2005). Multiple systems for spatial learning: dead reckoning and beacon homing in rats. *J Exp Psychol Anim Behav Process, 31*(2), 125-141. doi: 10.1037/0097-7403.31.2.125

Spetch, M. L. (1995). Overshadowing in landmark learning: touch-screen studies with pigeons and humans. *J Exp Psychol Anim Behav Process, 21*(2), 166-181.

Spetch, M. L., & Mondloch, M. V. (1993). Control of pigeons spatial search by graphic landmarks in a touch-screen task. *Journal of Experimental Psychology-Animal Behavior Processes, 19*(4), 353-372. doi: 10.1037/0097-7403.19.4.353

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral Cortex, 17*(3), 679-690.

Twyman, A. D., & Newcombe, N. S. (2010). Five reasons to doubt the existence of a geometric module. *Cogn Sci, 34*(7), 1315-1356. doi: 10.1111/j.1551-6709.2009.01081.x

Wilson, P. N., & Alexander, T. (2008). Blocking of spatial learning between enclosure geometry and a local landmark. *Journal of Experimental Psychology-Learning Memory and Cognition, 34*(6), 1369-1376. doi: 10.1037/a0013011

Zhao, M., & Warren, W. H. (2010, November). Path integration and visual landmarks: Optimal combination or multiple systems? . *Annual Meeting of Psychonomics Society, St. Louis, MO.*

Table 1. Bayes factors for comparisons between double-cue conditions and model predictions. Probability values (p) of t-tests in parentheses.

	models	**condition**	**rich**	**poor**	**Pooled**
Exp. 1A	Bayesian integration	combination	4.92 (.614)	0.65 (.037)	2.07 (.157)
		conflict	2.65 (.220)	5.31 (.748)	5.28 (.735)
	Bayesian alternation	combination	0.02 (.001)	1.09 (.069)	0.08 (.003)
		conflict	0.03 (.001)	0.00 (.000)	0.00 (.000)
	models	**condition**	**no rotation**	**rotation**	**Pooled**
Exp. 1B	Bayesian integration	combination	0.22 (.011)	5.59 (.989)	1.12 (.072)
		conflict	4.53 (.515)	2.32 (.184)	3.35 (.309)
	Bayesian alternation	combination	0.40 (.021)	0.00 (.000)	0.00 (.000)
		conflict	0.00 (.000)	0.03 (.001)	0.01 (.000)
	models	**condition**	**vision positive**	**motion positive**	
Exp. 2	Bayesian integration	combination	5.56 (.922)	0.58 (.032)	
		conflict	5.58 (.970)	0.39 (.021)	
	Bayesian alternation	combination	0.06 (.002)	0.00 (.000)	
		conflict	0.01 (.000)	0.00 (.000)	
	models	**condition**	**unstable**	**stable**	
Exp. 3	Bayesian integration	combination	0.03 (.001)	0.15 (.006)	
		conflict	0.06 (.002)	2.65 (.199)	
	Bayesian alternation	combination	0.51 (.026)	0.32 (.015)	
		conflict	1.42 (.088)	0.42 (.021)	
	models	**condition**	**group1**	**group2**	**group1&2**
Exp. 4	Bayesian integration	combination	0.297 (.137)	0.063 (.002)	0.004 (.000)
		conflict	0.069 (.003)	0.000 (.000)	0.000 (.000)
	Bayesian alternation	combination	0.013 (.000)	0.369 (.020)	0.001 (.000)
		conflict	0.456 (.022)	4.678 (.523)	1.057 (.043)

Table 2. Pearson correlations between response relative proximity and cue relative reliability. Probability values (p) in parentheses.

Exp. 1A	**condition**	**rich**	**poor**	**pooled**
	combination	0.38 (.122)	0.41 (.091)	0.51 (.029)
	conflict	0.72 (.001)	0.67 (.002)	0.80 (.000)
	averaged	0.69 (.002)	0.59 (.011)	0.77 (.000)
Exp. 1B	**condition**	**no rotation**	**rotation**	**pooled**
	combination	0.64 (.004)	0.62 (.006)	0.47 (.050)
	conflict	0.66 (.003)	0.55 (.017)	0.66 (.003)
	averaged	0.72 (.001)	0.61 (.008)	0.58 (.012)
Exp. 2	**condition**	**vision positive**	**motion positive**	
	combination	0.590 (.010)	0.508 (.032)	
	conflict	0.505 (.032)	0.235 (.348)	
	averaged	0.611 (.007)	0.402 (.098)	
Exp. 3	**condition**	**unstable**	**stable**	
	combination	.327 (.147)	.564 (.008)	
	conflict	.551 (.010)	.483 (.027)	
	averaged	.478 (.029)	.594 (.004)	
Exp. 4	**condition**	**group1**	**group2**	**group1&2**
	combination	0.210 (.360)	- 0.044 (.858)	0.109 (.502)
	conflict	0.022 (.924)	- 0.036 (.883)	0.073 (.653)
	averaged	0.217 (.344)	- 0.051 (.836)	0.119 (.463)

Table 3. Slopes of the linear regressions of response relative proximity on cue relative reliability. T- tests compare the slopes to 1.

Experiment	condition/group	slope	t	P
Exp. 1A	rich	0.413	5.219	< .001
	poor	0.367	5.000	< .001
	pooled	0.664	2.462	< .05
Exp. 1B	no rotation	0.463	4.837	< .001
	rotation	0.558	2.419	< .05
	pooled	0.495	2.907	< .02
Exp. 2	vision positive	0.421	4.250	< .001
	motion positive	0.370	3.035	< .01
Exp. 3	unstable	0.394	3.645	< .002
	stable	0.361	5.697	< .001

APPENDIX

Appendix 1

Mathematical Demonstration

Suppose there are two noise levels of a cue, rich and poor. v_1 and m_1 represent the vision and self-motion conditions in the rich cue condition, and v_2 and m_2 represent the vision and self-motion conditions in poor cue condition. c_1 represents combination condition derived from v_1 and m_1, and c_2 represents combination condition derived from v_2 and m_2. If people are optimal integrator at both noise levels, response variability of combination condition will be,

$$\sigma_{c_1}^2 = \frac{\sigma_{v_1}^2 \times \sigma_{m_1}^2}{\sigma_{v_1}^2 + \sigma_{m_1}^2} \qquad \text{-- (1)}$$

$$\sigma_{c_2}^2 = \frac{\sigma_{v_2}^2 \times \sigma_{m_2}^2}{\sigma_{v_2}^2 + \sigma_{m_2}^2} \qquad \text{-- (2)}$$

We then collapse c1 and c2. This is equivalent to a mixture of c1 and c2 distributions (assume c1 and c2 are independent of each other). Since there are equal number of responses in c1 and c2, the mixture ratio is ½. The collapsed variance is expressed as,

$$\sigma_{c_{1,2}}^2 = \frac{1}{2} \times \sigma_{c_1}^2 + \frac{1}{2} \times \sigma_{c_2}^2 \qquad \text{-- (3)}$$

Replace and in equation (3) with equations (1) and (2), and we get,

$$\sigma_{c_{1,2}}^2 = \frac{\sigma_{v_1}^2 \sigma_{m_1}^2 \left(\sigma_{v_2}^2 + \sigma_{m_2}^2\right) + \sigma_{v_2}^2 \sigma_{m_2}^2 \left(\sigma_{v_1}^2 + \sigma_{m_1}^2\right)}{2 \times (\sigma_{v_1}^2 + \sigma_{m_1}^2) \times (\sigma_{v_2}^2 + \sigma_{m_2}^2)} \qquad \text{-- (4)}$$

We also collapse v1 and v2, s1 and s2, assuming they are independent of each other. Since there are equal numbers of responses in v1 and v2, s1 and s2, the mixture ratios are also ½.

$$\sigma_{v_{1,2}}^2 = \frac{1}{2} \times \sigma_{v_1}^2 + \frac{1}{2} \times \sigma_{v_2}^2 \qquad \text{-- (5)}$$

$$\sigma_{m_{1,2}}^2 = \frac{1}{2}\times\sigma_{m_1}^2 + \frac{1}{2}\times\sigma_{m_2}^2 \qquad \text{-- (6)}$$

In order to get the Bayesian integration prediction for the collapsed combination distribution, we apply the Bayesian equations to collapsed vision distribution and collapsed self-motion distribution.

$$\sigma_{c_{pred}}^2 = \frac{\sigma_{v_{1,2}}^2\times\sigma_{m_{1,2}}^2}{\sigma_{v_{1,2}}^2+\sigma_{m_{1,2}}^2} \qquad \text{-- (7)}$$

We then replace $\sigma_{v_{1,2}}^2$ and $\sigma_{m_{1,2}}^2$ in equation (7) with equations (5) and (6),

$$\sigma_{c_{pred}}^2 = \frac{(\sigma_{v_1}^2+\sigma_{v_2}^2)\times(\sigma_{m_1}^2+\sigma_{m_2}^2)}{2\times(\sigma_{v_1}^2+\sigma_{v_2}^2+\sigma_{m_1}^2+\sigma_{m_2}^2)} \qquad \text{-- (8)}$$

Our hypothesis is that if people are optimal at both noise levels, when responses are collapsed, the actual variance of the combination distribution would be smaller than what we predict from the two collapsed single-cue distributions. This is equivalent to demonstrating,

$$\sigma_{c_{1,2}}^2 < \sigma_{c_{pred}}^2 \qquad \text{-- (9)}$$

Replacing $\sigma_{c_{1,2}}^2$ and $\sigma_{c_{pred}}^2$ in equation (9) with equations (4) and (8), we actually need to demonstrate,

$$\frac{\sigma_{v_1}^2\sigma_{m_1}^2\left(\sigma_{v_2}^2+\sigma_{m_2}^2\right)+\sigma_{v_2}^2\sigma_{m_2}^2\left(\sigma_{v_1}^2+\sigma_{m_1}^2\right)}{2\times(\sigma_{v_1}^2+\sigma_{m_1}^2)\times(\sigma_{v_2}^2+\sigma_{m_2}^2)} < \frac{(\sigma_{v_1}^2+\sigma_{v_2}^2)\times(\sigma_{m_1}^2+\sigma_{m_2}^2)}{2\times(\sigma_{v_1}^2+\sigma_{v_2}^2+\sigma_{m_1}^2+\sigma_{m_2}^2)} \qquad \text{-- (10)}$$

To expand equation (10), without changing the sign of the inequality, we multiply the numerator on the left with the denominator on the right, and also multiply the numerator on the left with the denominator on the left. After cancelling out items shared by the two sides. Equation (10) is finally reduced to,

$$2\times\sigma_{v_1}^2\sigma_{v_2}^2\sigma_{m_1}^2\sigma_{m_2}^2 < (\sigma_{v_1}^2)^2(\sigma_{m_2}^2)^2 + (\sigma_{v_2}^2)^2(\sigma_{m_1}^2)^2 \quad \text{-- (11)}$$

Let a stand for $\sigma_{v_1}^2\sigma_{m_2}^2$, and b stand for $\sigma_{v_2}^2\sigma_{m_1}^2$, we can simplify equation (11) as,

$$0 < (a - b)^2 \quad \text{-- (12)}$$

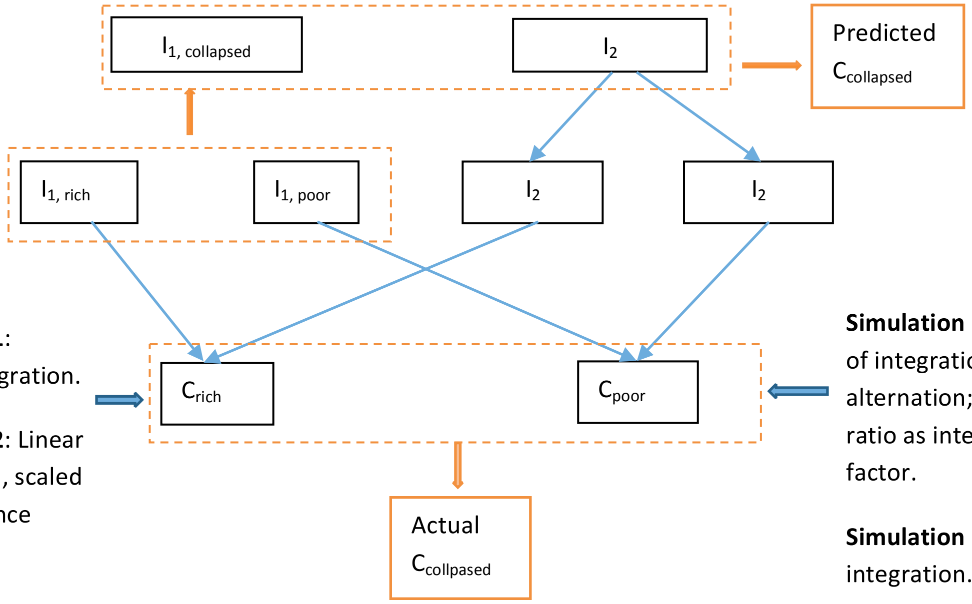
It is quite obvious now that, as long as a is not equal to b, equation (12) holds true. What does it mean when a equals to b? It means $\sigma_{v_1}^2\sigma_{m_2}^2 = \sigma_{v_2}^2\sigma_{m_1}^2$, which can be transformed to,

$$\frac{\sigma_{v_1}^2+\sigma_{m_1}^2}{\sigma_{m_1}^2} = \frac{\sigma_{v_2}^2+\sigma_{m_2}^2}{\sigma_{m_2}^2} \quad \text{-- (13)}$$

The left part is equal to the inverse of cue relative reliability at noise level 1, and the right part is equal to the inverse of cue relative reliability at noise level 2. Together, the entire mathematical demonstration shows that, as long as the cue relative reliability is not equal between the two noise levels, the actual variance of collapsed combination distribution would be smaller than predicted from Bayesian integration theory, leading to supra-optimal performance at the collapsed level. When the two noise levels have the same cue relative reliability, people can be optimal both at individual noise levels and at the overall level. In our experiments, the manipulation of visual cues and self-motion cues was significant and cue reliability differed between noise levels of the manipulated cue. Therefore, the former case fits our situation.

Simulations of Interference between Noise Levels

The diagram shows the procedure of the simulations. First, we assume there are two independent sources of spatial information, I_1 and I_2. I1 is manipulated in reliability, so I_1, rich stands for the more reliable information and I_1, poor stands for the less reliable information. I_2 is not manipulated. For rich and poor conditions, I_1 and I_2 are combined in various ways to generate combined-cue distributions C_{rich} and C_{poor} respectively. To get the actual collapsed variance $C_{collpased}$, we collapsed C_{rich} and C_{poor}. To get the predicted collapsed variance by Bayesian integration theory, we first collapsed $I_{1,rich}$ and $I_{1,poor}$, then apply Bayesian equations to I_{1}, collapsed and I_{2} and get predicted variance for combined-cue distribution. To measure prediction error, we take the ratio of predicted $C_{collapsed}$ over actual $C_{collapsed}$. Prediction error less than 1 means sub-optimal integration; greater than 1 means supra-optimal integration; equal to 1 means optimal integration.



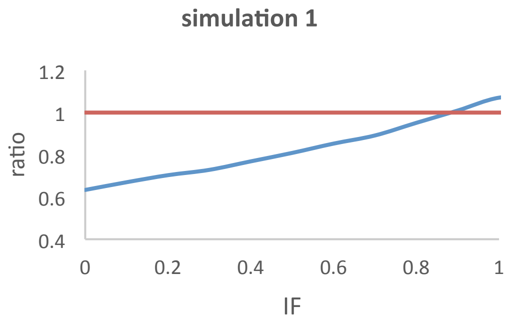
Simulation 1: optimal integration.

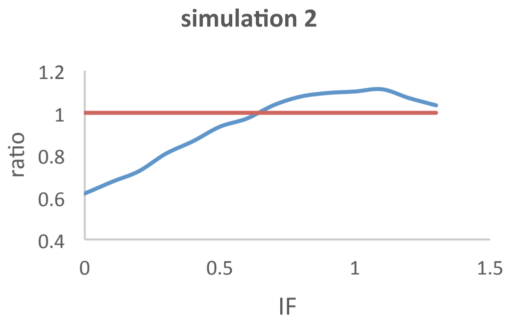
Simulation 2: Linear combination, scaled by interference factor.

Simulation 1: mixture of integration and alternation; mixture ratio as interference factor.

Simulation 2: optimal integration.

In the first simulation, for rich condition, two cues are integrated optimally; for poor condition, there is a mixture of optimal cue integration and cue alternation, and the mixture ratio corresponds to the interference factor (IF). In the second simulation, for rich condition, two cues are integrated linearly, but with weights that could deviate from Bayesian predictions. This deviation is expressed as the interference factor (IF), which scales the weights; IF equal to 1 means no deviation and optimal integration; IF different from 1 means deviation from Bayesian integration rules. For the poor condition, two cues are integrated optimally. Conceptually, the first simulation corresponds to our experiment 1 and the second conceptually corresponds to our experiment 2. We then plot prediction error against IF as in figures below.





Blue curve represents prediction error. Red line represents optimal integration. IF of 1 means no interference exists and cues are integrated optimally at both individual noise levels. One observation is that optimal integration at the collapsed level does not happen when IF equals to 1. Rather, it happens when IF deviates from 1. However, sub-optimal integration happens sometimes when IF deviates too much from 1. When IF is equal to 1, supra-optimal integration occurs. These simulations suggest people have to compromise to some degree at individual noise levels in order to be just optimal at the collapsed level. However, they also have to choose the correct parameters, since too much compromise would lead to sub-optimal integration at the collapsed level.

Appendix 2

In the pioneering study conducted by Nardini and his colleagues (Nardini, Jones, Bedford, & Braddick, 2008), only 4 trials were included in each condition. We speculate that having only 4 trials for each condition and calculating variance based on the 4 trials might be problematic. The problem could be either making data too noisy or even introducing some intrinsic biases, the latter of which might not be corrected by simply increasing the number of subjects. So we conducted the following simulations to examine how the number of observations per subject and the sample size would influence the observed data, and also how the way we frame the question would affect the results.

There are 4 simulations depending on different assumptions and the model fitting methods. The primary assumption is whether we consider subjects to come from a single population or different populations. Coming from different populations means each subject has his/her own response variances in the two single-cue conditions (vision and self-motion conditions). The opposite is that all subjects share common variances in the two single-cue conditions. There are two model fitting strategies, one is to model the data as though the subjects come from a single population and share a constant landmark weight, and the other is to model the data as though the subjects come from different populations and set their own landmark weights. A factorial combination of the two variables gives us 4 different simulations. It seems reasonable to predict that if the assumption mismatches the modeling strategy, some intrinsic bias might be observed.

Assumption 1: Subjects from a Single Population

For responses in the landmark and self-motion conditions, data points were sampled randomly from two normal distributions with means of 0 and variances equal to the mean variances obtained from my pilot study, $N(0, \sigma^2_{lm})$, $N(0, \sigma^2_{sm})$, $\sigma^2_{lm} = .42$, $\sigma^2_{sm} = 1.17$. Then, at the population level, I derived the variance of the combined-cue condition based on the Bayesian principles, $\sigma^2_{com} = \sigma^2_{conf} = \sigma^2_{lm} * \sigma^2_{sm} / (\sigma^2_{lm} + \sigma^2_{sm})$. For the combination condition, since the two cues are consistent, the distribution should be unbiased with a mean of 0, $N(0, \sigma^2_{comb})$, $\sigma^2_{comb} = 0.31$. For the conflict condition, there is a conflict and its mean should be, μ=conflict *

$\sigma^2_{sm}/(\sigma^2_{lm}+\sigma^2_{sm})+0*\sigma^2_{lm}/(\sigma^2_{lm}+\sigma^2_{sm})=$ conflict $*\ \sigma^2_{sm}/(\sigma^2_{lm}+\sigma^2_{sm})$. And here I set the conflict to be the same amount as in the pilot study, conflict=.46. The following is a summary of the response distributions for the 4 conditions: $r_{lm} \sim N(0, .42)$, $r_{sm} \sim N(0,1.17)$, $r_{com} \sim N(0,.31)$, $r_{conf} \sim N(.34, .31)$.

Assumption 2: Subjects from Different Populations

For the two single-cue conditions, different subjects have their own different response variances. For *i*th subject, the response in the landmark condition conforms to $N(0, \sigma^2_{i,lm})$, where $\sigma_{i,lm}$ itself was randomly sampled from a uniform distribution [0.13, 1.43]to create individual differences in response variability, and the response in self-motion condition conforms to $N(0, \sigma^2_{i,sm})$, where $\sigma_{i,sm}$ itself was randomly sampled from a uniforml distribution [0.216, 2.376] to create individual differences in response variability. The normal distributions of the variances of individuals' responses in the two single-cue conditions distributions were set arbitrarily. Then I calculated the variances and means for the combination and conflict conditions based on Bayesian principles for each subject. For responses in combined-cue conditions, $\sigma^2_{i,com}=\sigma^2_{i,conf}=\sigma^2_{i,lm}*\sigma^2_{i,sm}/(\sigma^2_{i,lm}+\sigma^2_{i,sm})$. The mean of conflict condition is $\mu_i=$conflict $*\ \sigma^2_{i,sm}/(\sigma^2_{i,lm}+\sigma^2_{i,sm})$, in which conflict=.46.

We randomly sampled data points from these population distributions, and estimated the variance of the combined-cue condition and the landmark relative proximity using the sampled data. To test whether the modeling works well, we compared the estimated values to the actual values we get from the sampled data.

Modeling Strategy 1: Subjects Use a Constant Landmark Weight

If we assume that subjects use a single constant landmark weight, we conduct the following steps: **1**, for the *i*th subject, calculate the actual variances for the 4 conditions ($S^2_{i,lm}$, $S^2_{i,sm}$, $S^2_{i,comb}$, $S^2_{i,conf}$) and the landmark relative proximity ($prox_{i,lm}$) in the conflict condition, using the sampled data; **2**, estimate the population variances for the two single-cue conditions and the two combined-cue conditions by averaging across subjects the individual variances in these conditions, $S^2_{lm}=mean(S^2_{i,lm})$, $S^2_{sm}=mean(S^2_{i,sm})$, $S^2_{comb}=mean(S^2_{i,comb})$, $S^2_{conf}=mean(S^2_{i,conf})$, and also estimate the population landmark relative proximity by averaging, $prox_{lm}=mean(prox_{i,lm})$; **3**, based on Bayesian equations, use the variance estimators from step 2 to calculate the predicted variance of the combined-cue condition, $\hat{S}^2_{comb}= S^2_{lm} * S^2_{sm}/(S^2_{lm}+ S^2_{sm})$, and the landmark relative proximity in the conflict condition, $rr_{lm}= S^2_{sm}/(S^2_{lm}+ S^2_{sm})$; **4**, compare the estimated variance, $\hat{S}^2_{comb}$, and landmark relative proximity, rr_{lm}, from step 3 to the actual variance of combined-cue condition, S^2_{comb}, and the actual landmark relative proximity from step 2, $prox_{lm}$, by taking the ratio of the predicted value over the actual value, $ratio1= \hat{S}^2_{comb} / S^2_{comb}$, $ratio2= rr_{lm}/ prox_{lm}$. **5**, this “bias ratio” is then plotted as a function of number of trials per condition and sample size.

Modeling Strategy 2: Subjects Establish Their Own Weights

If we assume subjects set their own weights, we conduct the following steps: **1**, for the *i*th subject, calculate the variances for the 4 conditions ($S^2_{i,lm}$, $S^2_{i,sm}$, $S^2_{i,comb}$, $S^2_{i,conf}$) and the landmark relative proximity ($prox_{i,lm}$) in the conflict condition, using the sampled data; **2**, for each subject, based on Bayesian equations, calculate the prediction of the population variance of the combined-cue condition and the prediction on the landmark relative proximity in the conflict condition, using the sampled data, $\hat{S}^2_{i,comb} = S^2_{i,lm} * S^2_{i,sm}/(S^2_{i,lm}+ S^2_{i,sm})$, $rr_{i,lm}= S^2_{i,sm}/(S^2_{i,lm}+ S^2_{i,sm})$; **3**, take the means of the predicted combined-cue variance and the predicted landmark relative proximity, $\hat{S}^2_{comb} =mean(\hat{S}^2_{i,comb})$, $rr_{lm} =mean(rr_{i,lm})$; **4**, compare the predicted values from step 3 to the means of the actual values calculated from step 1, $ratio1= \hat{S}^2_{comb} /S^2_{comb}$, $ratio2= rr_{lm} /prox_{lm}$, where $S^2_{comb}=mean(S^2_{i,comb})$ and $prox_{lm}=mean(prox_{i,lm})$. **5**, The bias ratio is then plotted as a function of number of trials per condition and sample size.

Algorithmically, the two modeling strategies are very similar. The only difference lies in the order of step 2 and step 3. In strategy 1, we first take the means of the variances and then apply the Bayesian equations to the means to get the predicted values, whereas in strategy 2, we first apply the Bayesian equations to individual subjects’ variances and then take the means of the predicted values. Mathematically, however, these strategies produce different results.

Results

The Figures below depict the bias ratio1 and bias ratio2 as a function of number of trials and number of subjects. The horizontal axis represents number of trials, and different colored lines represent different number of subjects. The horizontal dark thin line represents bias=1, which means no bias and the predicted value matches the actual value. Figures 1a & 2a show results for simulation 1, in which we assume subjects come from a single population and model as if they share a constant weight (assumption 1 & modeling strategy 1); Figures 1b & 2b show results for simulation 2, in which we assume subjects come from a single population, and model as if they set their own weights (assumption 1 & modeling strategy 2); Figures 1c & 2c show results for simulation 3, in which we assume subjects come from different populations, and model as if they share a constant weight (assumption 2 & modeling strategy 1); Figures 1d & 2d show results for simulation 4, in which we assume subjects come from different populations, and model as if they set their own weights (assumption 2 & modeling strategy 2). Note that in Figure 1c, the plot is on a different vertical scale; this is because the bias here is relatively large compared to other simulations

As described previously, the bias ratio is equal to the predicted value divided by the actual value. A distortion ratio of 1 means there is no bias. A ratio greater than 1 means the predicted value is greater than the actual value (i.e., positive bias in estimation) and a ratio less than 1 means the predicted value is less than the actual value (i.e., negative bias in estimation). Bias ratios are calculated for both the variance and the landmark relative weight. As for the variance, the ratio is equal to the predicted variance of the combined-cue condition derived from Bayesian equations divided by the actual variance of the combined-cue condition. As for the landmark relative weight, the ratio is equal to the landmark relative reliability derived from Bayesian equations divided by the landmark relative proximity observed in the conflict condition.

With only 4 trials per condition, the degree of distortion is very small in simulation 1, but the distortion is modest in all other 3 simulations. Such distortion does not seem to disappear with increasing number of subjects. As the number of trials increases, the distortion fades away quickly, except for simulation 3.

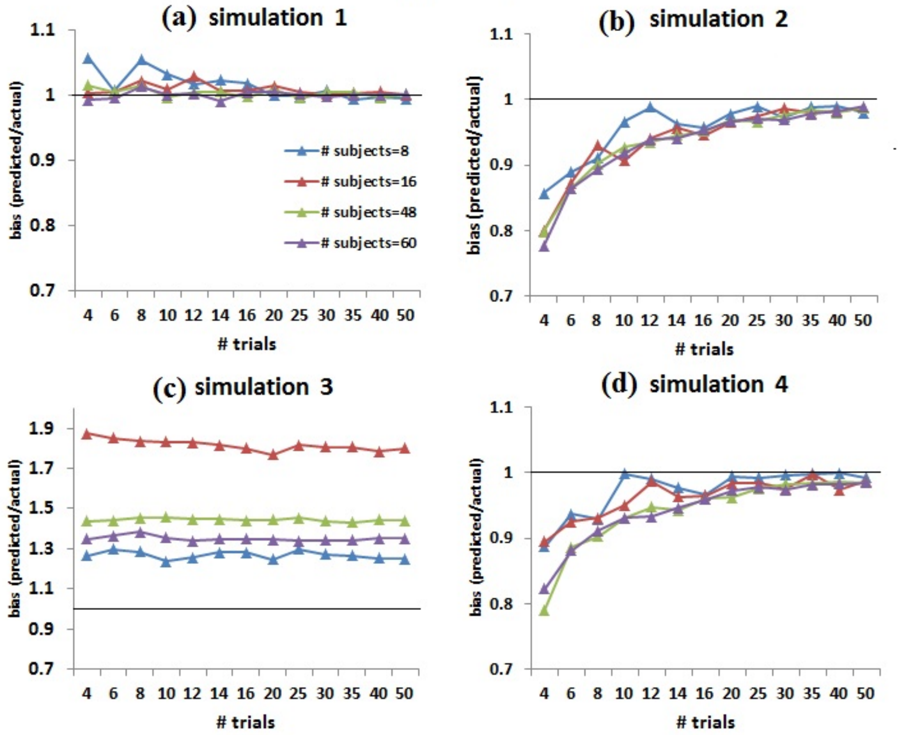


figure 1, ratio1

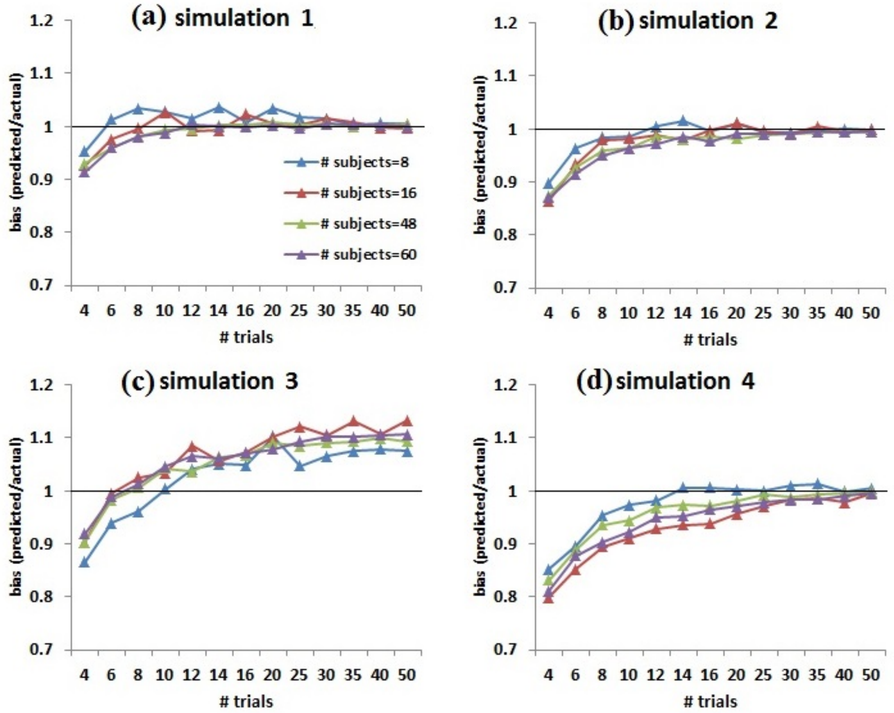


figure 2, ratio2

Nardini and his colleagues (2008) adopted the assumption and modeling strategy of simulation 1. They assumed that different subjects shared common population distributions. As a result, when they integrated the two types of cues, they adopted constant relative weights. This assumption is doubtful, given that different subjects might have interpreted the visual environment differently and had various levels of path integration ability. Even though there is trivial distortion observed for combined variance in the simulation regardless how many trials there are, there is still a small degree of bias for landmark relative weight when there are only 4 trials (bias ratio ≈.91).

When the assumption and the modeling strategy do not match, as in simulation 2 and 3, there is a modest degree of bias, which fades away with increasing number of trials in simulation 2 but does not in simulation 3. What we learn from simulation 3 is that if it is true that subjects have different response distributions and set their own weights to cues, then using the constant weight modeling strategy will cause intrinsic bias, which cannot be corrected empirically either by increasing observations per subject or increasing the sample size.

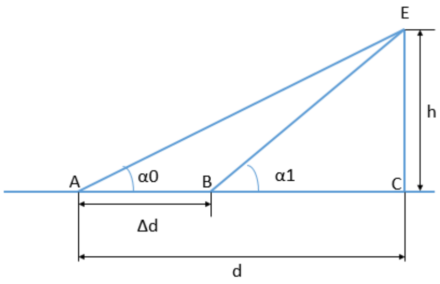
In simulation 4, where we assume different populations and individual relative weights, a modest degree of bias still exits for both estimators when the number of trials is small. But the bias goes away when we increase the number of trials. Increasing the number of subjects does not help solve the bias problem. Assuming subjects have different response variances in the two single-cue conditions, and thus should have their own optimal weights, is more realistic, in our opinion, than the assumption of simulation 1, which was adopted by Nardini and his colleagues (2008).

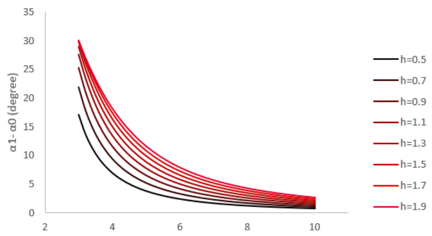
References

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Curr Biol, 18*(9), 689-693. doi: 10.1016/j.cub.2008.04.021

Appendix 3

We will demonstrate in this thought experiment that cue salience is probably positively related to cue reliability. As shown in the figure below, A represents the initial location of the person, CE represents a landmark with a height of h. When at A, the person is at a distance of d from the landmark. The view taken at point A relative to the landmark can be represented by the view angle α_0. Then the person moves to point B by a distance of Δd, and the view angle changes to α_1. Precision of location representation is positively related to the difference between α_0 and α_1, so that a larger difference between the two views leads to better differentiation between locations A and B. Therefore, we can assess the precision of the location presentation by examining how the view changes given a certain amount of change in location. We fixed the distance Δd to be 2m and varied the landmark's height (h) and distance (d). We plotted the difference between view angles (α_1- α_0) as a function of h and d.





The line graph shows that as the distance increases and as the height decreases, view change decreases. This means short and distal landmarks are poorer in cue reliability and will cause noisier location representation than tall and proximal landmarks. This simulation can be easily generalized to landmark's width on the horizontal plane and give us a similar conclusion that as landmarks becomes wider, location representation becomes more precise.

For the case of visual richness, we can conceptually consider a visual environment as consisting of a set of visual elements. Each visual element can be utilized to define locations, and its cue reliability at one specific location can be calculated. Then, multiple visual elements can be combined in a certain way, like linear combination, to refine the representation of that location. Cue combination usually results in greater precision of searches. A good example is demonstrated in Bayesian theory of cue integration, in which cues are integrated to optimally improve response reliability. Suppose there are a series of steps of cue combination, and each step results in a more refined location representation than before. The larger number of visual elements is combined, the more precise the final representation is. Compared to poor environments, rich environments consist of more visual elements and therefore lead to more refined location representations.