Cognitive Psychology 158 (2025) 101734

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

# Processing spatial cue conflict in navigation: Distance estimation

Xiaoli Chen [a,*], Yingyan Chen [a], Timothy P. McNamara [b]

[a] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou 310058, PR China
[b] Department of Psychology, Vanderbilt University, Nashville, TN 37240, USA

ARTICLE INFO

*Keywords:*
Spatial cognition
Spatial navigation
Causal inference
Cue conflict
Cue integration
Bayesian models
Path integration

ABSTRACT

Spatial navigation involves the use of various cues. This study examined how cue conflict influences navigation by contrasting landmarks and optic flow. Participants estimated spatial distances under different levels of cue conflict: minimal conflict, large conflict, and large conflict with explicit awareness of landmark instability. Whereas increased cue conflict alone had little behavioral impact, adding explicit awareness reduced reliance on landmarks and impaired the precision of spatial localization based on them. To understand the underlying mechanisms, we tested two cognitive models: a Bayesian causal inference (BCI) model and a non-Bayesian sensory disparity model. The BCI model provided a better fit to the data, revealing two independent mechanisms for reduced landmark reliance: increased sensory noise for unstable landmarks and lower weighting of unstable landmarks when landmarks and optic flow were judged to originate from different causes. Surprisingly, increased cue conflict did not decrease the prior belief in a common cause, even when explicit awareness of landmark instability was imposed. Additionally, cue weighting in the same-cause judgment was determined by bottom-up sensory reliability, while in the different-cause judgment, it correlated with participants' subjective evaluation of cue quality, suggesting a top-down metacognitive influence. The BCI model further identified key factors contributing to suboptimal cue combination in minimal cue conflicts, including the prior belief in a common cause and prior knowledge of the target location. Together, these findings provide critical insights into how navigators resolve conflicting spatial cues and highlight the utility of the BCI model in dissecting cue interaction mechanisms in navigation.

## 1. Introduction

Spatial navigation, a fundamental ability crucial for both human and animal survival, depends on the ability to combine spatial cues (e.g., landmarks in the environment, proprioceptive cues from self-motion) to estimate location. Navigators must estimate their own locations in the environment, as well as the locations of goals. According to traditional models of spatial learning (e.g., Gallistel, 1990; O'Keefe & Nadel, 1978; Siegel & White, 1975), navigators can develop a mental representation of an environment through repeated explorations. This representation, commonly referred to as "cognitive map", includes straight-line distances and directions between locations (Tolman, 1948). It has been argued that cognitive maps enable efficient and flexible navigation in familiar environments (for alternative views, see the cognitive graph theory in Ericson & Warren, 2020 and the cognitive collages hypothesis in Tversky, 1993).

A key challenge in spatial navigation is resolving conflicts between different spatial cues. This problem has been extensively studied

* Corresponding author.
*E-mail addresses:* chenxiaoli54@gmail.com, cxlpsy@zju.edu.cn (X. Chen).

https://doi.org/10.1016/j.cogpsych.2025.101734
Received 15 May 2024; Received in revised form 6 April 2025; Accepted 16 April 2025
Available online 9 May 2025

across disciplines. Much work has examined conflicts between landmarks and self-motion cues. Navigation with self-motion cues, such as proprioceptive inputs, vestibular signals, and optic flow, requires continuous integration of self-movement to determine one's location, a process known as path integration (Etienne & Jeffery, 2004; Mittelstaedt & Mittelstaedt, 1980). In contrast, landmarks are prominent environmental features that provide direct spatial information. These two navigation modes recruit distinct and independent cognitive-neural mechanisms (Chen et al., 2017, 2019, 2024; Shettleworth & Sutton, 2005), raising an interesting question of how these two cue types interact during navigation. Behavioral studies in humans have reported mixed findings: some suggest a predominance of landmarks over self-motion cues when conflicts are large (Zhao & Warren, 2015a), others suggest the opposite (Sjolund et al., 2018), and some show no change in cue weighting until the conflicts become extreme (Zhao & Warren, 2015b). Neuroscience studies in non-human animals typically reveal that spatially-modulated neurons respond to both cue types (Campbell et al., 2018; Chen et al., 2013; Gothard et al., 1996), but preferences vary by brain region: the retrosplenial cortex favors landmarks, whereas the entorhinal cortex favors self-motion cues (Campbell et al., 2021).

Beyond landmark vs. self-motion conflicts, research has also examined how geometric and featural cues interact across multiple disciplines (see review papers, Cheng, 2008; Cheng et al., 2013; Cheng & Newcombe, 2005; Lew, 2011; Newcombe, 2023). Geometric cues refer to environmental features related to shape, layout, and spatial structure, such as the shape of a room. Featural cues refer to distinct, identifiable aspects of an environment, such as an isolated landmark at one of the room corners. Using the reorientation paradigm, Cheng demonstrated that rats predominantly relied on geometric cues rather than featural cues for reorientation, supporting a geometric module hypothesis (Cheng, 1986). However, later studies have shown that navigators make use of both geometric and featural cues, with cue reliance varying based on factors such as cue salience, navigation history, and language use (see a recent review, Newcombe, 2023). These findings have led to the adaptive cue combination hypothesis, which posits that spatial cue utilization is flexible and depends on contextual demands (Xu et al., 2017).

Additional studies have examined conflicts between other spatial cue types, such as an individual landmark vs. multiple landmarks in an array (Jetzschke et al., 2017; Roy et al., 2023) and distal vs. proximal landmarks (Knierim, 2002; Qi & Mou, 2024; Shapiro et al., 1997; Tanila et al., 1997; Yoganarasimha et al., 2006).

Across these studies, cue conflict mainly serves as an experimental tool to assess navigators' relative reliance on different spatial cues. Researchers evaluate cue reliance by analyzing response distributions. When responses are continuous, cue weighting is inferred from the relative proximities of the response centroid to the target locations defined by conflicting cues. The closer the response centroid to the location defined by a particular cue, the greater the reliance on that cue (see Chen et al., 2017, for a review). When responses are discrete, such as in the reorientation paradigm, cue weighting is assessed based on the proportion of trials in which participants choose the location defined by a given cue (Ratliff & Newcombe, 2008). While this approach has provided valuable insights, it does not fully reveal the cognitive processes that navigators use. Specifically, it remains unclear (a) how navigators decide whether conflicting sensory-perceptual information is informative about the world (i.e., there are different sources or causes) or should be ignored (i.e., the conflicts are caused by sensory-perceptual error), and (b) how they select a goal location when they have determined that discrepant spatial cues should not be integrated.

### *1.1. Cognitive models accounting for navigation behavior in cue-conflicting situations*

Several models have been proposed to explain navigation behavior in cue-conflicting situations. While these models offer valuable insights, they have limitations, such as lacking mechanistic explanations of cue detection and resolution, being constrained to specific spatial cue types, and failing to generalize across tasks.

Jetzschke and colleagues proposed a probabilistic model to explain continuous spatial localization in a 2D environment, with a landmark conflicting with other landmarks in an array (Jetzschke et al., 2017). In the standard maximum-likelihood-estimation (MLE) model of cue integration, the likelihood distributions of individual cues are assumed to be Gaussian. Cue integration involves multiplying individual likelihood distributions, which results in a more precise joint likelihood distribution (Bromiley, 2013; McNamara & Chen, 2022, Appendix A). Unlike the MLE model, Jetzschke's model assumes that each individual likelihood distribution is a mixture of two Gaussians, one of which has very heavy tails. Multiplying these mixture distributions does not result in a more precise joint likelihood distribution, eliminating the typical gain from cue integration. However, this model remains primarily descriptive and does not explain how cue conflict is detected and resolved.

Harootonian and colleagues tested models for head direction estimation, considering body-based self-motion cues and visual feedback (Harootonian et al., 2022). Their findings support a hybrid model: cues are integrated when cue consistency is assumed, but only body-based cues are used when cue inconsistency is assumed. However, this model does not incorporate a mechanism for detecting cue conflict; instead, this treats the proportion of trials for cue integration as an independent, freely varying parameter.

The adaptive cue combination model aims to explain conflicts between geometric cues and featural cues in the reorientation paradigm (Xu et al., 2017, Case study 3). This model follows the principles of the standard MLE model of cue integration (Rohde et al. 2016). The MLE model, however, was originally developed to explain cue combination behavior in scenarios with minimal or no cue conflicts, making it theoretically unsuitable for situations with substantial cue conflicts (French & DeAngelis, 2020; Newman et al., 2023). Consequently, this model fails to distinguish between conditions conducive or inconducive to cue integration, where different navigation strategies should be adopted (Sjolund et al., 2018; Zhao & Warren, 2015b).

Similarly, Wang and colleagues propose a model for a reorientation task (Wang et al., 2018), which contrasted two intersecting streets (geometric cue) and trees placed at the intersection (landmark cue). This model is also based on the standard MLE model of cue integration, but logit-transformed behavioral accuracy was used as a proxy of cue reliability, due to the discrete nature of target location and response. Consequently, this model has the same limitations of the adaptive cue combination model (Xu et al., 2017).

Furthermore, the experimental design differs from typical reorientation tasks, as participants were restricted to choosing from two of the four street ends. Consequently, this model's generalizability to standard reorientation tasks remains uncertain.

Beyond these models, the view-matching model (Cheung et al., 2008) and the associative learning model (Miller & Shettleworth, 2007) offer mechanistic explanations for detecting and resolving conflicts between geometric and featural cues in the reorientation paradigm. However, the view-matching model is inapplicable to tasks involving self-motion cues, as spatial locations defined by such cues do not correspond to specific views. The associative learning model relies on feedback to adjust the association strength of the cues with the reward, but this type of feedback is absent in many navigation tasks.

In summary, several cognitive models have been proposed to explain navigation behaviors under conditions with cue conflicts, but they face significant limitations, including a lack of mechanistic explanations and restricted applicability across tasks. To address these issues, a more comprehensive model is needed – one that incorporates a mechanism for detecting and resolving cue conflicts, accounts for situations that either support or hinder cue integration, and applies broadly to navigation tasks. The Bayesian causal inference (BCI) model represents such a model.

### *1.2. Bayesian causal inference model*

The Bayesian causal inference (BCI) model provides a framework for understanding how the brain combines and processes information from multiple sensory sources to create a coherent and accurate perception of the external world (Körding et al., 2007). In the context of spatial navigation, the BCI model posits that the perceived location is inherently corrupted by intrinsic sensory noise, meaning that the perceived location often does not correspond to the true location as defined by the spatial cues. For example, the observer would perceive different self-locations at different times even when the same location is occupied. Therefore, the navigator cannot determine the true location at which they are actually positioned. Instead, the navigator infers the true self-location based on the perceived one, with a certain degree of uncertainty that is proportional to the amount of sensory noise inherent in the spatial input. This uncertainty inherent in perceiving self-location makes the causal structure inference a non-trivial problem, that is, determining whether different spatial inputs stem from the same location or different locations. For example, the perceived locations from different cues can be in large disparity even when the cues are congruent with each other in the physical space; conversely, the perceived locations from different cues can be close in space even when the cues are in substantial conflict in the physical space. Hence, the causal structure inference is not definitive and subject to uncertainty.

Fig. 1 illustrates the conceptual structure of the BCI model. The BCI model addresses two key questions. First, how do navigators judge whether spatial cues are congruent (i.e., sensory inputs originate from the same cause) or incongruent with each other (i.e., sensory inputs originate from different causes)? Second, what strategies do they employ to handle spatial cues based on this cue-congruence judgment?

Regarding the first question of cue-congruency judgment, the BCI model posits that an ideal observer combines the following three information sources: prior belief in a common source, prior knowledge about possible target locations, and sensory inputs stemming from different cues. The combination of sensory inputs and prior knowledge about target location generates the likelihood of a common cause, which is then combined with the prior belief in a common cause to generate the posterior belief in a common cause. All

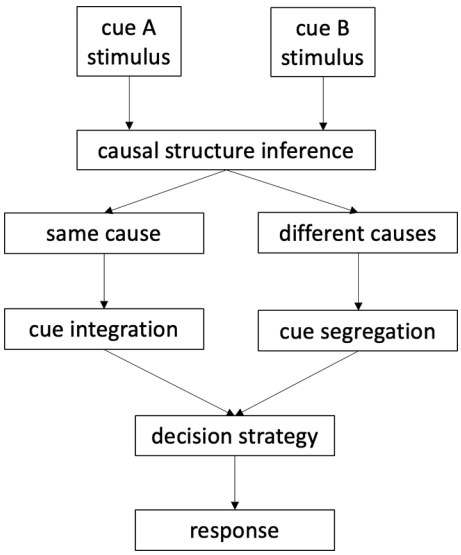



**Fig. 1. Conceptual framework of BCI model** In the BCI model, first the observer performs causal structure inference, judging whether the two stimulus inputs (cue A and cue B) come from the same cause or different causes. If the two cues are judged to come from the same cause, the cues are integrated. If the two cues are judged to come from different causes, the cues are segregated. The outcomes from the same-cause judgment and different-cause judgments are combined through a certain decision rule to generate the response.

else equal, the probability of making a same-cause judgment increases with higher prior belief in common source, more widespread prior location distribution, and more similar sensory inputs.

Regarding the second question of cue-handling strategies, first, the BCI model conceives different sub-models corresponding to different cue-congruency judgments: cue integration for the common-cause judgment and cue segregation for the different-cause judgment (Körding et al., 2007; Wozny et al., 2010). In the integration sub-model, different spatial inputs are judged to be in congruency and thus are integrated, following the MLE principles. This cue integration yields the joint likelihood distribution, which is then integrated with the prior knowledge about the target location distribution to generate the posterior distribution, embodying the Bayesian theorem of combining likelihood and prior information. In the cue segregation sub-model, spatial inputs from the two cues are judged to be in conflict and thus not integrated. Typically, the task-relevant cue type is selected, whose likelihood distribution is then integrated with the prior knowledge about target location distribution to generate the posterior distribution. Next, the two posterior distributions derived from the cue integration sub-model and the cue segregation sub-model are combined using different decision rules, depending on the goal prioritized by the observer (Wozny et al., 2010).

To illustrate the benefits of applying the BCI model to understand navigation in cue-conflicting situations, consider the conflicts that are created when a landmark is relocated in space to be in conflict with other stable cues (commonly referred to as landmark instability). A frequent finding is that navigators rely less on unstable landmarks (Auger et al., 2015; Chen et al., 2017; Roy et al., 2023; Sjolund et al., 2018; but see Zhao & Warren 2015a). However, at the process level, it remains unclear why unstable landmarks reduce navigators' reliance on them and what cognitive processes landmark instability affects. From the perspective of the BCI model, one possible explanation is that navigators gradually acquire knowledge about landmark instability statistics through experience, leading to a decreased prior belief in a common cause for unstable landmarks (Roy et al., 2023). A decreased prior belief in common-cause leads to a decreased posterior probability of a common cause, which entails more frequent different-cause judgments. If navigators assign a lower weight to landmarks in the different-cause judgment than in the same-cause judgment, the observed landmark reliance would be lessened.

However, other possibilities exist. For example, reduced landmark reliance observed in behavior could be caused by a lower weight assigned to landmarks in the different-cause judgment, while the prior belief in a common cause remains unchanged. Additionally, landmark instability may increase sensory noise of landmarks, which is typically reflected in poorer performance associated with landmarks (Auger et al., 2015; Biegler & Morris, 1993; Chen et al., 2017). The increased sensory noise lowers the weight assigned to landmarks in the common-cause judgment, as stipulated by the MLE rule of weighting cues by their relative reliabilities (Rohde et al., 2016). Reduced weight for landmarks in the common-cause judgment translates to reduced reliance on landmarks observed in behavior. In both cases, reduced reliance on landmarks emerges without changes in the prior belief in common cause or the frequency of making this judgment.

In summary, in this concrete example, spatial cue conflict potentially influences multiple cognitive processes to cause a decrease in navigators' dependence on unstable landmarks. The BCI model comprises parameters that reflect distinct cognitive processes, allowing us to pin down the specific processes that are affected. Furthermore, the BCI model conceptualizes information as probability distributions that can be either continuous or discrete (see Section 4 "Cognitive Modeling" for details), it generalizes well across various navigation scenarios.

### *1.3. BCI model as a framework for understanding cue combination suboptimality*

The BCI model not only accounts for navigation behavior in cue-conflict situations but also provides insights into cue combination suboptimality observed even with minimal cue conflicts present. Cue combination suboptimality refers to the situation when the observed response precision is lower than the prediction of the MLE model (Rohde et al., 2016). While some navigation studies suggest that people can integrate visual spatial cues (featural landmark cues or geometric cues) and body-based self-motion cues in an optimal or near-optimal manner (Chen et al., 2017; Nardini et al., 2008; Sjolund et al., 2018), others report suboptimal cue combination effects. For example, cue combination suboptimality occurs between different types of self-motion cues (visual optic flow vs. proprioceptive cues) (Chrastil et al., 2019) and between different visual landmarks (Newman & McNamara, 2022). Suboptimal cue combination behavior is also commonly observed in other perceptual domains (refer to Section 2.7.2 in Rahnev & Denison, 2018 for a summary).

One potential factor contributing to cue combination suboptimality is prior knowledge of stimulus distribution. Because this prior knowledge is shared across cue conditions, it causes correlated errors, reducing cue integration benefits (Oruç et al., 2003). The more precise the prior distribution of the target location, the stronger its influence, and the lower the gain in response precision from cue integration. When the stimulus distribution spans a relatively wide range and continuous responses are required, utilizing prior knowledge of stimulus distribution leads to a well-documented phenomenon known as the central tendency effect, wherein observers' responses are biased towards the mean of the stimulus distribution (Hollingworth, 1910; Petzschner et al., 2015; Petzschner & Glasauer, 2011). Aston et al. quantified prior knowledge's influence based on this effect and then excluded it from responses, uncovering the sensory cue integration process. However, this approach is constrained by the detectability of central tendency, which diminishes when the target range is narrow. Furthermore, with discrete distributions, such as those obtained in the reorientation paradigm (Cheng, 1986), the central tendency effect is challenging to quantify. In this case, the distribution mode should represent the central tendency, which can be complicated by multimodal distributions. Prior knowledge also shapes behavior in categorical tasks (Ratliff & Newcombe, 2008). In contrast, the BCI modeling approach offers broader applicability by accommodating prior knowledge beyond conditions that elicit the central tendency effect.

Besides prior knowledge of target distribution, some studies sought to explain suboptimal cue combination by attributing what cannot be explained by the MLE model to other forms of prior knowledge or prior preference (Byrne & Crawford, 2010; Kersten &

Yuille, 2003; Qi & Mou, 2024). However, these studies often lack independent data to verify the use of the claimed prior knowledge or preference. The BCI model overcomes these problems by providing a unified framework that incorporates the dynamic interplay among multiple factors, including those contributing to suboptimal cue combination, with prior knowledge and prior belief as critical contributing factors. In doing so, the BCI model enhances our understanding of the broader question of spatial cues interaction, a central focus of navigation research.

### 1.4. A non-Bayesian alternative model

Although the BCI model offers a valuable framework for understanding spatial navigation in both cue-conflicting and cue-congruent situations, its validity and robustness require evaluation through comparisons with alternative models. Previous studies have compared the BCI model to subsets of this model (e.g., full segregation or full integration) (de Winkel et al., 2017, 2018; Körding et al., 2007), to other Bayesian models (Körding et al., 2007), or to alternative variants of the BCI model (Badde et al., 2020; Wozny et al., 2010). However, none of these studies has compared the BCI model to a non-Bayesian model. A crucial element of Bayesian models is the use of prior information, which corresponds to the prior belief about causal structure and prior knowledge of the target distribution in the BCI model. In contrast, a non-Bayesian model should exclude prior information.

To address this gap, we propose the sensory disparity model, which employs a non-Bayesian causal inference mechanism (for details, see the Methods section). The primary distinction between this model and the BCI model lies in how causal structure judgments are made. The BCI model incorporates both sensory inputs and prior information (i.e., prior belief about causal structure and prior knowledge of the target distribution) in making causal structure judgments. In contrast, the sensory disparity model only relies on sensory inputs for making such judgments: the greater the absolute distance between sensory measurements from the two cue types, the lower the likelihood of a common-cause judgment. Furthermore, the sensory disparity model retains one key feature of the BCI model – the incorporation of sensory noise. Therefore, comparing the BCI model with the sensory disparity model provides a targeted test of the primary tenet of the BCI model, namely the use of prior information.

### 1.5. Present study

The overarching objective of the current study is to investigate spatial cue conflicts in navigation. We were especially interested in how navigators decide whether discrepancies between spatial inputs arise from sensory-perceptual error or indicate distinct environmental causes, and how they select goals accordingly. To accomplish this objective, we applied the BCI model to a spatial navigation task and compared it with a non-Bayesian sensory disparity model.

To test these models, we developed a novel cue combination paradigm along a linear track, building on paradigms established in our previous work (Chen et al., 2019, 2024; Kuehn et al., 2018). The task required participants to localize target locations by using either a visual landmark or visual self-motion cues (i.e., optic flow). When using the visual landmark, they need to estimate their distance to the landmark; when relying on visual self-motion information, they need to estimate their distance from the starting position of self-movement.

This task is limited compared to real-life spatial navigation, as it only probes one aspect of spatial navigation – distance estimation. Terrestrial spatial navigation is typically carried out in a two-dimensional space, involving angular estimation, distance estimation, and vector computations. Even so, distance estimation is an essential element for spatial navigation. For example, straight-line distances between locations are an essential component of survey knowledge, or a cognitive map. The importance of distance estimation extends beyond navigation. For example, time estimation is closely intertwined with spatial distance estimation (Riemer et al., 2022; Umbach et al., 2020), as it is essentially distance estimation in the temporal domain. Hence, investigating one-dimensional spatial distance estimation can help understand basic processes of spatial navigation and other related topics such as time perception.

Additionally, distance estimation is ubiquitous in real-life navigation, where cue conflicts often occur. Imagine navigating an unfamiliar city to find a café. You first follow a specific route, judging the distance you've traveled in a fixed direction to determine when to make a turn. Along the way, you also use a landmark, such as a vender's booth in outdoor market, to confirm you are nearing your destination by estimating your distance to the landmark. Confusion arises when the vender moves to a different location. At this point, you must decide whether to rely on the distance you believe you've traveled along the route or adjust your judgment based on the perceived distance to the landmark. This scenario illustrates the challenges of reconciling conflicting spatial cues in distance estimation during navigation.

Moreover, our recent fMRI studies have demonstrated that linear navigation tasks engage key brain areas for spatial navigation, including the retrosplenial cortex, hippocampus, and entorhinal cortex (Chen, Vieweg, & Wolbers, 2019; Chen, Wei, & Wolbers, 2024, 2025; Chen et al., 2022). Linear navigation tasks are also widely used in electrophysiological studies on spatial navigation (Fischer et al., 2020; Mao et al., 2017; Saleem et al., 2018, to name a few). Hence, investigating spatial distance estimation in humans can facilitate inter-species comparisons and enhance our understanding of the cognitive-neural mechanisms underlying spatial navigation.

Finally, the use of the linear track navigation task allowed us to collect a substantial amount of data, which is essential for distinguishing complex cognitive models of a task (Lerche et al., 2017). This strength of the paradigm aligns with one of our objectives: to rigorously evaluate competing models for spatial cue conflicts.

In summary, this study examined how individuals resolve spatial cue conflicts during navigation in a linear track navigation task using the cognitive modeling approach. Our aim was to provide insights into the mechanisms underlying spatial cue interactions.

## 2. Methods

### 2.1. Participants

A total of 142 participants were recruited, all with normal or corrected-to-normal vision and no history of neurological diseases. Participants gave written informed consent prior to the experiment, and received course credits or monetary compensation after the experiment. The study was approved by the local ethics committee of Zhejiang University, Hangzhou, China.

Participants were randomly assigned to three groups. The large-conflict-absent group (LC-absent) included 60 participants (30 female, mean age (SD) = 21.05 (2.012) year). The large-conflict-present group (LC-present) included 42 participants (27 female, mean age (SD) = 22.238 (1.750) year). One participant was excluded for not following the experimental instructions, leaving 41 participants for analyses. The large-conflict-informed group (LC-informed) included 40 participants (22 female, mean age (SD) = 21.400 (1.780) year). The rationale behind the group naming can be found in the following Procedure section.

The LC-absent group included more participants as it also served as a baseline group for other experiments in our lab. All participants in this group were included in the analyses for the purpose of maximizing statistical power. Furthermore, results remained consistent when analyses were restricted to the first 40 participants in this group.

We started the study with a plan to run a relatively large sample size of participants (i.e., around 40 per group), based on common practice in behavioral cognitive experiments. Our primary interest was whether cue conflict reduces weight assigned to landmarks. Therefore, we performed a statistical power analysis for this effect based on raw data from a previous study, which employed a similar task paradigm in the 2D space using immersive VR and showed that participants experiencing unstable landmarks assigned lower behavioral weight to landmarks compared to those experiencing stable landmarks (Chen et al., 2017, Experiment 2). We expected two possible scenarios. In the first scenario, the LC-informed group assigned lower behavioral weight to landmarks compared to the other two groups. In the second scenario, both the LC-present and LC-informed groups assigned lower behavioral weight to landmarks compared to the LC-absent group. In both scenarios, the probability of detecting significant differences was greater than 90 %, indicating an adequate level of statistical power. To further ensure the reliability of our results, we also reported Bayes factors, which, among many strengths (Rouder, 2014; Rouder et al., 2009), can assess strength of evidence for both alternative and null hypotheses.

### 2.2. Apparatus and materials

Participants completed a spatial localization task in a desktop virtual reality environment rendered in Worldviz 5.0 (https://www.worldviz.com) on a 24-inch monitor screen (Fig. 2a). In the task, the participant traveled along a linear track, meaning only translation was permitted and rotation was disabled. A set of three arrows and the landmark (a tree) were positioned at the two ends of the linear track. The set of arrows was fixed at the position of 0 virtual meter (vm) on the linear track and served as the anchoring point for path integration. The set of arrows consisted of three identical red arrows positioned at the same horizontal position but at different heights from the ground, for the purpose of optimizing the visibility of the path integration anchor from varying distances. The landmark, whose original position was at 51 vm, served as the anchoring point for landmark-based navigation. The landmark was relocated to different positions, creating conflict between landmarks and self-motion cues in the conflict conditions (see the following "Task and Design" section). The floor was covered with a collection of life-limited white dots (life duration = 1 s), whose positions were randomly determined. This collection of dots could provide optic flow information as a form of path integration cue during movement. The target location was randomly sampled from a very narrow uniform distribution [-0.8 vm, 0.8 vm] centered at the position of 14 vm on a trial-by-trial basis.

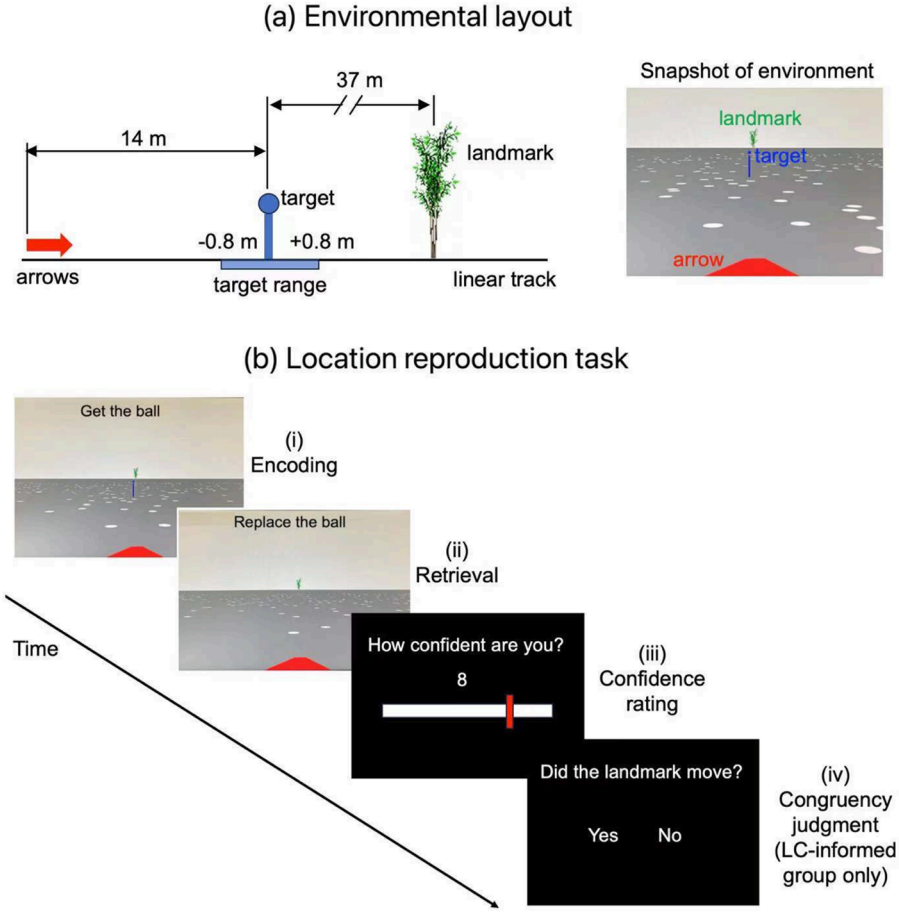
### 2.3. Task and design

The main task participants completed was a location reproduction task based on one-dimensional distance estimation (Fig. 2b). Participants were restricted to linear movements along the linear track on the horizontal plane, allowing for only forward and backward translation without any lateral (side-to-side) movement. Rotational movements, such as turning left or right, were also disabled, meaning that participants' orientation remained fixed throughout the navigation task.

Each trial consisted of an encoding phase and a retrieval phase. During the encoding phase, participants traveled to the location of the target, which was a blue ball positioned at the top of a pole. The target disappeared once participants had reached its position. During the retrieval phase, the target remained invisible, and participants needed to travel to the remembered target location. Participants were instructed to be as accurate as possible, but not to spend unnecessary extra time in making the response. Afterward, participants rated self-confidence on a scale from 1 (least confident) to 10 (most confident), in increments of 1. Participants were encouraged to distribute confidence ratings across the entire scale.

There were five cue conditions: landmark, self-motion, combination, small conflict, and large conflict. The first two were single-cue conditions, wherein only one of the spatial cues was available in a given trial. The latter three were double-cue conditions, wherein both landmark and self-motion cues were present in each trial.

In the landmark condition, the landmark was visible, so that participants could localize themselves by estimating their distance to the landmark. The arrows were invisible. In both the encoding and retrieval phases, the starting position of participants' movement was randomly sampled on a trial-by-trial basis from a uniform distribution, ranging from −6 vm to + 6 vm around the arrows' position. In this way, participants could not perform path integration to estimate the traveled distance to infer self-position. In addition, once participants had started to move, the flashing dots on the floor moved quickly along or against participants' movement direction,




(caption on next page)

**Fig. 2. Environmental layout and experimental task (a)** The left panel shows a schematic of the environmental layout. The arrow symbol represents a set of three red arrows, which occupied the same horizontal position but differed in height. The set of arrows served as the fixed starting position for the movement in the self-motion condition and the double-cue conditions. The target was a blue ball positioned on the top of a blue post, whose position was randomly sampled from a narrow uniform distribution [-0.8 vm, 0.8 vm] with a center 14 m away from the arrows. The bamboo tree served as a landmark, positioned 37 vm away from the ends of the arrows. For better illustration, the landmark is depicted closer to the target than its actual position in the experiment. The arrows and the target were located on an imaginary straight line in the middle of the track. The landmark was located near this imaginary straight line but with a small offset of 0.5 vm to the right. Participants' movement was restricted to the imaginary straight line. The right panel shows a snapshot of the environment taken from the participant's first-person perspective when located at the position of the arrows. **(b)** Time course of the location reproduction task. The task consisted of five phases. In phase (i) "Encoding", the participant saw the target – a blue ball on the top of a post. The participant needed to move to the location of the target, which disappeared once the participant arrived. In the landmark condition, the starting position of movement was randomized from trial to trial around the arrows within a range [-6 vm, +6 vm], and the arrows remained invisible. In other cue conditions, the starting position of movement was fixed at the arrows, which remained visible. The landmark was invisible in the self-motion condition but visible in other cue conditions. In phase (ii) "Retrieval", the participant was required to travel to the remembered target location from the "Encoding" phase. The target remained invisible. The manipulation of the starting position of movement was the same as in the encoding phase. In phase (iii) "Confidence Rating", the participant rated their confidence in the accuracy of their response on a scale from 1 (least confident) to 10 (most confident) in increments of 1. A small red vertical bar on the screen could be adjusted to indicate the confidence level. In phase (iv) "Congruency Judgment", the participant had to decide whether the landmark had moved to a different position during the retrieval phase. This phase was presented only to the LC-informed group, but not to the LC-absent and LC-present groups. **(c)** Snapshots of the landmark taken from the participant's first-person perspective when the participant stood at the center of the target location range (14 vm) during the retrieval phase in the three double-cue conditions: the combination condition (left panel; conflict = 0 vm), the small conflict condition (middle panel, conflict = 5 vm), and the large conflict condition (right panel, conflict = 20 vm). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

substantially degrading the optic flow information that could be used for path integration (refer to Chen et al., 2019, the low-reliability self-motion condition). The movement speeds of the dots were randomly sampled from a normal distribution with a mean of 0 vm/s and a standard deviation of 6 vm/s. Before the movement started and once the movement had stopped, the dots remained in place on the floor.

In the self-motion condition, the arrows were visible, serving as the anchoring point for path integration. In each trial, participants traveled from the arrows' position, so they could infer self-position by performing path integration (i.e., estimating the traveled distance from the starting position). The flashing dots remained in place throughout the trial to provide stable optic flow for path integration. The landmark remained invisible throughout the trial, eliminating landmark-based navigation. To prevent participants from adopting a pure timing-based strategy, the movement speed was randomly selected from a uniform distribution ranging from 2 to 5 vm per second on a trial-by-trial basis. Speed randomization was also employed in other cue conditions.

In the combination condition, the landmark stayed at its original position and the two cue types were congruent with each other during both the encoding and retrieval phases, meaning both cue types could be utilized for localization. In the small conflict and large conflict conditions, the landmark was moved from its original position to be farther away from the arrows by a certain distance during the retrieval phase, so that the target position defined by self-motion cues was different from that defined by the landmark. The landmark was moved by 5 vm in the small conflict condition and 20 vm in the large conflict condition (Fig. 2c).

Participants were randomly assigned to three groups: large-conflict-absent group (LC-absent), large-conflict-present group (LC-present), and large-conflict-informed group (LC-informed). For all the three groups, the procedure included the landmark, self-motion, combination, and small conflict conditions. In the LC-absent group, the large conflict condition was absent. In the LC-present group, the large conflict condition was included. In these two groups, to ensure that the two cue types were at equal footing, participants were informed at the beginning of the experiment that both the arrows (i.e., the anchoring point of path integration) and the landmark (i.e., the anchoring point of landmark-based navigation) would remain stable throughout the procedure. Therefore, when a conflict was detected between them, it could be interpreted as arising from the instability of either the landmark or self-motion cues.

In the LC-informed group, the large conflict condition was included, and participants were made explicitly aware of the possibility of cue conflict. Specifically, participants were told at the beginning of the experiment that the landmark might move during retrieval, and when this occurred, it should not be trusted when displaced. Participants were also required to judge whether the landmark had been shifted at the end of each double-cue trial (Fig. 2). This design encouraged participants to attribute cue conflict to landmark instability. The inclusion of the LC-informed group was motivated by previous findings that conflicts between landmarks and self-motion cues alone may not induce changes in cue weighting (Zhao & Warren, 2015a, 2015b).

### *2.4. Procedure*

The experiment took place in a single session. Trials were organized into six runs, each containing six blocks of four trials. Blocks were randomized in each run. For the LC-absent group, each run consisted of one block of the landmark condition, one block of the self-motion condition, two blocks of the combination condition, and two blocks of the small conflict condition. The four trials in a block always belonged to the same cue condition. The procedure included a total of 144 trials, with 24 trials for each single-cue condition and 48 trials for each double-cue condition.

For the LC-present and LC-informed groups, there were five cue conditions: landmark, self-motion, combination, small conflict, and large conflict. Each run contained one block for each of the five cue conditions, plus a mixture block with trials from all three double-cue conditions The total number of trials was matched among the three double-cue conditions across the six runs. Like the LC-absent

group, these groups also completed 144 trials, with 24 trials for each single-cue condition and 32 trials for each double-cue condition.

Participants took a 2-min rest between the runs. At the end of each run, they received feedback on their mean unsigned distance error averaged across all the trials in the run, excluding conflict trials where no correct target location could be defined. This vague feedback served only to maintain participants' attention on the task and should not alter their navigation strategies.

Prior to the experimental task, participants completed a preparation phase, during which they were familiarized with the virtual environment and keyboard controls. They navigated a visually rich 2-dimensional open-field environment. Participants were required to remember the colors of five balls scattered around the environment. Afterwards, they needed to travel to the location of each ball, recalling its color. This environment gradually transitioned to the sparse environment employed in the location reproduction task based on one-dimensional distance estimation (Fig. 2). Finally, participants practiced six experimental trials (two each from the landmark, self-motion, and combination conditions). The preparation phase lasted about 10 min.

## 3. Behavioral analysis and results

### 3.1. Behavioral analysis

The behavioral data analysis followed the same procedure adopted in our previous study (Chen et al., 2017), testing the MLE principles of optimal cue integration (Rohde et al., 2016). One critical difference is that in the current study, responses were recorded in a one-dimensional space rather than a two-dimensional space. The main rationale is to evaluate whether participants combined landmarks and self-motion cues in a statistically optimal manner in the double-cue conditions in two aspects: whether the response precision was improved in double-cue conditions compared to single-cue conditions in a statistically optimal manner and whether the weights assigned to different cues complied with the MLE rules (i.e., weighting cues by their relative cue reliability). The names and meanings of all the variables involved in the behavioral analysis are listed in Table 1.

Because the target location was randomly sampled around 14 vm from a uniform distribution with a very narrow range [-0.8 vm, +0.8 vm], in each trial, the response was transformed into a spatial coordinate with the current target location as the origin by subtracting the current target location from the response location. Next, the transformed responses were pooled across trials for each cue condition. In each cue condition, outlier responses were defined as responses whose distance from the centroid of all responses exceeded the 3rd quartile by 3 * interquartile range (IQR). Response variability was calculated as the standard deviation of the response distribution:

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

where $d$ is the Euclidean distance of each response to the centroid of the response distribution, and $n$ is the number of responses in the distribution.

If landmarks and self-motion cues were optimally integrated according to the MLE model of cue integration, the predicted response variability in the double-cue condition is:

$$S_{opt,d} = \sqrt{S_{obs,l}^2 S_{obs,m}^2/(S_{obs,l}^2 + S_{obs,m}^2)}$$

Cue reliability is equal to the inverse of the squared response variability (i.e., response variance). Expressed in terms of response variabilities, cue relative reliability (for landmark cues) is:

$$rr = S_{obs,m}^2/(S_{obs,l}^2 + S_{obs,m}^2)$$

Cue relative reliability $rr$ also represents the behavioral optimal weight ($w_{opt,beh}$), which is the optimal weight that should be assigned to landmarks in optimal cue integration. Typically, this term is named "optimal weight" in the literature. Here, the adjective "behavioral" is added to the term to convey the idea that this term is calculated from participants' behavioral responses (Aston et al.,

**Table 1**
Glossary table of all variables in the behavioral analysis testing the MLE model of cue integration.

| Variable | Description |
|---|---|
| $S_{obs}$ | Observed response variability (the standard deviation of the response distribution) |
| $S_{obs,l}$ | Observed response variability in the landmark condition |
| $S_{obs,m}$ | Observed response variability in the self-motion condition |
| $S_{opt,d}$ | Predicted response variability in the double-cue condition based on the MLE model of cue integration (i.e., the optimal response variability). |
| $S_{alt,d}$ | Predicted response variability in the double-cue condition based on the cue alternation model |
| $rr$ | Cue relative reliability for landmarks (=$w_{opt,beh}$) |
| $w_{opt,beh}$ | Behavioral optimal weight (=$rr$) |
| $w_{obs}$ | Observed weight assigned to landmarks in a double-cue condition |
| $\overline{X_l}$ | Centroid of the response distribution in the landmark condition |
| $\overline{X_m}$ | Centroid of the response distribution in the self-motion condition |
| $d_{l-d}$ | Euclidean distance between the response centroids in the landmark condition and the double-cue condition |
| $d_{m-d}$ | Euclidean distance between the response centroids in the self-motion condition and the double-cue condition |

2022). This term should be distinguished from "sensory optimal weight" ($w_{opt,sen}$), which is calculated from the estimated values of the sensory noise levels of the cues in the cognitive modeling (see the following Cognitive Modeling section).

$$w_{opt,beh} = rr$$

The observed weights assigned to spatial cues are calculated using the relative distances between the response centroid in the double-cue condition and the response centroids in the two single-cue conditions. The observed weight for landmarks ($w_{obs}$) is calculated as:

$$w_{obs} = d_{m-d}/(d_{l-d} + d_{m-d})$$

where $d_{m-d}$ represents the Euclidean distance between the response centroid in the self-motion condition and the response centroid in

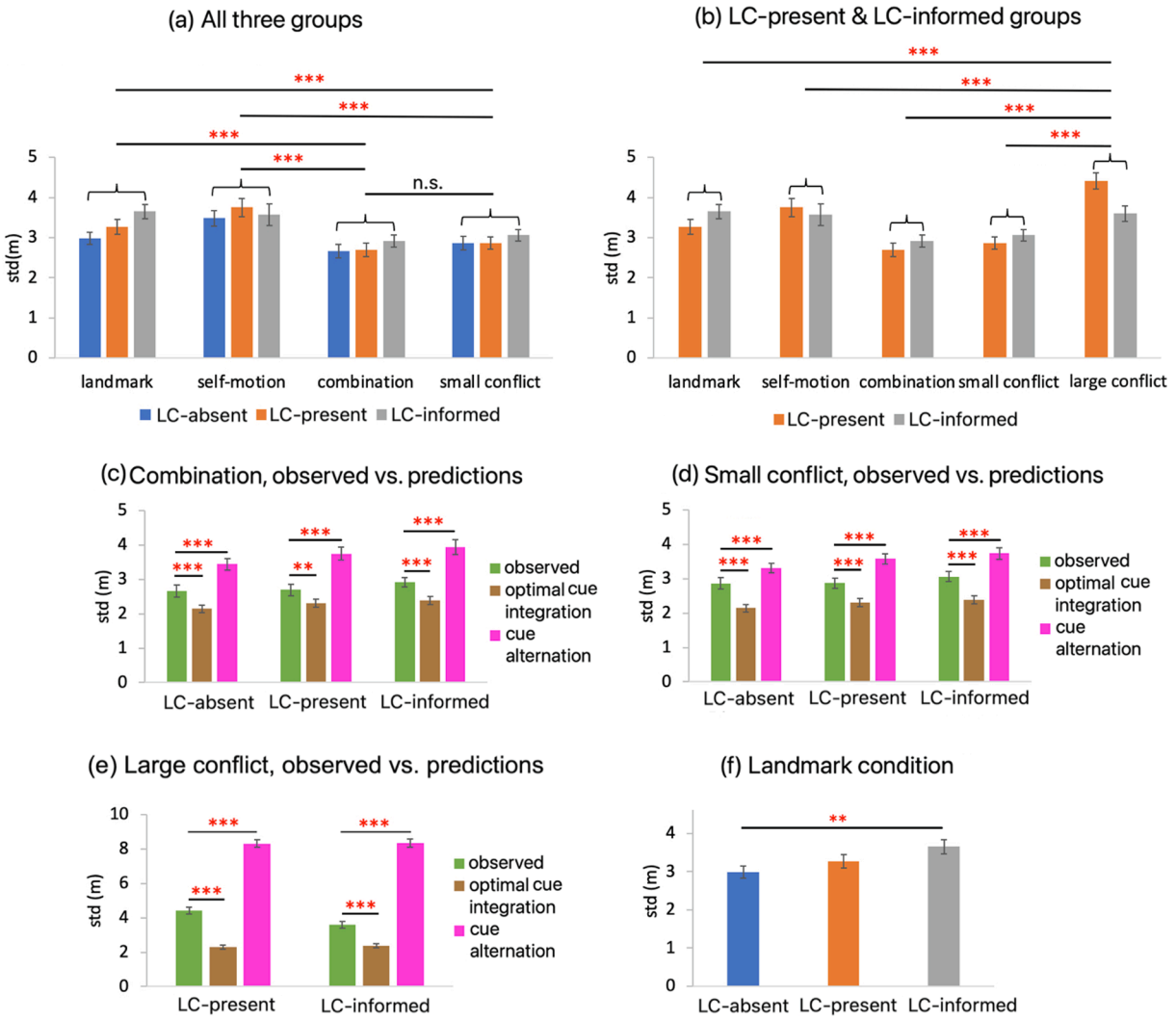



**Fig. 3. Behavioral results of response variability (a)** Response variability across the four cue conditions (landmark, self-motion, combination, small conflict) in the three groups (LC-absent, LC-present, LC-informed). **(b)** Response variability across the five cue conditions (landmark, self-motion, combination, small conflict, and large conflict) in the LC-present and LC-informed groups. **(c)** Response variability in the combination condition (light green bars), along with predictions from the MLE cue integration model (optimal cue integration, brown bars) and the cue alternation model (purple bars) in the three groups. **(d)** Response variability in the small conflict condition (light green bars), along with predictions of the MLE cue integration model (optimal cue integration, brown bars) and the cue alternation model (purple bars) in the three groups. **(e)** Response variability in the large conflict condition (light green bars), along with predictions of the MLE cue integration model (optimal cue integration, brown bars) and the cue alternation model (purple bars) in the LC-present and LC-informed groups. **(f)** Response variability in the landmark condition across the three groups. "n.s." indicates $p > 0.05$; "**" indicates $p < 0.01$; "***" indicates $p < 0.001$. Error bars represent standard error of the mean (S.E.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the double-cue condition, and $d_{l\_d}$ represents the Euclidean distance between the response centroid in the landmark condition and the response centroid in the double-cue condition.

Besides the MLE model of cue integration, we also compared participants' responses to the cue alternation model, which posits that participants alternated between landmarks and self-motion cues on a trial-by-trial basis, with the ratio determined by the optimal weight ($rr$) (Nardini et al., 2008). Its prediction on double-cue response variability is:

$$S_{alt,d} = \sqrt{\begin{array}{c}(1-rr)\left(\overline{X}_m^2 + S_m^2\right) + rr\left(\overline{X}_l^2 + S_l^2\right) - \\ \left((1-rr)\overline{X}_m + rr\overline{X}_l\right)^2\end{array}}$$

where $\overline{X}_l$ represents the response centroid in the landmark condition, and $\overline{X}_m$ represents the response centroid in the self-motion condition.

### 3.2. Behavioral results

Behavioral analyses were conducted using Matlab_R2021a and JASP (Version 0.17.1; JASP Team, 2023). Statistical outliers were identified within each group for each variable, i.e., > 3rd quartile + 3*IQR or < 1st quartile – 3*IQR. Outliers were winsorized to the nearest inlier within each group (Reifman & Keyton, 2010).

#### 3.2.1. Response variability

Because the LC-absent group did not experience the large conflict condition, first, we analyzed response variability in the other four cue conditions across groups (Fig. 3a). Observed response variability was submitted to a two-way mixed ANOVA, with cue condition (landmark vs. self-motion vs. combination vs. small conflict) as a within-participant independent variable and group (LC-absent vs. LC-present vs. LC-informed) as a between-participant independent variable. The only significant effect was the main effect of cue condition ($F(3,414) = 25.470$, $p < 0.001$, $\eta_p^2 = 0.156$, $BF_{inclusion} = 5.816*10^{11}$). Planned comparisons revealed that both the combination condition and the small conflict condition showed the cue integration effect: response variability in the combination was smaller than both the landmark condition ($t(414) = 5.532$, $p < 0.001$) and the self-motion condition ($t(414) = 7.874$, $p < 0.001$); response variability in the small conflict condition was smaller than both the landmark condition ($t(414) = 3.774$, $p < 0.001$) and the self-motion condition ($t(414) = 6.116$, $p < 0.001$). There was no significant difference between the combination condition and the small conflict condition ($t(414) = 1.797$, $p = 0.073$). The main effect of group was not significant ($F(3,138) = 1.395$, $p = 0.251$, $\eta_p^2 = 0.020$). Neither was the interaction between cue condition and experiment ($F(6,414) = 1.382$, $p = 0.237$, $\eta_p^2 = 0.020$).

As shown in Fig. 3c&d, in each group, although both the combination condition and the small conflict condition showed the cue integration effect (i.e., lower response variability than the two single-cue conditions), their response variabilities were larger than the prediction of the MLE cue integration model (i.e., optimal cue integration), indicating non-optimal cue integration in these two double-cue conditions (ts > 2.9, ps < 0.006, $BFs_{10} > 7.5$). As expected, in each group, the response variability was lower than the prediction of the cue alternation model in both the combination condition and the small conflict condition (ts > 3.5, ps < 0.001, $BFs_{10} > 80$).

Next, we analyzed the LC-present group and the LC-informed group together by including the large conflict condition in the analysis (Fig. 3b). The main effect of cue condition was significant ($F(4, 316) = 23.188$, $p < 0.001$, $\eta_p^2 = 0.227$, $BF_{inclusion} = 1.820 * 10^{14}$). In the large conflict condition, participants did not show the cue integration effect, as its response variability was significantly larger than both single-cue conditions (vs. landmark, $t(316) = 3.574$, $p < 0.001$; vs. self-motion, $t(316) = 2.357$, $p = 0.019$). As expected, the response variability in the large conflict condition was greater than the prediction of the MLE cue integration model (i.e., optimal cue integration; ts > 7, ps < 0.001, $BFs_{10} > 1*10^6$, Fig. 3e). Nevertheless, the response variability in the large conflict condition in each group was substantially smaller than the prediction of the cue alternation model (ts > 14, ps < 0.001, $BFs_{10} > 7*10^{14}$, Fig. 3e). Actually, in each group, the observed response variability in the large conflict condition was considerably closer to the prediction of the MLE cue integration model than that of the cue alternation model.

Although the preceding omnibus statistical tests did not reveal any significant influences of group on response variability, given that we had a prior hypothesis that the response variability in the landmark condition would be affected by landmark instability (Chen et al., 2017), we conducted a separate analysis to compare the three groups on response variability in the landmark condition (Fig. 3f). The main effect of group was significant ($F(2,138) = 4.233$, $p = 0.016$, $\eta_p^2 = 0.058$, $BF_{inclusion} = 2.465$). Follow-up post-hoc comparisons showed that the response variability was significantly larger in the LC-informed group than the LC-absent group ($t(138) = 2.910$, $p = 0.004$); other differences were not significant ($|ts| < 1.6$, ps > 0.11). These results indicate that heightened cue conflict, combined with explicit awareness of landmark instability, increased response variability associated with landmarks.

#### 3.2.2. Cue weight

First, we analyzed cue weight in the small conflict condition by considering all three groups together, as the LC-absent group did not include the large conflict condition. Observed landmark weights were submitted to ANOVA, with group as a between-participant independent variable. As shown in Fig. 4a, The main effect of group was significant ($F(2,138) = 5.215$, $p = 0.007$, $\eta_p^2 = 0.070$, $BF_{inclusion} = 5.248$). Follow-up comparisons showed that the observed landmark weight was significantly lower in the LC-informed group than the LC-absent group ($t(138) = 2.878$, $p = 0.005$) and the LC-present group ($t(138) = 2.809$, $p = 0.006$), whereas there was no significant difference between the LC-absent group and the LC-present group ($t(138) = -0.182$, $p = 0.856$). These results

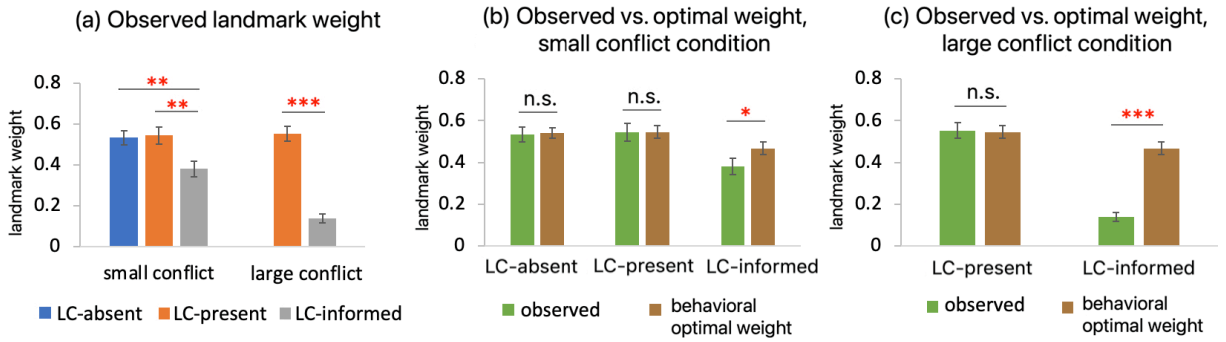



**Fig. 4. Behavioral results of cue weight (a)** Observed weight assigned to landmarks in the two conflict conditions (small conflict and large conflict) across the three groups (LC-absent, LC-present, and LC-informed). Note that the LC-absent group did not include the large conflict condition. **(b)** Observed weight assigned to landmarks in the small conflict condition (light green bars), compared to the behavioral optimal weight (brown bars) across the three groups. **(c)** Observed weight assigned to landmarks in the large conflict condition (light green bars), compared to the behavioral optimal weight (brown bars) in the LC-present and LC-informed groups. "n.s." indicates $p > 0.05$; "*" indicates $p < 0.05$; "**" indicates $p < 0.01$; "***" indicates $p < 0.001$. Error bars represent standard error of the mean (S.E.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

indicate that while heightened cue conflict alone did not reduce the weight navigators assigned to the landmark in the small conflict condition, the addition of explicit awareness of landmark instability achieved this effect. This finding is particularly intriguing, because the experimental setup in the small conflict condition was exactly the same between the three groups; the key difference lies in whether participants also experienced large conflict trials at the same time and had explicit awareness of landmark instability.

We wondered whether the significant differences in observed landmark weight between groups could be simply explained by the aforementioned significant differences in the response variability in the landmark condition between the groups (Fig. 3f), because reduced landmark precision predicts lower behavioral optimal weight assigned to landmarks according to the MLE principles. As shown in Fig. 4b, observed landmark weight was significantly lower than the behavioral optimal weight in the LC-informed group ($t(39) = -2.519$, $p = 0.016$, $BF_{10} = 2.742$), whereas the two variables did not differ from each other in either the LC-absent group ($t(59) = -0.167$, $p = 0.868$, $BF_{10} = 0.143$) or the LC-present group ($t(40) = -0.053$, $p = 0.958$, $BF_{10} = 0.169$). These results indicate that the differences in the observed landmark weights could not be simply explained by the differences in the landmark response variability between the groups.

Next, we analyzed the weight data in the small conflict condition and the large conflict condition by considering the LC-present group and the LC-informed group together (Fig. 4a). The interaction between group and cue condition was significant ($F(1,79) = 20.374$, $p < 0.001$, $\eta_p^2 = 0.205$, $BF_{inclusion} = 3751$). Follow-up t tests showed that the observed landmark weight was significantly lower in the large conflict condition than in the small conflict condition in the LC-informed group ($t(79) = 6.115$, $p < 0.001$), but did not differ between the two conflict conditions in the LC-present group ($t(79) = -0.232$, $p = 0.817$). Furthermore, in the LC-present group, the observed weight in the large conflict condition did not differ from the behavioral optimal weight ($t(40) = 0.157$, $p = 0.876$, $BF_{10} = 0.171$), whereas in the LC-informed group, the observed weight in the large conflict condition was significantly lower than the behavioral optimal weight ($t(39) = -10.518$, $p < 0.001$, $BF_{10} = 1.347*10^{10}$) (Fig. 4c).

### *3.2.3. Summary of behavioral results*

Overall, increasing cue conflict alone did not cause any discernable changes in participants' behavior, as the LC-absent group and the LC-present group did not differ from each other in any behavioral aspects under examination. The LC-informed group differed from the LC-absent group and the LC-present group in several aspects, indicating that explicit awareness of landmark instability was necessary to cause changes in participants' navigation behavior. Specifically, the LC-informed group showed larger response variability in the landmark condition than the LC-absent group, implying that the participants informed about landmark instability performed worse when exclusively relying on landmarks for navigation. In addition, the LC-informed group assigned lower weights to landmarks than the other two groups. Importantly, the observed landmark weight was lower than the behavioral optimal weight in the LC-informed group but did not differ from the behavioral optimal weight in the other two groups, indicating that lower weight assigned to landmarks in the LC-informed group could not be simply explained by the increased response variability in the landmark condition in this group.

Despite these differences, all three groups exhibited cue integration effects in both the combination and small conflict conditions, as response variability was lower in these two double-cue conditions than in the two single-cue conditions. However, response variability was larger than what the MLE cue integration model predicted, indicating suboptimal cue integration. The large conflict condition did not show cue integration effects, as its response variability was larger than the two single-cue conditions. However, the response variability in the large conflict condition was considerably lower than predicted by the cue alternation model, suggesting a mixture of cue integration and cue alternation strategies. This finding is consistent with previous findings on spatial cue combination (Chen et al., 2017; Qi & Mou, 2024; Sjolund et al., 2018).

## 4. Cognitive modeling

### 4.1. Modeling methods

We tested two cognitive models of the task: A Bayesian causal inference (BCI) model used in multisensory perception tasks (Körding et al., 2007; Wozny et al., 2010) and a sensory disparity model, which did not incorporate any prior information. We adopted the maximum likelihood estimation approach for model fitting, as detailed in Supplemental information (Section A).

#### 4.1.1. BCI model

Here, we provide a conceptual overview of the BCI model, with detailed descriptions available in Supplemental information (Section A).

The BCI model involves four different information sources: the prior belief in sensory cues originating from a common source ($p_{cc,pr}$), the prior knowledge about possible locations of the target ($N\left(\mu_{pr}, \sigma_{pr}\right)$, target location prior), the sensory information from landmark cues ($N(\mu_l, \sigma_l)$, sensory measurement distribution of landmarks), and the sensory information from self-motion cues ($N(\mu_m, \sigma_m)$, sensory measurement distribution of self-motion cues).

First, this model utilizes these information sources to make the causal structure judgment (i.e., whether the sensory inputs stem from a common cause, or equivalently, whether the cues are in congruency). In the context of our navigation task, the causal structure judgment corresponds to whether the optic flow information was consistent with the landmark in indicating the target location (Fig. 2b). Same-cause judgment means participants determined that the two cues were consistent with each other, and any disparities between their sensory inputs were due to sensory noise. Conversely, different-cause judgment means participants determined that the two cues were inconsistent with each other, and any disparities between their sensory inputs were due to the fact that they did not indicate the same target location, that is, not sharing a common spatial source.

Next, the model combines or segregate spatial cues based on the causal judgment. When a different cause judgment is made ($C = 2$), the cue segregation sub-model is implemented. Because the experimental task only required one response rather than two, we implemented a cue alternation strategy (de Winkel et al., 2017, 2018), which has either been demonstrated (Nardini et al., 2008) or implicated (Chen et al., 2017) in human spatial navigation. First, the likelihood distribution indicated by the landmarks is selected with probability $w_{dc}$ and the likelihood distribution indicated by self-motion is selected with probability $(1 - w_{dc})$. Next, the selected likelihood distribution is optimally integrated with the spatial location prior distribution to generate the posterior probability distribution according to the MLE principles, with weights determined by their relative reliabilities (Bromiley, 2013; McNamara & Chen, 2022, Appendix A).

When the cues are judged to originate from a common cause, the cue integration sub-model is implemented. The three location-relevant pieces of information – the likelihood distribution of landmarks $N(x_l, \sigma_l)$, the likelihood distribution of self-motion cues $N(x_m, \sigma_m)$, and the prior distribution of target location $N(\mu_{pr}, \sigma_{pr})$ – are integrated in a statistically optimal manner, each weighted by its relative reliability. To be parallel to the two-step cue-weighting scheme in the cue segregation sub-model, here, this optimal cue integration process can be conceived to consist of two steps too. In the first step, the two likelihood distributions of the two cue types are optimally integrated to generate the joint likelihood distribution according to the MLE principles. The weight assigned to the landmark in this step is equal to the inverse ratio of sensory noise levels between the two cues, $w_{cc} = \sigma_m^2/(\sigma_l^2 + \sigma_m^2)$. $w_{cc}$ is considered as the sensory weight assigned to landmarks and conceptually parallel to $w_{dc}$ in the cue segregation sub-model, because it is unrelated to the prior distribution on target location. Hence, $w_{cc}$ is also identified as the sensory optimal weight ($w_{opt,sen}$). Note that the sensory optimal weight is different from the behavioral optimal weight mentioned in the behavioral analysis ($w_{opt,beh}$). The behavioral optimal weight is calculated from the observed responses in the single-cue conditions, which reflect mixed influences of sensory noise, target location prior, and motor noise. By contrast, the sensory optimal weight is recovered via the modeling analysis, and is determined by the sensory noise levels of the two cues (Aston et al., 2022). In the second step, the joint likelihood distribution is optimally integrated with the prior distribution of target location to generate the posterior distribution.

Finally, three decisions strategies are constructed to combine estimates from the cue segregation sub-model and cue integration sub-model, based on the posterior probabilities of common-cause and different-cause judgments (Koerding et al., 2007; Wozny et al., 2010). In the model averaging strategy, to obtain the final positional estimate, the positional estimates from the cue integration sub-model and the cue segregation sub-model are linearly combined, with weights equal to the posterior probabilities of the common-cause and different-cause judgments, respectively. This strategy minimizes the uncertainty of the positional estimate. The model selection strategy adopts a winner-take-all approach. When the posterior probability of a common cause exceeds a criterion, the positional estimate from the cue integration sub-model is selected; otherwise, the positional estimate from the cue segregation sub-model is chosen. The criterion is set to 0.5, which entails that the sub-model with the larger posterior probability is selected. In the probability matching strategy, the observer alternates between the positional estimates of the cue integration sub-model and the cue alternation sub-model, with the alternation ratio determined by the posterior probabilities of the common-cause and different-cause judgments.

Motor noise was also modeled. In each trial, a number was randomly sampled form an unbiased normal distribution $N(0, \sigma_{mt})$, which was then added to the final positional estimate of the target location.

To facilitate the model optimization process, we minimized the total number of free parameters by estimating the sensory measurement distributions ($N(\mu_l, \sigma_l)$ $N(\mu_m, \sigma_m)$) from participants' responses recorded in the single-cue conditions, instead of allowing all four parameters freely to vary. We assumed that the prior distribution of target location $N(\mu_{pr}, \sigma_{pr})$, which is used for the causal

structure judgment and integrated into the positional estimate in the double-cue conditions, also functions in the two single-cue conditions. $\mu_l$ and $\mu_m$ can be unambiguously expressed as functions of other model parameters and response variables. However, this is not true for $\sigma_l$ and $\sigma_m$. Therefore, $\sigma_l$ and $\sigma_m$ were still included as model parameters.

For each of the three decision strategies, there are seven free parameters: the prior belief in the probability that the sensory inputs originate from the same source ($p_{cc,pr}$), the standard deviation ($\sigma_{pr}$) and the mean ($\mu_{pr}$) of the prior distribution of target location, the weight assigned to landmarks for the different-cause judgment ($w_{dc}$), the standard deviations of the sensory measurement distributions for landmarks ($\sigma_l$) and self-motion cues ($\sigma_m$), and motor noise ($\sigma_{mt}$).

#### 4.1.2. Sensory disparity model

Similar to the BCI model, the sensory disparity model incorporates sensory noise by sampling sensory measurements from spatial cues on a trial-by-trial basis. However, unlike the BCI model, this model does not include any form of prior knowledge. Therefore, comparing the BCI model to the sensory disparity model allows us to evaluate whether prior knowledge – the prior distribution of target locations and the prior belief about causal structure – influences navigation behavior.

In the sensory disparity model, causal judgments are based on the absolute distance between the two sensory measurements from the two spatial cues. In each trial, two sensory measurements are sampled, each associated with one of the two cue types: $x_l$ for landmarks, and $x_m$ for self-motion cues. First, we normalize the absolute distance between the two sensory measurements by the pooled sensory noise level from the two cue types (i.e., the pooled standard deviation calculated from $\sigma_l$ and $\sigma_m$).

$$d = \frac{|x_l - x_m|}{\sqrt{\frac{\sigma_l^2+\sigma_m^2}{2}}}$$

Next, the normalized distance is transformed to inverse distance:

$$d_{inv} = \frac{2}{1 + e^{t \times d}}$$

where $t$ is a scaling parameter that controls the rate of change. The larger the magnitude of $t$, the less likely the common-cause judgment. The variable $d_{inv}$ stands for inverse distance and ranges from 0 to 1. $d_{inv}$ is used as a counterpart of posterior probability of common cause in the BCI model. $d_{inv} > 0.5$ leads to the same-cause judgment and $d_{inv} \leq 0.5$ leads to the different-cause judgment. According to the model, when the sensory measurements from two cues are perfectly aligned with each other (i.e., $d$ is equal to 0), $d_{inv}$ is equal to 1, the highest probability of common-cause. Conversely, when the two sensory measurements are extremely far away from each other (i.e., $d$ is extremely large), $d_{inv}$ is equal 0, the lowest probability of common-cause. Hence, this process of causal structure inference reflects sensory uncertainties, because the standard deviations of the sensory measurement distributions ($\sigma_l$ and $\sigma_m$) are incorporated into the calculation of $d_{inv}$: the larger the sensory uncertainties, the less likely the different-cause judgment. However, this process does not incorporate any form of prior information.[1]

The following steps are exactly the same as the BCI model, except that in each cue condition the likelihood distribution(s) is(are) no longer integrated with the target location prior distribution. There are five free parameters in total: scaling parameter ($t$), weight assigned to landmark for the different-cause judgment ($w_{dc}$), standard deviations of the sensory measurement distributions for landmarks ($\sigma_l$) and self-motion cues ($\sigma_m$), and motor noise ($\sigma_{mt}$). Hence, the sensory disparity model has two fewer free parameters than the BCI model.

### 4.2. Modeling results

#### 4.2.1. Model comparison

We compared the model fit between the BCI model and the sensory disparity model. Since we could not reliably recover the decision strategy used to generate the synthetic data among all three strategies (Supplemental information, Section B), we considered all three decision strategies together. AIC was highly correlated among the three decision strategies within each model (rs $> 0.85$), so we compared the mean AIC averaged across decision strategies. Given that the mean AIC differences sometimes deviated from the normal distribution and had outliers (see scatterplots in Fig. 5), we conducted the nonparametric Wilcoxon signed-rank tests to compare model fit between the two models.

The results are depicted in Fig. 5. When analyzing all three groups together, mean AIC was significantly lower for the BCI model than the sensory disparity model ($z = -6.346$, $p < 0.001$, $BF_{10} = 279495$). The pattern of results was consistent across groups. In the LC-absent group, mean AIC was numerically lower for the BCI model than the sensory disparity model, although the difference did not reach statistical significance ($z = -1.016$, $p = 0.311$, $BF_{10} = 0.473$) (Fig. 5a, left panel). In the other two groups, mean AIC was significantly lower for the BCI model than the sensory disparity model (LC-present, $z = -4.995$, $p < 0.001$, $BF_{10} = 57442$, Fig. 5b, left

[1] An alternative approach for making the causal structure judgment is to set a criterion on the absolute distance between the two sensory measurements directly (Badde et al., 2020). For example, when $|x_l - x_m|$ is larger than the criterion, the different-cause judgment is made. However, this approach is only suitable for the model selection strategy, but not suitable for the model averaging strategy or the probability matching strategy, because the latter two strategies require a probability-like statistic to linearly combine or sample from the two spatial estimates from the cue integration sub-model and the cue segregation sub-model.

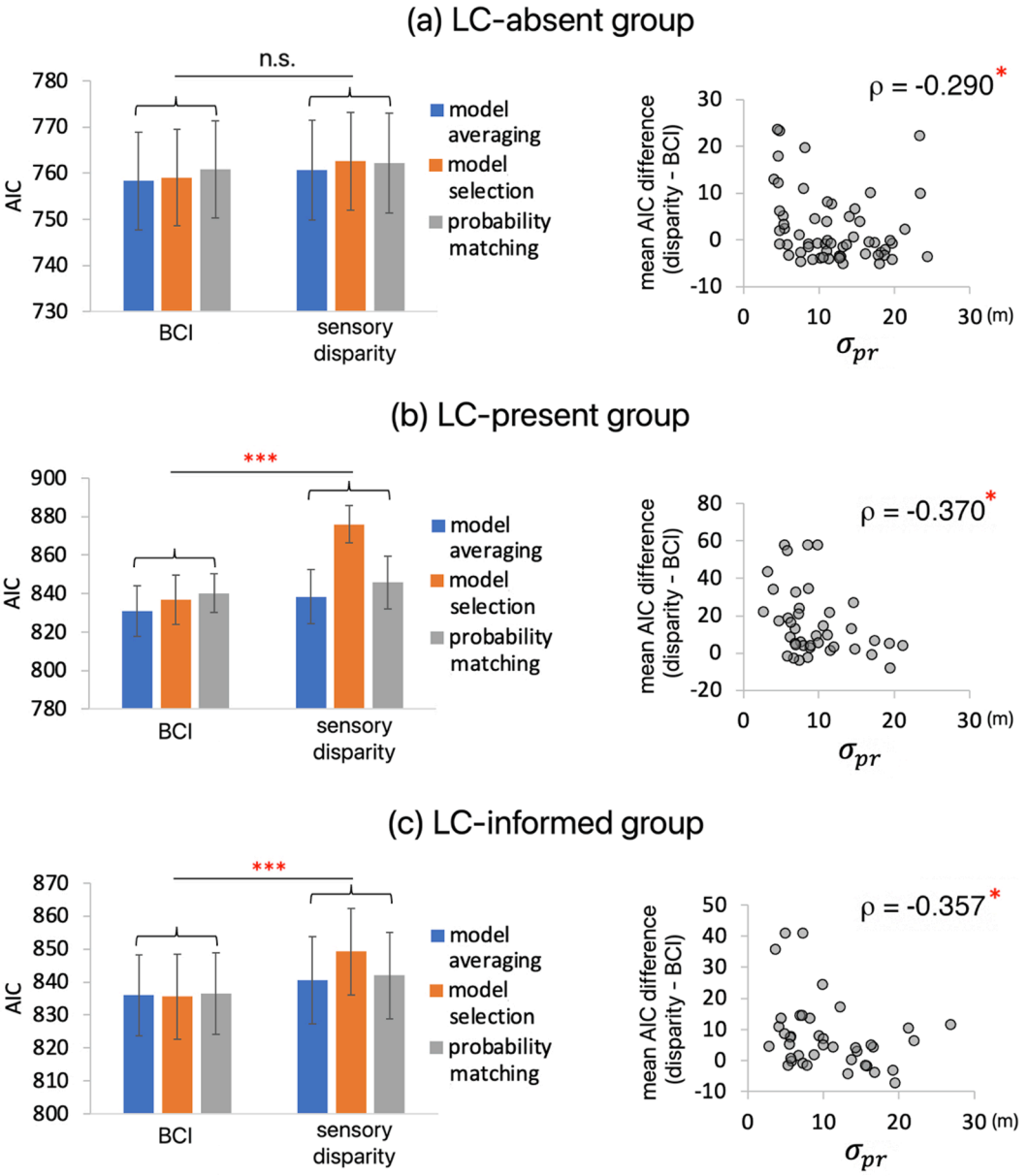



**Fig. 5. Comparisons of goodness-of-fit between BCI model and sensory disparity model** Results are displayed separately for the LC-absent group **(a)**, the LC-present group **(b)**, and the LC-informed group **(c)**. For each group, the left panel shows AIC for the BCI model and the sensory disparity model across the three decision strategies. AIC was first averaged across the three decision strategies, which was then compared between the two models. For each group, the right panel shows the across-participant correlation between the mean AIC difference between the two models (sensory disparity − BCI) and the estimated standard deviation of the prior knowledge about target location ($\sigma_{pr}$). Spearman rank correlation coefficients ($\rho$) are displayed. "*" indicates $p < 0.05$. Error bars represent standard deviation of the mean (S.E.).

panel; LC-informed, $z = -4.180$, $p < 0.001$, $BF_{10} = 7950$, Fig. 5c, left panel). These results suggest that the advantage of the BCI model was more pronounced when the large conflict condition was included.

Further analyses showed that model fit difference negatively correlated with the standard deviation of prior knowledge about target location (aggregated across groups: Spearman $\rho = -0.398$, $p < 0.001$; LC-absent group: $\rho = -0.318$, $p = 0.014$; LC-present group: $\rho = -0.371$, $p = 0.017$; LC-informed group: $\rho = -0.358$, $p = 0.024$), meaning that the less precise the prior knowledge about target location, the smaller the relative advantage of the BCI model over the sensory disparity model (Fig. 5, right panels).

Similar results were obtained if we used BIC as the model fit index (across groups, $z = -2.316$, $p = 0.021$, $BF_{10} = 15.030$).

To validate these findings, we performed a four-fold cross-validation analysis (Hastie, 2009), which reduces risks of over-fitting or under-fitting compared to conventional methods like AIC (Browne, 2000). For each participant, we split the data into four equal parts,

which contained the $(4n + 1)^{th}$ trials, the $(4n + 2)^{th}$ trials, the $(4n + 3)^{th}$ trials, and the $(4n + 4)^{th}$ trials (n is an integer), respectively. In each round, three quarters served as the training data and the remaining quarter served as the test data. First, we fitted the model to training data, obtaining the best-fitting parameters. Next, we used these parameter estimates to generate simulated data, with 5000 responses for each cue condition. We applied the Kernel smoothing function ("ksdensity" command in Matlab) to these simulated responses, resulting in a response density distribution for each cue condition. After that, we calculated log-likelihood of the test data based on the response density distributions, which were summed across all cue conditions. This procedure was repeated four times, with each quarter serving as the test data once. The log-likelihood of the test data was summed cross the four rounds and then averaged across decision strategies, serving as an index of model fit.

The cross-validation results showed that the BCI model provided a superior fit compared to the sensory disparity model (median log-likelihood = -618.063 vs. −619.279; Wilcoxon signed-rank test, $z = 4.793$, $p < 0.001$, $BF_{10} = 106685$; all groups included). Additionally, the difference in log-likelihood was negatively correlated with the standard deviation of prior knowledge about target location ($\sigma_{pr}$) ($\rho = -0.360$, $p < 0.001$). Furthermore, the BCI model showed better fit than the sensory disparity model when the decision strategies were considered separately (model averaging, $z = 5.239$, $p < 0.001$, $BF_{10} = 1483$; model selection, $z = 7.462$, $p < 0.001$, $BF_{10} = 454342$; probability matching, $z = 5.899$, $p < 0.001$, $BF_{10} = 1544$; all groups considered). Finally, the best-fitting decision strategy from the BCI model (i.e., model averaging) outperformed all decision strategies from the sensory disparity model (vs. model averaging, $z = 5.239$, $p < 0.001$, $BF_{10} = 1483$; vs. model selection, $z = 7.799$, $p < 0.001$, $BF_{10} = 591527$; vs. probability matching, $z = 6.089$, $p < 0.001$, $BF_{10} = 3612$). In sum, the cross-validation analysis yielded results consistent with the conventional AIC method, supporting the robustness of our model comparison findings.

Taken together, our results showed that the BCI model outperformed the sensory disparity model in fitting the data. The BCI model's advantage was greater when the precision of prior knowledge about target location was higher. These findings indicate that participants utilized prior information when navigating in a multi-cue environment.

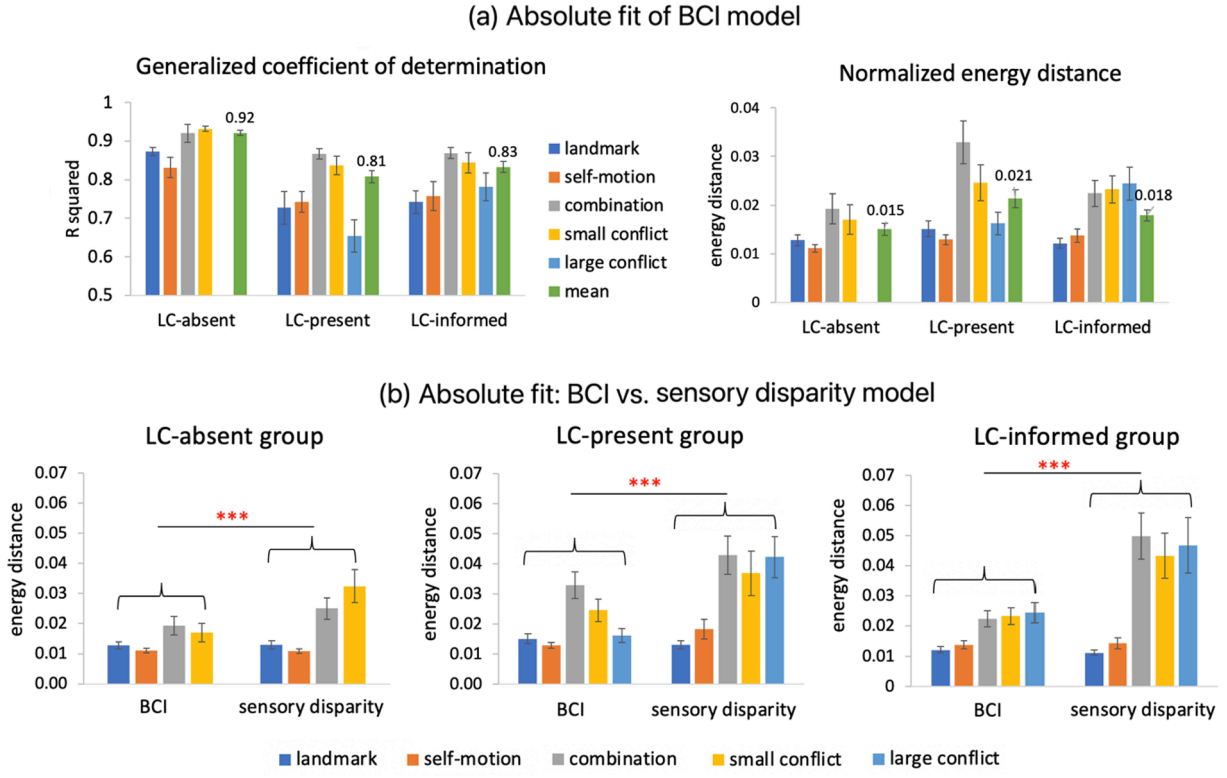



**Fig. 6. Absolute fit for BCI model and sensory disparity model (a)** Two indices of absolute fit – the generalized coefficient of determination (left panel) and the normalized energy distance (right panel) – are displayed for the BCI model across the three experimental groups. These two indices measure the similarity between the actual responses and the simulated responses. **(b)** Comparison of the BCI model and the sensory disparity model in terms of normalized energy distance. Simulated data were generated by sampling synthetic responses from the three decision strategies (model averaging, model selection, and probability matching), with proportions determined by their relative marginal likelihoods. Both measurements range from 0 to 1. Low values of normalized energy distance and high values of generalized coefficient of determination indicate good model fit. These two measurements are displayed for each cue condition, along with mean values across all cue conditions. Error bars represent standard error of the mean (S.E.).

#### 4.2.2. Absolute fit of BCI model

The preceding section evaluated the comparative fit of the BCI model relative to the sensory disparity model. To evaluate the absolute fit of the BCI model, we calculated the generalized coefficient of determination, which measures the proportion of variance in responses explained by the model compared to a null model (Nagelkerke, 1991). The null model was set as a uniform distribution, bounded by the minimal and maximum response values across all participants in the group. As shown in Fig. 7a (left panel), the BCI model explains a high proportion of response variance: mean (SD) = 92 % (5.02 %), 81 % (9.88 %), and 83 % (8.78 %) for the LC-absent, LC-present, and LC-informed groups, respectively.

Additionally, we calculated "normalized energy distance", which quantifies a model's absolute fit by measuring the similarity between actual and simulated response distributions (Rizzo and Székely, 2016). This index ranges from 0 (identical distributions) to 1 (completely different distributions). To obtain the simulated response distributions, we simulated 5000 responses per cue condition for each decision strategy. Next, we sampled from the simulated responses of the three decision strategies, with proportions equal to their relative marginal likelihoods (see Section 4.2.3 for calculation). This approach is consistent with the Bayesian model averaging approach used for model parameter estimation and model predictions (see Section 4.2.3 and Section 4.2.4).

As shown in Fig. 6a (right panel), the normalized energy distance was quite low across groups (mean (SD) = 0.015 (0.01), 0.020 (0.012), 0.019 (0.007) for the LC-absent, LC-present, and LC-informed groups, respectively), indicating strong alignment between model predictions and observed data.

We also leveraged the absolute fit measurement – normalized energy distance – to examine why the BCI model fit the data better than the sensory disparity model. As expected, the BCI model showed smaller normalized energy distance than the sensory disparity model across groups (Fig. 6b): LC-absent, $z = -4.689$, $p < 0.001$, $BF_{10} = 496$; LC-present, $z = -4.011$, $p < 0.001$, $BF_{10} = 1720$; LC-informed, $z = -4.637$, $p < 0.001$, $BF_{10} = 3052$. Importantly, the superior fit of the BCI model mainly manifests in the double-cue conditions.

Finally, we visualized the model fit by plotting raw responses of individual participants alongside the BCI model predictions. Fig. 7 depicts two example participants from each group, demonstrating strong qualitative alignment between observed and predicted responses. Notably, the BCI model captures the bimodal distributions that emerge in the large conflict condition in some participants (Fig. 7, panels b.1, b.2, and c.1). Additional examples are provided in Supplemental information (Section F).

Collectively, both quantitative metrics and qualitative visualizations indicate the BCI model's strong fit to the data in the current experiment.

#### 4.2.3. Model fit evaluation for BCI model

The preceding two sections have demonstrated that the BCI model fit the data well. Here, to further evaluate the goodness-of-fit of the BCI model, we examine whether the BCI model could predict key variables of the actual data and replicate key aspects of the behavioral findings.

For each decision strategy, we obtained the predictions for key response variables: response centroid and variability in each double-cue condition, cue weights in each conflict condition. Next, we calculated the Bayesian average of the predictions on these variables by linearly combining the predictions of the three decisions strategies, with weights determined by their respective posterior probabilities ($p(M)$). This Bayesian model averaging approach was also adopted for model parameter estimation (see a later section "Influences of Cue Conflict on Model Parameters").

We employed the Bayesian model averaging approach for both model prediction and parameter estimation, as it typically outperforms approaches based on a single model (Raftery, 1995). By integrating over a range of possible parameter values, rather than relying on a single estimate, this method enhances out-of-sample predictions. This process accounts for model uncertainty derived from the Bayesian information criterion (BIC). BIC gives a good approximation of model evidence (Raftery, 1995), resulting in more robust and generalizable predictions (Raftery et al., 1996). Furthermore, the model recovery analysis revealed that the three decision strategies could not be reliably distinguished with our data (Supplemental information, Section B), underscoring the need to combine these strategies for parameter estimation and model prediction.

With no priors assumed on model evidence, the posterior probability of a decision strategy $p(M)$ equals to its marginal likelihood, which can be calculated from BIC as follows:

$$p(M) = exp(-0.5 \times n - 0.5 \times BIC)\ exp(-BIC)$$

$$BIC = -2LL_{sum} + plog(n)$$

in which $n$ represents the total number of trials ($\approx 144$), and $p$ represents the total number of free parameters ($= 7$). The marginal likelihood of a decision strategy is proportional to the exponential of its negative BIC. In both model prediction and parameter estimation, the weights of the three decision strategies correspond to their relative marginal likelihoods.

Model fit is depicted in Figs. 8-10 for the three groups separately. Outliers were identified within each group for each variable, i.e., $>$ 3rd quartile + 3*IQR or $<$ 1st quartile – 3*IQR. Outliers were winsorized to the nearest inlier within each group (Reifman & Keyton, 2010).

First, we examined how well the BCI model predicted the aforementioned key variables of the behavioral data for each group, by comparing the simulated values to the observed values of these variables (Figs. 8-10, panels a, b, and c). The results showed that our models fit the data considerably well by capturing major trends of the data in terms of group means (left panels). In addition, the simulated values showed strong correlations with the observed values for the key variables ($rs > 0.6$, $ps < 0.01$) (middle and right

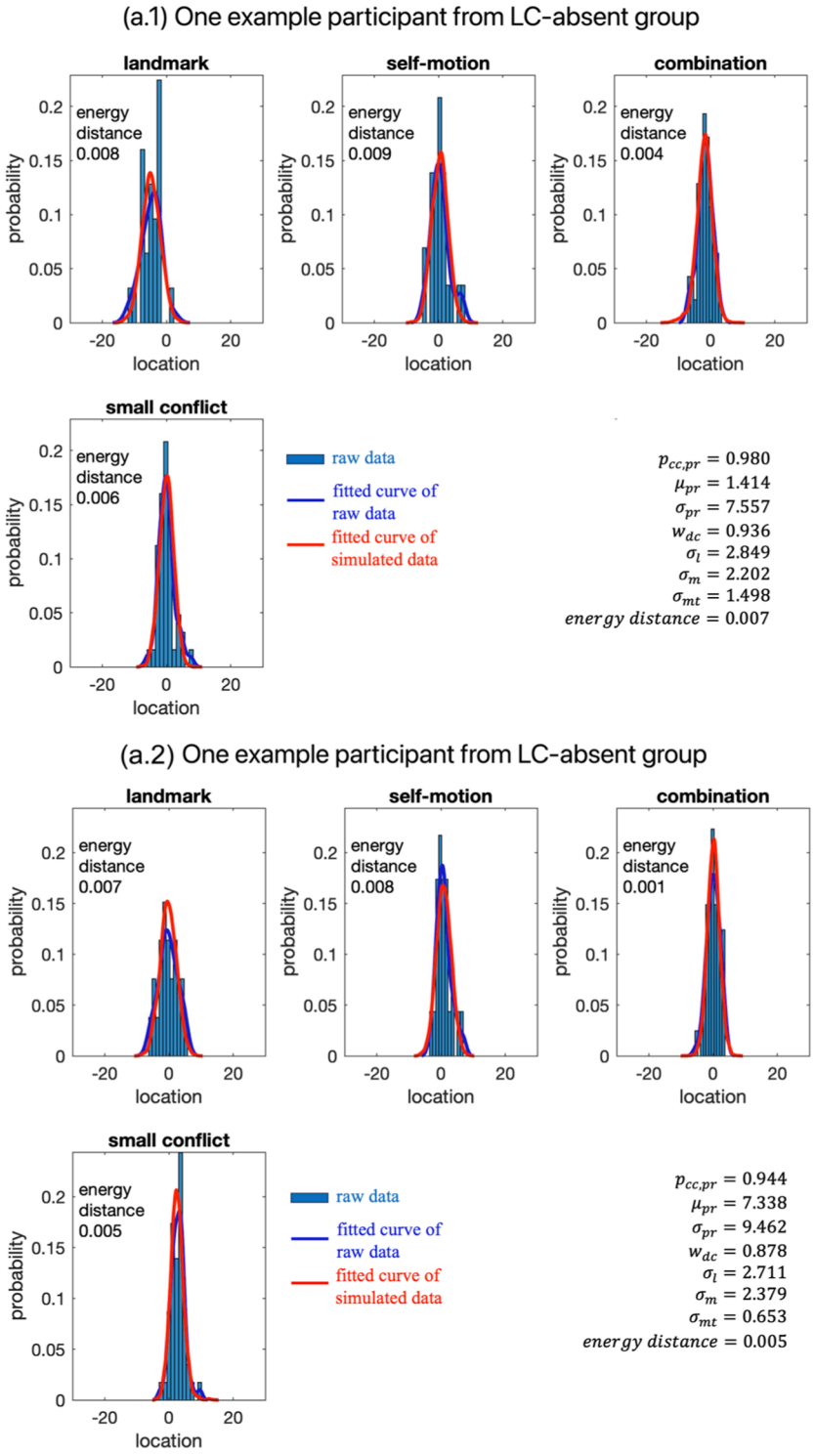
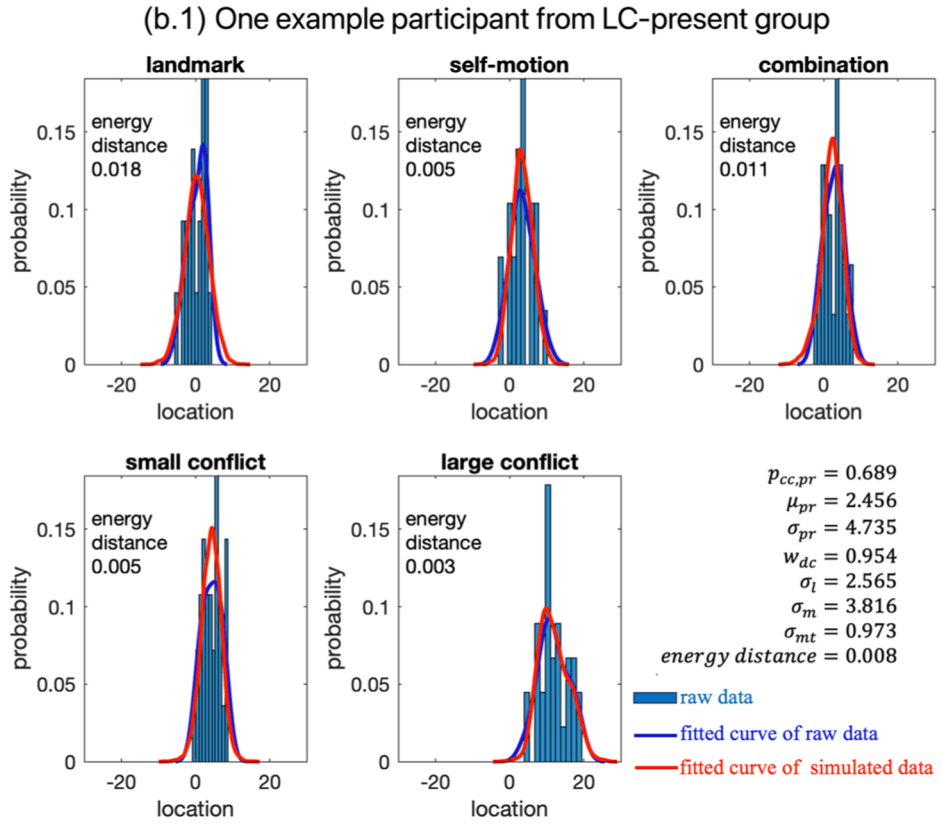



**Fig. 7. Illustrating goodness-of-fit of BCI model for example participants** For individual participants, the raw behavioral data are plotted as histograms, alongside the fitted curves to the raw data and the simulated data based on the BCI model. Simulated data were generated by sampling synthetic responses from the three decision strategies (model averaging, model selection, and probability matching), with proportions determined by their relative marginal likelihoods. The estimated values of the model parameters (Bayes average) are also displayed for each participant. **(a)** Two example participants from the LC-absent group. **(b)** Two example participants from the LC-present group. **(c)** Two example participants from the LC-informed group.



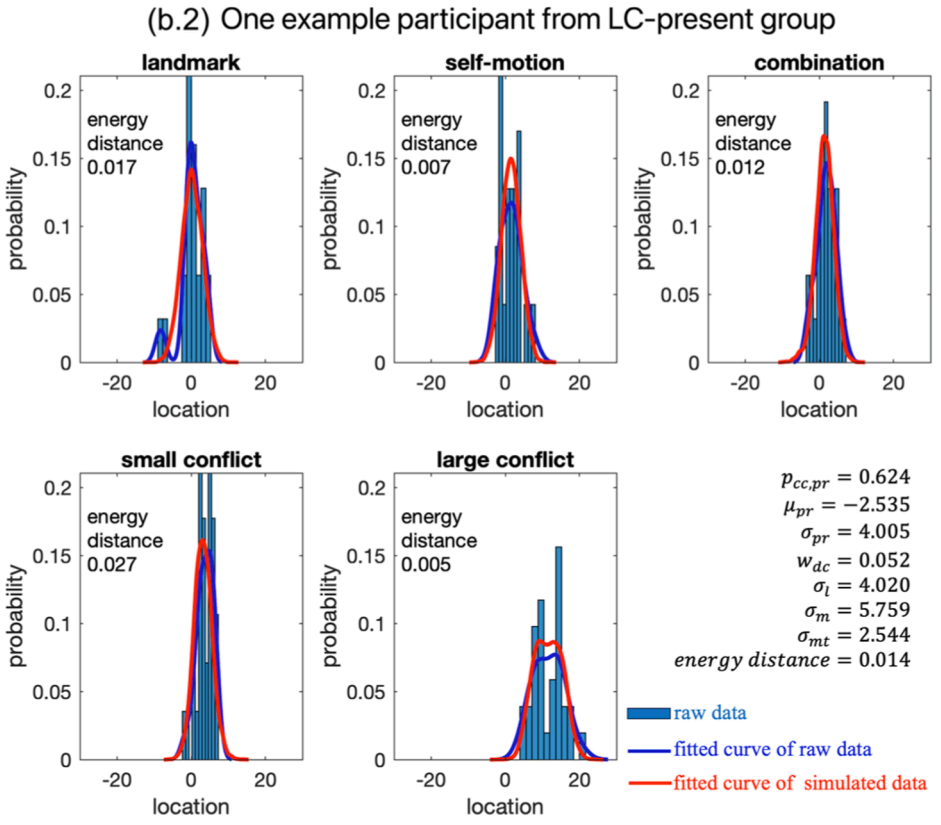
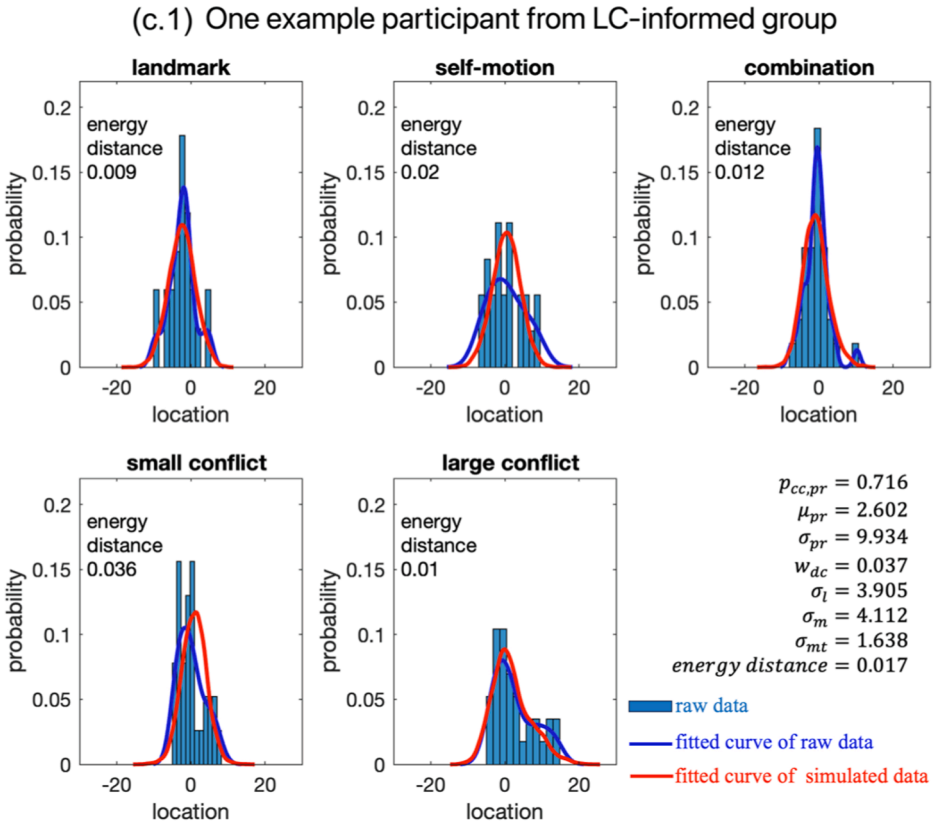




**Fig. 7.** (*continued*).



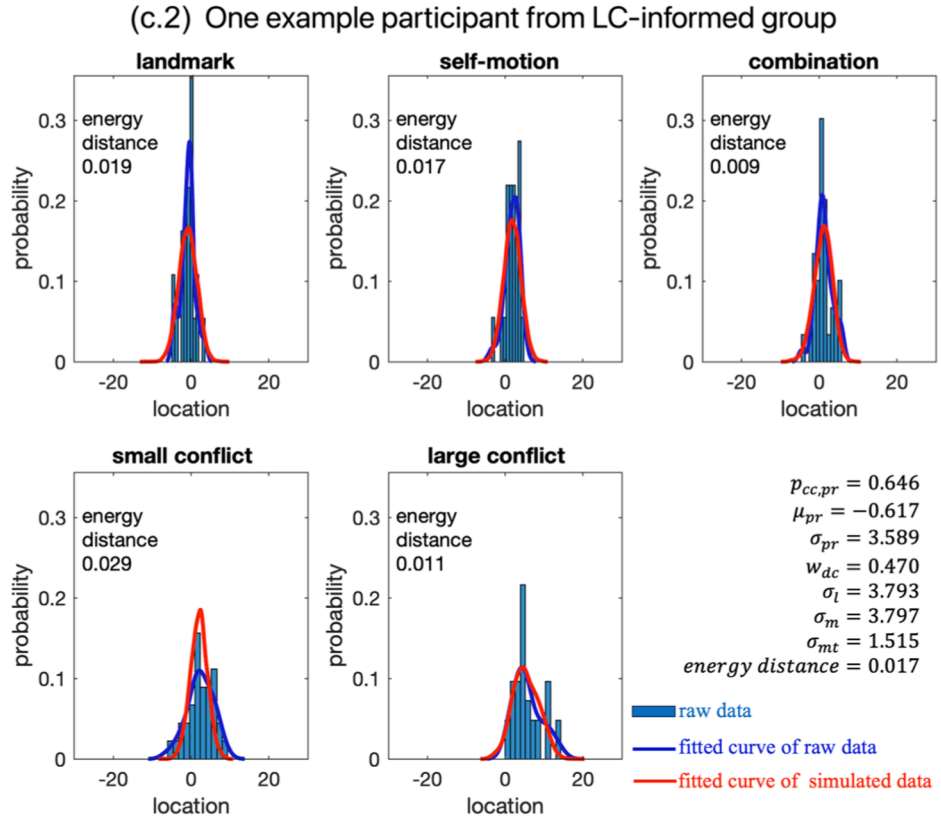




**Fig. 7.** (*continued*).

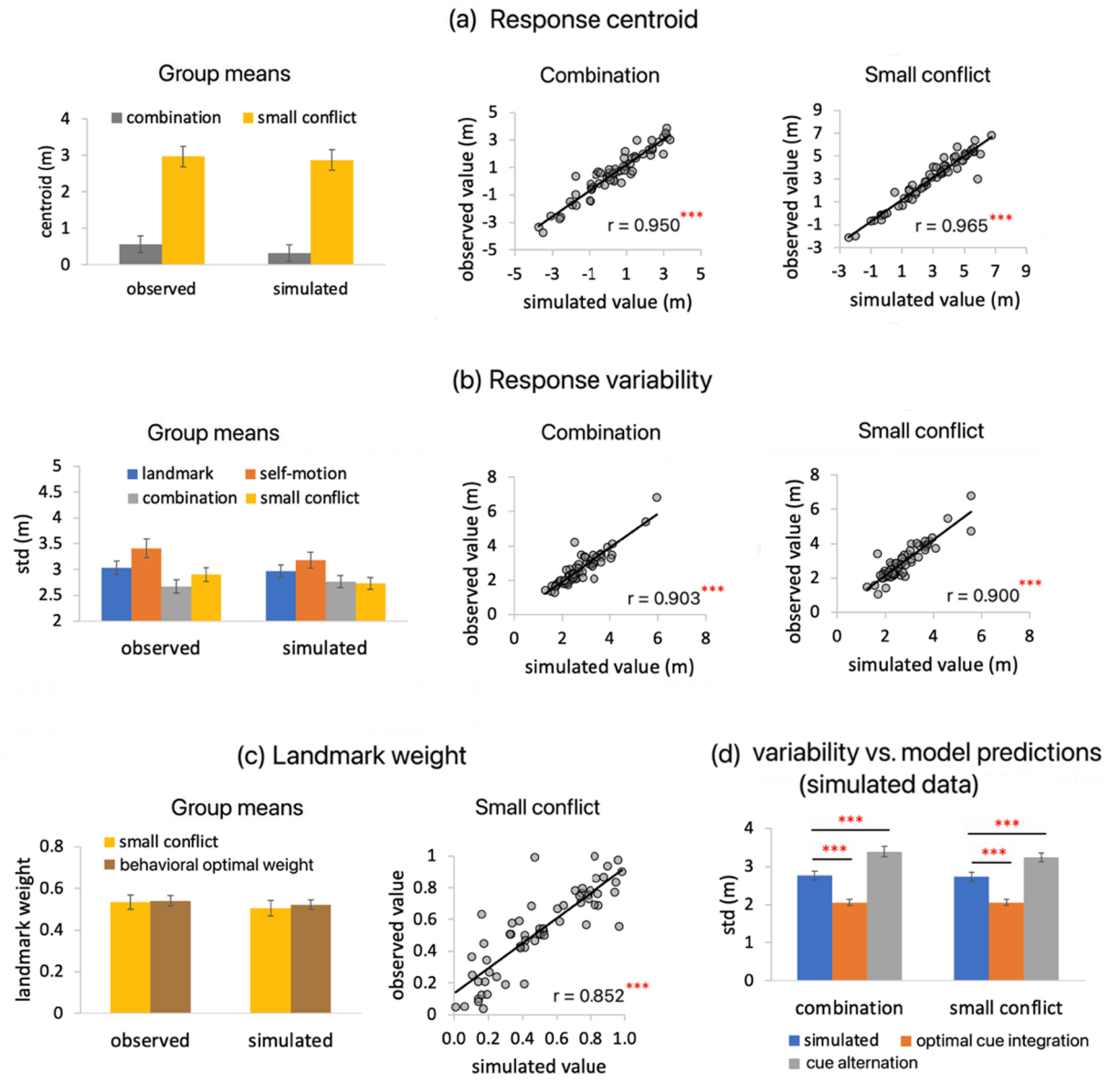



**Fig. 8. Goodness-of-fit of BCI model in LC-absent group** From **(a)** to **(c)**, actual data and model predictions are displayed for the following key variables: response centroids in the two double-cue conditions **(a)**; response variability in the two double-cue conditions (b); and the behavioral weight assigned to landmarks in the small conflict condition, compared with the behavioral optimal weight **(c)**. Model predictions were generated via simulations and represent the Bayes average across the three decision strategies. In each subplot, the left panel shows group means of the actual data and model predictions; the middle and right panels show across-participant correlations between actual data and model predictions. Pearson r correlation coefficients are shown along with the scatterplots. **(d)** shows simulated data comparing response variability in the double-cue conditions with predictions from the MLE model of cue integration (i.e., optimal cue integration) and the cue alternation model. All data shown in this subplot are simulated rather than observed. "***" indicates $p < 0.001$. Error bars represent standard error of the mean (S.E.).

panels).

Next, we found that the simulated data reproduced the key patterns of behavioral results obtained in the observed data in terms of cue weight and response variability in the double-cue conditions. First, the simulated landmark weight was not different from the simulated behavioral optimal weight in the LC-absent group (small conflict, $t(59) = -0.504$, $p = 0.616$, $BF_{10} = 0.160$) and the LC-present group (small conflict, $t(40) = 0.564$, $p = 0.576$, $BF_{10} = 0.196$; large conflict, $t(40) = 1.387$, $p = 0.173$, $BF_{10} = 0.409$), but was significantly smaller than the simulated behavioral optimal weight in the LC-informed group (small conflict, $t(39) = -3.136$, $p = 0.003$, $BF_{10} = 10.771$; large conflict, $t(39) = -11.453$, $p < 0.001$, $BF_{10} = 1.528*10^{11}$) (Figs. 8-10, c, left panel).

Second, in all double-cue conditions, the simulated response variability fell between the simulated prediction of the MLE cue integration model and the simulated prediction of the cue alternation model ($|ts| > 6$, $ps < 0.001$, $BFs_{10} > 8*10^{4}$; Figs. 8-10, panel d).

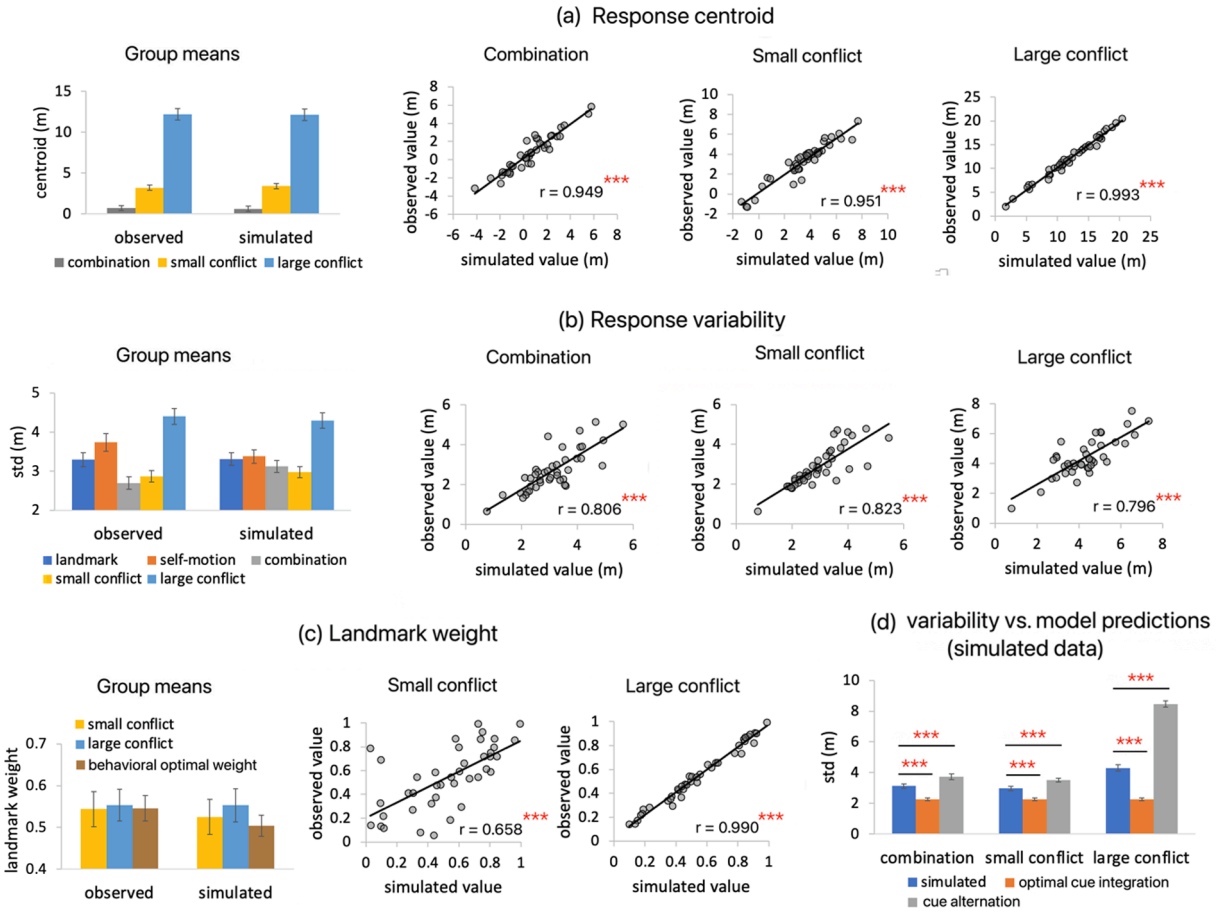



**Fig. 9. Goodness-of-fit of BCI model in LC-present group** From **(a)** to **(c)**, actual data and model predictions are displayed for the following key variables: response centroids in the three double-cue conditions **(a)**; response variability in the two double-cue conditions **(b)**; and the behavioral weight assigned to landmarks in the two conflict conditions, compared with the behavioral optimal weight **(c)**. Model predictions were generated via model simulations and represent Bayes average across the three decision strategies. In each subplot, the left panel shows group means of the actual data and model predictions; the middle and right panels show across-participant correlations between actual data and model predictions. Pearson r correlation coefficients are shown along with the scatterplots. **(d)** shows simulated data comparing response variability in the double-cue conditions with predictions from the MLE model of cue integration (i.e., optimal cue integration) and the cue alternation model. All data shown in this subplot are simulated rather than observed. "***" indicates $p < 0.001$. Error bars represent S.E. of the mean.

Even in the large conflict condition, the simulated response variability was numerically far lower than the simulated prediction of the cue alternation model (Figs. 9 and 10, panel d). This simulation replicates our behavioral findings (Fig. 3, panels c-e) and aligns with previous studies (Chen et al., 2017; Qi & Mou, 2024; Sjolund et al., 2018). These results suggest a mixture of cue integration and cue alternation strategies in double-cue conditions, a key prediction of the BCI model.

#### 4.2.4. *Influences of cue conflict on model parameters*

The preceding results have established that the BCI model provided a considerably good fit to the data. Here, we proceeded to assess how cue conflict affected the cognitive processes underlying spatial navigation, by comparing the three groups on the free parameters of the BCI model. To reiterate, The Bayesian model averaging approach was adopted for estimating the model parameters across the three decision strategies.

We adopted the ANOVA test, with Welch correction if the assumption of variance homogeneity was violated. Outliers were identified within each group for each variable, i.e., > 3rd quartile + 3*IQR or < 1st quartile – 3*IQR. Outliers were winsorized to the nearest inlier within each group (Reifman & Keyton, 2010).

As shown in Fig. 11, the three groups differed significantly on two model parameters: the weight assigned to landmarks in the different-cause judgment ($w_{dc}$, $F(2,138) = 21.816$, $p < 0.001$, $\eta_p^2 = 0.240$, $BF_{inclusion} = 1.79 * 10^6$; Fig. 11d), and the sensory noise level of landmarks ($\sigma_l$, $F(2,138) = 14.121$, $p < 0.001$, $\eta_p^2 = 0.170$, $BF_{inclusion} = 6493$; Fig. 11e). Regarding $w_{dc}$, post-hoc comparisons showed that the LC-informed group was lower than both the LC-absent group ($t(138) = 5.702$, $p < 0.001$) and the LC-present group ($t(138) = 5.992$, $p < 0.001$), whereas the LC-absent group and the LC-present group did not differ from each other ($t(138) = -0.754$, $p = 0.452$).

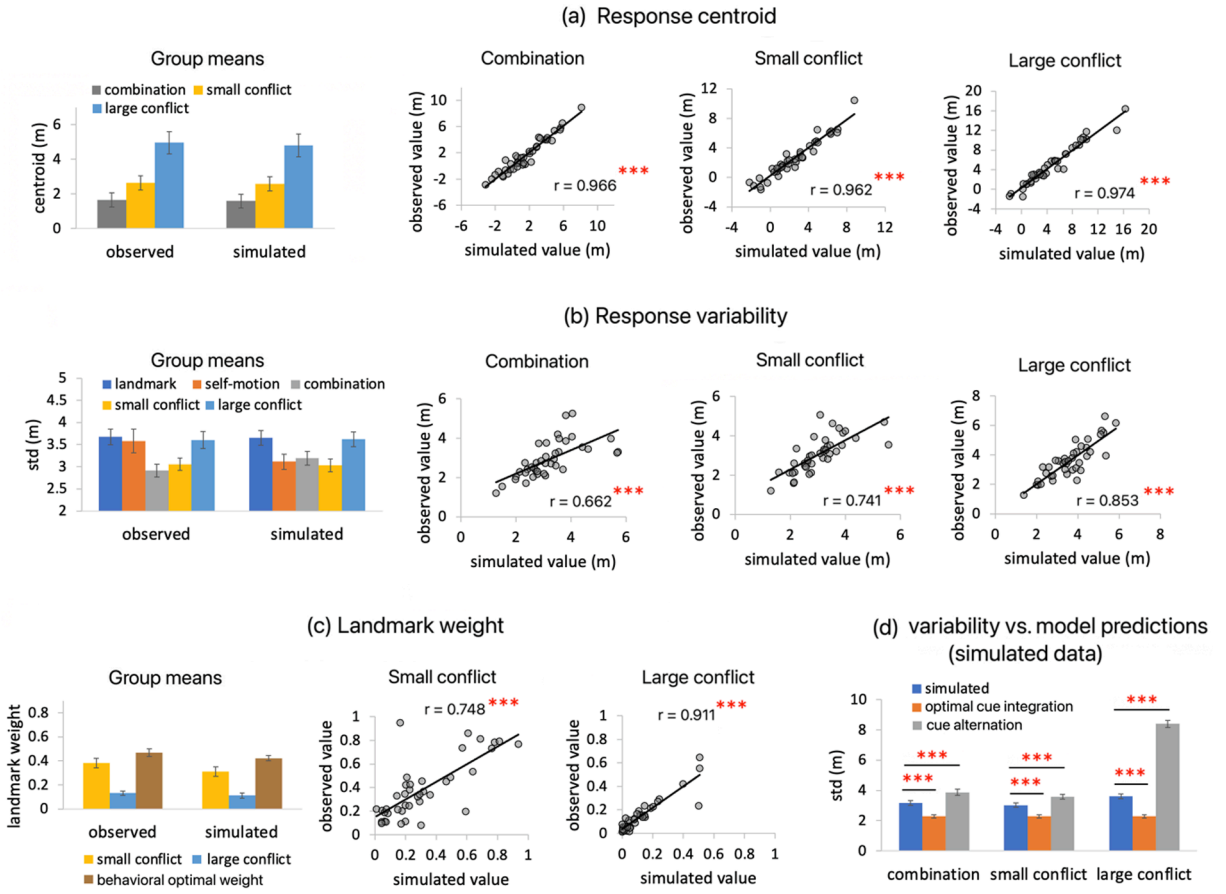



**Fig. 10. Goodness-of-fit of BCI model in LC-informed group** From **(a)** to **(c)**, actual data and model predictions are displayed for the following key variables: response centroids in the three double-cue conditions **(a)**; response variability in the two double-cue conditions **(b)**; and the behavioral weight assigned to landmarks in the conflict conditions, compared with the behavioral optimal weight **(c)**. Model predictions were generated via model simulations and represent Bayes average across the three decision strategies. In each subplot, the left panel shows group means of the actual data and model predictions; the middle and right panels show across-participant correlations between actual data and model predictions. Pearson r correlation coefficients are shown along with the scatterplots. **(d)** shows the simulated data comparing response variability in the double-cue conditions with predictions form the MLE model of cue integration (i.e., optimal cue integration) and the cue alternation model. All data shown in this subplot are simulated rather than observed. "***" indicates $p < 0.001$. Error bars represent S.E. of the mean.

Regarding $\sigma_l$, post-hoc comparisons showed that the LC-informed group was larger than both the LC-absent group ($t(138) = -5.234$, $p < 0.001$) and the LC-present group ($t(138) = -3.651$, $p < 0.001$), whereas the LC-absent group and the LC-present group did not differ from each other ($t(138) = -1.268$, $p = 0.207$). Additionally, the three groups did not differ significantly on the sensory noise level of self-motion cues ($\sigma_m$; Fig. 11f). Accordingly, the sensory optimal weight for landmarks, which is determined by the ratio of sensory noise levels of the cues ($w_{opt,sen} = \sigma_m^2/(\sigma_l^2 + \sigma_m^2)$), was significantly lower in the LC-informed group than the other two groups ($F(2,138) = 8.409$, $p < 0.001$, $\eta_p^2 = 0.109$, $BF_{inclusion} = 70.147$; LC-informed vs. LC-absent, $t(138) = 3.759$, $p < 0.001$; LC-informed vs. LC-present, $t(138) = 3.439$, $p < 0.001$; LC-absent vs. LC-present, $t(138) = 0.015$, $p = 0.988$; Fig. 11*h*). This result means that the LC-informed group assigned lower weight to landmarks in the same-cause judgment, compared to the other two groups.

The three groups did not differ in other model parameters ($Fs < 3.1$, $ps > 0.05$, $\eta s_p^2 < 0.05$, $BFs_{incl} < 0.9$). In particular, the three groups were very close in the prior belief that the cues originate from the same spatial source ($mean p_{cc,pr} = 0.595$ vs. $0.588$ vs. $0.598$), and the ANOVA test showed no significant differences between the groups ($F(2, 138) = 0.015$, $p = 0.985$, $\eta_p^2 = 2.159*10^{-4}$, $BF_{inclusion} = 0.072$; 11a).

The preceding results have shown that the LC-informed group had lower weight assigned to landmarks in the sensory cue alternation process in the different-cause judgment ($w_{dc}$) and higher sensory noise level of landmarks ($\sigma_l$) than the other two groups. Because increased sensory noise of landmarks $\sigma_l$ predicts lower weight assigned to landmarks for the same-cause judgment ($w_{opt,sen}$), we wondered whether the decreased $w_{dc}$ in the LC-informed group could be explained by the decreased $w_{opt,sen}$ in this group; that is, whether decreased sensory weight given to landmarks in the different-cause judgment was a consequence of decreased sensory weight given to landmarks in the same-cause judgment. In addition, it is of theoretical significance to understand whether distinct sensory-

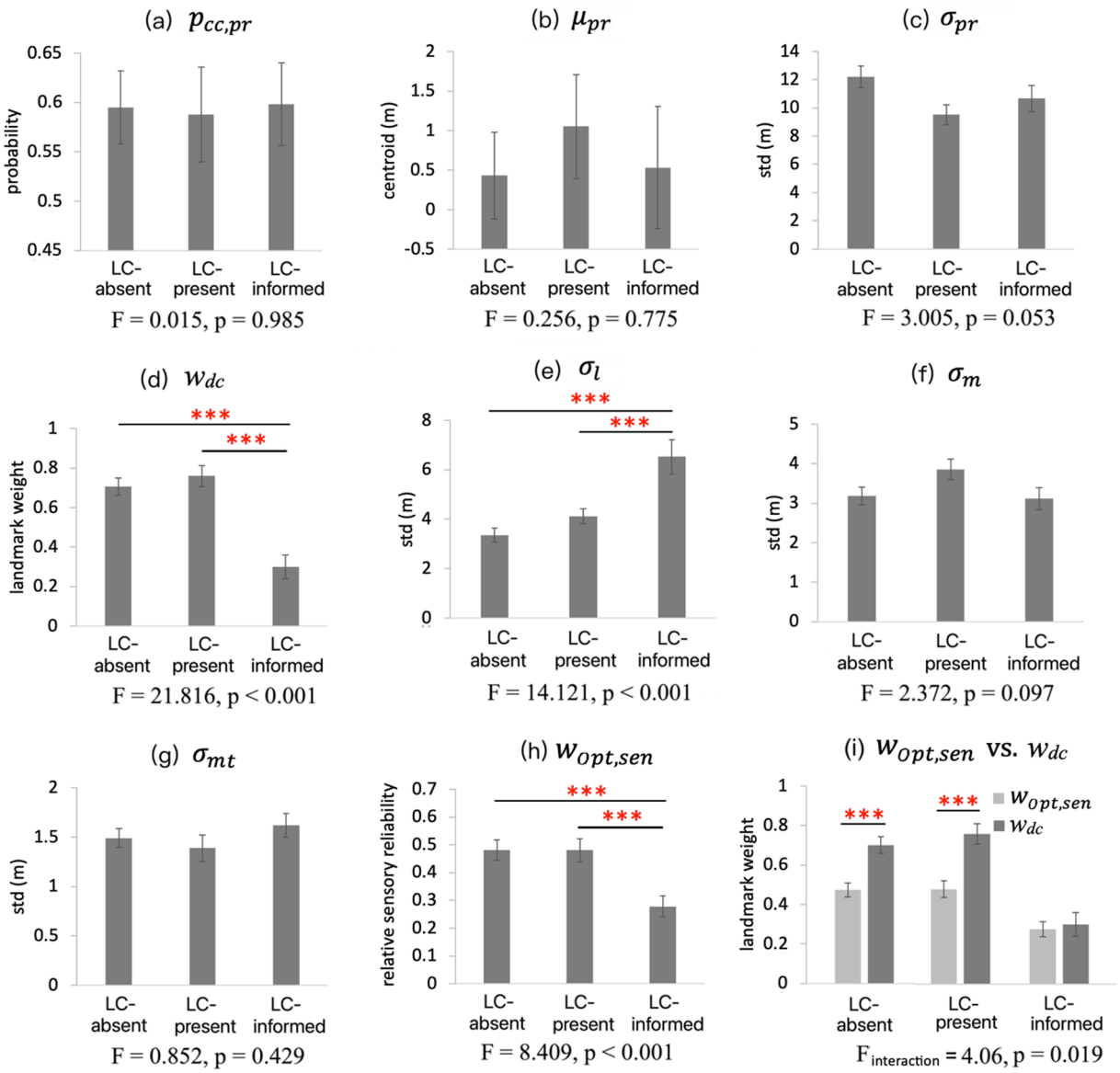



**Fig. 11. Comparisons of BCI model parameters among experimental groups** From **(a)** to **(g)**, the three groups were compared on each of the seven model parameters. **(h)** The sensory optimal weight ($w_{opt,sen}$) was compared among the three groups. **(i)** The two sensory cue weight parameters, $w_{Opt,sen}$ and $w_{dc}$, were compared among the three groups. These two parameters are conceptually parallel to each other, as they both represent the sensory weight assigned to landmarks in the sub-models for the same-cause judgment and the different-cause judgment respectively, before the prior distribution of target location is considered. $p_{cc,pr}$ – prior belief in a common sense; $\mu_{pr}$ – mean of the prior distribution of target location; $\sigma_{pr}$ – standard deviation of the prior distribution of target location; $w_{dc}$ – sensory weight assigned to landmarks in the different-cause judgment; $\sigma_l$ – sensory noise level of landmarks; $\sigma_m$ – sensory noise level of self-motion cues; $\sigma_{mt}$ – motor noise; $w_{opt,sen}$ – optimal sensory weight to landmarks in the same-cause judgment. "***" indicates $p < 0.001$. Error bars represent standard error of the mean (S.E.).

cue-weighting strategies are adopted by different cause judgments, by comparing $w_{dc}$ and $w_{opt,sen}$.

To address this question, we conducted a mixed ANOVA test, with weight type ($w_{dc}$ vs. $w_{opt,sen}$) and group (the LC-absent vs. the LC-present vs. the LC-informed) as independent variables (Fig. 11i). This analysis revealed a significant interaction between weight type and group ($F(2,138) = 4.060$, $p = 0.019$, $\eta_p^2 = 0.056$, $BF_{inclusion} = 13.207$). Post-hoc comparisons showed that: $w_{dc}$ was larger than $w_{opt,sen}$ in both the LC-absent group ($t(138) = 4.050$, $p < 0.001$) and the LC-present group ($t(138) = 4.151$, $p < 0.001$), whereas in the LC-informed group, $w_{dc}$ did not differ from $w_{opt,sen}$ ($t(138) = 0.342$, $p = 0.734$). These results indicate that for different-cause judgments, participants in the LC-absent and LC-present groups assigned to landmarks a sensory weight greater than the sensory optimal weight, whereas participants in the LC-informed group assigned to landmarks a sensory weight equal to the sensory optimal weight. In other words, at the group-level, in the LC-informed group, $w_{dc}$ decreased beyond what was predicted by the increased $\sigma_l$ and decreased

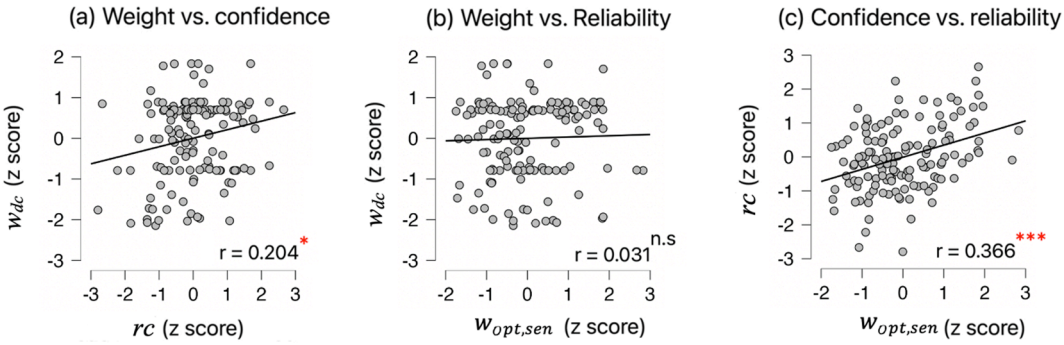



**Fig. 12.** Simple bivariate correlations among model parameters of cue weights and self-rated confidence Scatterplots depicts relationships among *rc* (cue relative confidence), $w_{opt,sen}$ (sensory weight assigned to landmarks in the same-cause judgment), and $w_{dc}$ (sensory weight assigned to landmarks in the different-cause judgment). Z-scores were calculated within each group. Outliers were identified within each group, i.e., > 3rd quartile + 3*IQR or < 1st quartile – 3*IQR. Outliers were winsorized to the nearest inlier within each group (Reifman & Keyton, 2010). "***" indicates $p < 0.001$; "n.s." indicates $p > 0.05$.

**Table 2**

**Predicting sensory cue weight in the different-cause judgment** A multiple linear regression was conducted to predict the sensory weight assigned to landmarks in the cue segregation sub-model for the different-cause judgment ($w_{dc}$). Results for the significant predictor *rc* are highlighted in bold. *rc* represents the relative confidence participants placed in landmarks compared to self-motion cues, reflecting their subjective evaluation of cue quality. $w_{Opt,sen}$ is the inverse ratio of sensory noise levels between landmarks and self-motion cues, reflecting the relative objective cue quality of the cues. $w_{Opt,sen}$ also corresponds to the sensory weight assigned to landmarks in the same-cause judgment ($w_{cc}$). Note that we did not include any group dummy variables (*isLC-absent*, *isLC-present*, and *isLC-informed*) in the multiple linear regression model, because prior to this analysis, the dependent variable $w_{dc}$ had already been standardized within each group and therefore had the same mean (= 0) in all the groups. Similarly, all independent variables were standardized within each group prior to the analysis.

| Predictor | ß | t | p | $BF_{inclusion}$ |
|---|---|---|---|---|
| ***rc*** | **0.210** | **2.356** | **0.020** | **1.400** |
| $\boldsymbol{w_{opt,sen}}$ | −0.047 | −0.526 | 0.600 | 0.279 |
| *isLC-informed*rc* | −0.007 | −0.082 | 0.934 | 0.260 |
| *isLC-informed**$\boldsymbol{w_{opt,s}}$ | **−0.221** | **−2.446** | **0.016** | **0.911** |
| *isLC-present*rc* | 0.007 | 0.082 | 0.935 | 0.259 |
| *isLC-present**$\boldsymbol{w_{opt,s}}$ | −0.142 | −1.565 | 0.120 | 0.457 |

$w_{opt,sen}$.

The parameter recovery analysis confirmed the validity of our estimation of the model parameters (see details in Supplemental information, Section B), as we found moderate or strong correlations across participants between the recovered values and the best-fitting values for the three parameters of interest ($p_{cc,pr}$, $w_{dc}$, and $\sigma_l$, rs > 0.44, ps < 0.001).

### *4.2.5. Self-rated confidence predicts cue weight in different-cause judgment*

The BCI model incorporates two cue-weighting schemes: in the cue integration sub-model for common-cause judgments, the sensory weights assigned to the two cue types are determined by the sensory optimal weight ($w_{opt,sen}$); in the cue alternation sub-model for different-cause judgments, the parameter $w_{dc}$ represents the sensory weight assigned to landmarks in the cue alternation process. While $w_{opt,sen}$ is determined by other sensory noise parameters ($\sigma_l^2$ and $\sigma_{mt}^2$), $w_{dc}$ is free to vary. A critical question is what influences $w_{dc}$?

Our previous study identified cue relative confidence (*rc*) as a key factor in cue weighting (Chen et al., 2017). Cue relative confidence (*rc*), was calculated as the relative mean confidence level between the landmark condition and the self-motion condition:

$$rc = \frac{\overline{cf}_l}{\overline{cf}_l + \overline{cf}_m}$$

in which $\overline{cf}_l$ and $\overline{cf}_m$ represent the mean confidence score averaged across all trials in the landmark condition and the self-motion condition, respectively. We found that cue relative confidence positively predicted the weight assigned to landmarks, even after controlling for relative objective performance.

Here, we investigated whether *rc* predicted the weight assigned to landmarks for the different-cause judgment ($w_{dc}$). As shown in Fig. 12, across the three groups, there existed a significant positive correlation between $w_{dc}$ and *rc* ($r = 0.204$, $p = 0.015$, $BF_{10} = 1.950$), meaning that in the different-cause judgment, participants assigned greater weight to the spatial cue with which they felt more confident. There was no significant correlation between $w_{dc}$ and $w_{opt,sen}$ ($r = 0.031$, $p = 0.714$, $BF_{10} = 0.113$). There was significant

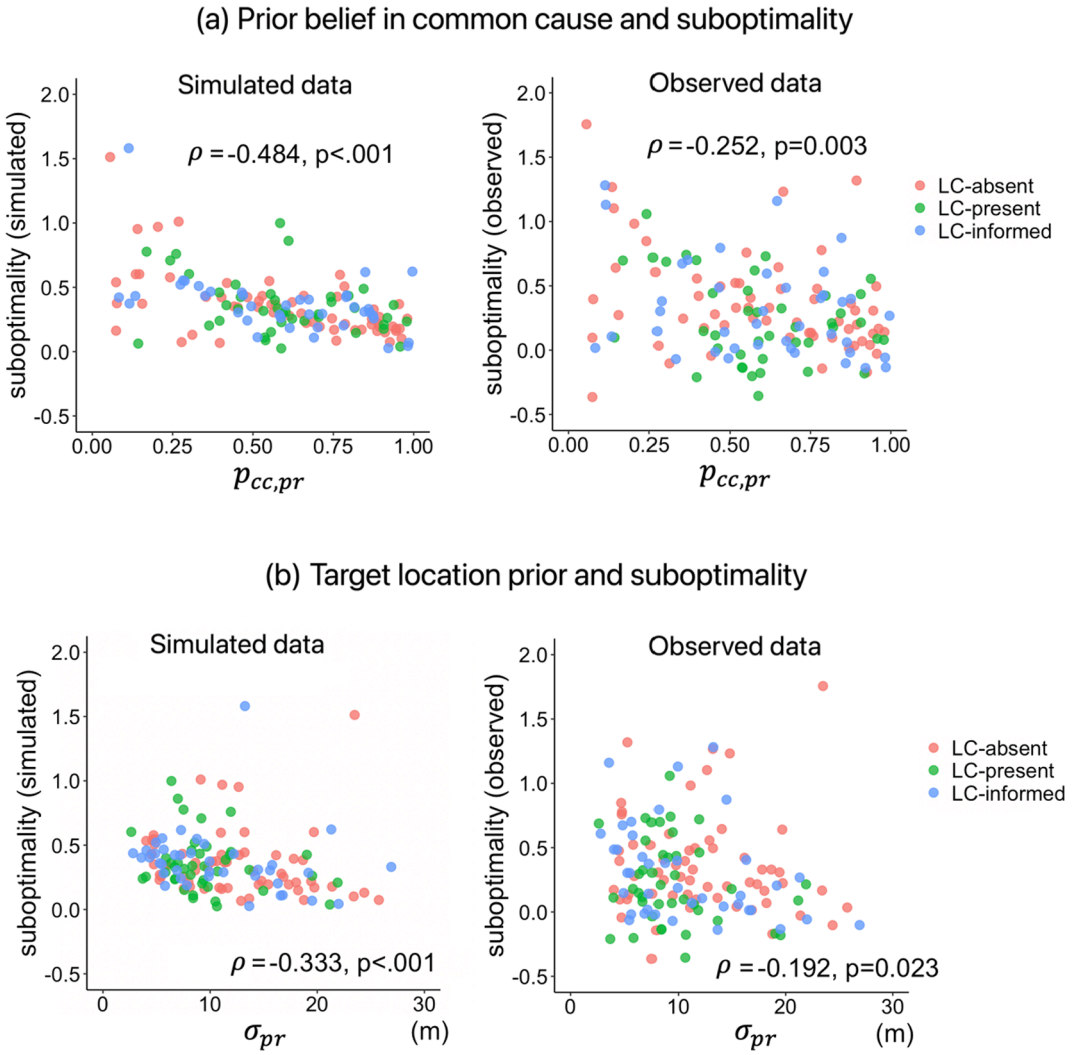



**Fig. 13. Relationships between prior information and cue combination suboptimality (a)** Correlation between prior belief in a common cause ($p_{cc,pr}$) and cue combination suboptimality in the simulated (left panel) and observed data (right panel). **(b)** Correlation between the standard deviation of the prior distribution of target location ($\sigma_{pr}$) and cue combination suboptimality in the simulated (left panel) and observed data (right panel). Dots of different colors represent individual participants from different groups in the main experiment.

correlation between *rc* and $w_{opt,sen}$ ($r = 0.366$, $p < 0.001$, $BF_{10} = 2030$), meaning that participants felt more confident using the spatial cue with lower sensory noise level.

To control for potential influences of $w_{opt,sen}$, we conducted a multiple linear regression analysis with $w_{dc}$ as the dependent variable, and *rc* and $w_{opt,sen}$ as the independent variables, along with terms involving the group dummy variables (*isLC-present* and *isLC-informed)*.

As shown in Table 2, *rc* was a significant predictor *(*standardized ß = 0.197, t = 2.209, p = 0.029). The Bayesian analysis of posterior summaries of coefficients showed that $BF_{inclusion}$ for *rc* was 1.400, meaning that the data have increased our prior odds for including *rc* as a predictor by a factor of 1.400 – positive, albeit small, evidence for including *rc* in the model.

Note that the interaction term "*isLC-informed* $w_{opt,sen}$*" was significant (standardized ß = −0.221, t = -2.446, p = 0.016, $BF_{inclusion}$ = 0.911). To interpret this effect, in the LC-absent group, the correlation between $w_{opt,sen}$ and $w_{dc}$ was positive and significant (r = 0.279, p = 0.031); in the LC-present group, this correlation was negative numerically (r = -0.087, p = 0.588); in the LC-informed group, this correlation was negative numerically (r = -0.223, p = 0.167). This pattern of results suggests that cue integration and alternation strategies are harder to distinguish in scenarios with minimal cue conflicts.

In summary, we found that *rc* uniquely predicted $w_{dc}$, whereas the role of $w_{opt,sen}$ depends on cue conflict contexts. These findings suggest that sensory cue weighting operates differently for common-cause vs. different-cause judgments.

## 5. Additional analyses and experiments

### 5.1. Modeling cue congruency judgments

Among previous studies applying the BCI model to multi-sensory perceptual tasks involving spatial localization of audio-visual stimuli, some fit the model only to localization responses (Körding et al., 2007; Odegaard et al., 2017; Wozny et al., 2010), whereas others fit the model to both localization responses and cue congruency judgments (Badde et al., 2020; Hong et al., 2022). Recall that participants in the LC-informed group also judged cue congruency by explicitly determining whether the landmark had been relocated (Fig. 2b). To our knowledge, no studies have implemented both approaches on the same dataset for comparison. Therefore, to examine the robustness of our modeling findings, we modeled both cue congruency judgments and spatial localization responses.

This subsidiary analysis not only ensures the robustness of our primary findings but also extends the scope of our investigation. In particular, this approach offers a more complete picture of how individuals resolve spatial uncertainty during navigation, particularly when cues conflict.

Due to a technical oversight, cue congruency judgments were not recorded for the first 10 participants in the LC-informed group, potentially resulting in insufficient statistical power. To address this, we included a separate group of 40 participants (LC-informed-additional group), whose experimental procedure closely matched that of the LC-informed group (see Supplemental information, Section C). Given that this additional group exhibited similar behavioral results, we conducted a pooled analysis of the two groups.

As detailed in Supplemental information (Section D), key modeling results remained consistent with those obtained from modeling localization responses alone: compared to the LC-absent group, the LC-informed and LC-informed additional groups assigned lower weight to landmarks in different-cause judgments ($w_{dc}$), exhibited higher sensory noise associated with landmarks ($\sigma_l$), and had lower optimal sensory weight for landmarks in same-cause judgments ($w_{opt,sen}$). The consistency underscores the robustness of our findings across different modeling methodologies. However, unlike when modeling localization responses alone, the LC-informed and LC-informed-additional groups showed greater motor noise ($\sigma_{mt}$) than the LC-absent group. Although the precise explanation for this result remains unclear to us, it does not affect our interpretation of the primary findings.

### 5.2. Conceptual Replication of Roy et al. (2023)

Our main experiment revealed that increasing cue conflict alone was not sufficient to change participants' reliance on different cues. This result appears to contradict Roy et al (2023), who found that increasing a landmark's instability decreased the weight assigned to it. We hypothesized that this discrepancy stemmed from the feedback participants received in this study, which was defined by other stable landmarks and therefore devalued the unstable landmark. This feedback mechanism resembles the instructions given to the LC-informed and LC-informed-additional groups in our study, where participants were explicitly told that the landmark could be relocated and would no longer provide valuable information about the target location.[2]

To test this hypothesis, we examined another group of 39 participants (LC-feedback group), using the same experimental setup as in the LC-present group but with a key difference: in double-cue trials, after participants made the localization response and rated confidence, the target appeared at a position defined by the self-motion cues. Participants were told that this feedback represented the correct target location. They needed to travel to this feedback position to complete the trial.

As detailed in Supplemental information (Section E), the LC-feedback group exhibited the same pattern of behavioral results as the LC-informed group. This group assigned overall lower weight to landmarks and showed larger response variability in the landmark condition (i.e., poorer performance), compared to the LC-absent group.

Modeling results also replicated key findings from the LC-informed group: the LC-feedback group showed lower weight assigned to landmarks in different-cause judgments ($w_{dc}$), higher sensory noise associated with landmarks ($\sigma_l$), and lower weight assigned to landmarks in same-cause judgments ($w_{opt,sen}$), compared to the LC-absent group. Surprisingly, however, unlike the LC-informed group, the LC-feedback group exhibited a stronger prior belief in common cause ($p_{cc,pr}$) than the LC-absent group, meaning that experiencing large cue conflict trials paradoxically reinforced a belief in a unified causal structure.

In summary, these findings support our hypothesis that the finding of reduced reliance on unstable landmarks in Roy et al.'s study was not solely driven by cue conflict but rather reflects the combined effect of cue conflict and feedback that deliberately devalued the landmark.

## 6. Explaining cue combination suboptimality within BCI model

Besides explaining navigation behavior in situations with substantial cue conflicts, the BCI model can also be leveraged to uncover underlying factors contributing to suboptimal cue combination behavior in situations with small or no cue conflicts. To this end, we examined which BCI model parameters correlated with cue combination suboptimality in the combination and small conflict conditions. We calculated a suboptimality index, which measures deviation of response variability in these two conditions from the

[2] Although in Roy et al. (2023) the feedback was only provided during training trials but not during test trials, the navigation strategy developed during training trials likely persisted into the test trials because these two types of trials were intermixed throughout the experiment.

prediction of the MLE model.

suboptimality $= (\sigma_{obs} - \sigma_{opt})/\sigma_{opt}$ where $\sigma_{obs}$ represents observed response variability and $\sigma_{opt}$ represents the optimal response variability. $\sigma_{opt}$ was calculated from responses in the single-cue conditions, following the MLE rules of cue integration (see Section 3.1 "Behavioral Analysis").

We analyzed all participants from the main experiment. Due to the high correlation between response variability in the combination and small conflict conditions, we averaged across these conditions. Both observed and simulated data were examined. Simulated data were obtained using the Bayesian averaging approach. Simulated data reflect the inner mechanisms of the BCI model and serve to assess its ability to capture cue combination suboptimality. Consistency between the observed and simulated data would indicate that the BCI model fit the data sufficiently well to capture nuanced aspects of participants' cue combination behavior.

We focused on two parameters that encapsulate the essence of the BCI model: prior belief in common cause ($p_{cc,pr}$) and uncertainty in the prior knowledge about target location ($\sigma_{pr}$). As shown in Fig. 13, both parameters were negatively correlated with cue combination suboptimality, in both the simulated and observed data. A stronger prior belief in common cause and lower uncertainty in the prior knowledge about target location corresponded to less severe cue combination suboptimality.

We investigated other model parameters for completeness. Motor noise ($\sigma_{mt}$) was positively correlated with cue combination suboptimality in the simulated data ($\rho = 0.204$, $p = 0.016$), and sensory noise level associated with self-motion cues ($\sigma_m$) was negatively correlated with cue combination suboptimality in both the simulated ($\rho = -0.180$, $p = 0.033$) and observed data ($\rho = -0.348$, $p < 0.001$).

In summary, prior belief in common cause and uncertainty in prior knowledge about target location contributed to cue combination suboptimality. The application of the BCI model offers insights into the underlying mechanisms of suboptimal cue integration in scenarios with minimal cue conflicts.

## 7. Discussion

The current study aimed to investigate cognitive processing of spatial cue conflict in spatial distance estimation during navigation. To accomplish this aim, we developed a novel cue combination paradigm conducted on a linear space and tested two cognitive models of the task: the Bayesian causal inference model (BCI) and a non-Bayesian sensory disparity model. Our contribution is multifaceted. First, the behavioral findings illustrate the crucial, yet nuanced, influences of spatial cue conflict on cue-weighting behavior. Second, the BCI model outperformed the non-Bayesian alternative model and effectively accounted for participants' navigation behaviors, demonstrating its applicability in the field of spatial navigation. Third, the model parameter analysis provided valuable insights into how cue conflict influences navigation, pinning down the cognitive processes specifically affected by cue conflict and elucidating the intricate interplay among these processes. Finally, our modeling results shed light on the root causes of suboptimal cue combination in situations with minimal cue conflicts.

### *7.1. Cognitive processes affected by cue conflict during navigation*

In behavior, the LC-absent group and the LC-present group showed no differences, indicating that increasing cue conflict alone exerted no discernable influences on navigation behavior. However, compared to the LC-absent group, the LC-informed group showed increased response variability in the landmark condition and decreased weight assigned to landmarks in the double-cue conditions, indicating that explicit awareness of cue conflict is necessary to behavioral changes. The LC- feedback group, which received self-motion-defined feedback to reinforce awareness of the landmark's instability and invalidity, showed results consistent with the LC-informed group. This consistency suggests that a cue must be explicitly devalued to reduce reliance on it.

The modeling work offered deeper insights into the cognitive processes at play behind these behavioral findings. Cue conflict, when coupled with explicit awareness of landmark instability, elevated sensory noise level of landmarks and reduced the weight assigned to landmarks when conflict was perceived. Surprisingly, navigators' prior belief about causal structure did not get updated in line with the cue conflict statistics: in most cases, the prior belief remained unchanged, even when participants were made explicitly aware of the cue conflict (i.e., LC-present, LC-informed, and LC-informed-additional groups); in one case (LC-feedback group), the prior belief in common cause even increased in the presence of greater cue conflict, contrary to the cue conflict statistics.

In the following sections, we will delve into detailed discussions on these three model parameters: sensory noise level of landmarks, cue weighting, and prior belief about causal structure.

#### *7.1.1. Sensory noise level of landmarks*

We observed increased response variability associated with landmarks when participants were explicitly aware of landmark instability. This finding is consistent with our earlier finding using the cue combination task in a 2-dimensional virtual reality space, which also showed increased response variability due to landmark instability (Chen et al., 2017, Experiment 2). Cognitive modeling further revealed that the increased response variability associated with landmarks was driven by elevated sensory noise in landmarks ($\sigma_l$).

Why would explicit awareness of landmark instability elevate the sensory noise of the landmark cue? We speculate that once participants recognized that landmarks could be invalid in defining the target location, they allocated less attention to them, leading to reduced cue efficiency and impaired behavioral performance. Intriguingly, previous studies on audio-visual spatial localization have documented the multisensory enhancement phenomenon, in which exposure to congruent audiovisual stimulus pairs improves

subsequent unisensory localization performance with auditory stimuli (Bruns et al., 2020). This phenomenon parallels our finding that unstable landmarks, when explicitly recognized as invalid, led to diminished behavioral performance. However, whether both effects stem from similar underlying mechanisms, such as altered attentional levels, remains to be investigated.

Importantly, our findings suggest that sensory noise level was influenced not only by a cue's physical properties but is also by cognitive factors. Indeed, prior research has shown that the sensory noise level of a cue can be modulated by attentional state (Badde et al., 2020), exposure to congruent/incongruent stimulus pairs (Hong et al., 2022), and perceptual training (Rohe & Noppeney, 2015). Therefore, it is essential to consider variations in sensory noise levels before jumping to conclusions about cue-weighting or prior beliefs.

Notably, the form of landmark instability differed between the current study and our previous study (Chen et al., 2017). In our previous study, a three-landmark layout constituted the landmark cue. Landmark instability was induced by changing the landmark configuration from trial to trial, while the landmark configuration remained unchanged between the encoding and retrieval stages within each trial. Thus, despite its instability, the landmark cue, still provided valid information for localizing the target. In contrast, in the current study, landmark instability was induced by repositioning the landmark during the retrieval stage within a trial, rendering it unreliable for defining the target location. While both forms of landmark instability increased sensory noise level, they had distinct effects on cue weighting. In our previous study, the observed weight assigned to unstable landmarks remained consistent with the behavioral optimal weight calculated from the MLE rules, whereas in the current study it was lower than the behavioral optimal weight. This difference suggests that the weight assigned to the landmark cue is determined by its perceived value rather than its instability. In the next section, we discuss how cue conflicts affect cue weighting.

#### 7.1.2. ***Weight*** *assigned to landmarks*

Our behavioral results showed that landmark instability, when made explicit, reduced participants' reliance on landmarks. Response variability in the landmark condition also increased, which theoretically resulted in a lower behavioral optimal weight for landmarks according to the MLE principles (i.e., relative response variabilities between the cues). Nevertheless, the observed behavioral weight assigned to landmarks was significantly lower than the behavioral optimal weight, suggesting additional influencing factors beyond reduced response precision.

Consistently, modeling results revealed two mechanisms behind this reduced reliance on unstable landmarks. First, explicit awareness of landmark instability increased sensory noise in landmarks ($\sigma_l$), inherently lowering their sensory weight in the same-cause judgment (i.e., $w_{opt,sen}$). Second, this awareness also reduced the sensory weight assigned to landmarks in the different-cause judgment ($w_{dc}$) when cue conflicts were perceived.

These two mechanisms operate independently, as indicated by two findings. First, in the LC-informed group, the sensory weight assigned to landmarks in the different cause-judgment ($w_{dc}$) was reduced beyond what is predicted by the increased sensory noise in landmarks ($w_{opt,sen}$) (Fig. 12i). This pattern of results was replicated in the LC-informed-additional and LC-feedback groups. Second, whereas cue-weighting in the same-cause judgment ($w_{opt,sen}$) was determined by relative objective cue quality (i.e., relative sensory noise levels of cues), cue-weighting in the different-cause judgment ($w_{dc}$) was positively correlated with subjective relative cue quality (i.e., relative cue confidence). Collectively, these findings indicate distinct cue-weighting processes between common-cause and different-cause judgments.

Additionally, the modeling work revealed a dissociation between behavioral and sensory cue weights (Aston et al., 2022). Behavioral cue weights (i.e., $w_{obs}$,$w_{opt,beh}$) are calculated from participants' behavioral responses in the behavioral analysis. Behavioral responses are jointly influenced by a multitude of factors, some of which are shared between different cue conditions, such as prior distribution of target location and motor noise. In contrast, sensory cue weights ($w_{opt,sen}$, $w_{dc}$) represent intrinsic weights, with components shared across cue conditions removed.

This dissociation was evident in our study. In the LC-absent and LC-present groups, when perceiving the landmark to be in conflict with optic flow, participants weighted landmarks heavily, with the sensory weight assigned to landmarks ($w_{dc}$) exceeding what is dictated by the relative sensory noise levels of cues – the sensory optimal weight ($w_{opt,sen}$). Yet, the observed behavioral weight assigned to landmarks ($w_{obs}$) was not different from the behavioral optimal weight ($w_{opt,beh}$). In the LC-informed group, when perceiving the landmark to be in conflict with path integration, participants weighted the landmark not differently from what is dictated by the relative sensory noise levels of cues (i.e., $w_{dc} = w_{opt,sen}$). However, the observed behavioral weight assigned to landmarks was lower than the behavioral optimal weight (i.e., $w_{obs} < w_{opt,beh}$). Consistent results were obtained in the LC-informed-additional and LC-feedback groups.

Given that sensory weights directly reflect participants' intrinsic reliance on cues, these findings suggest that when conflict between visual path integration and landmark-based navigation was not made explicit to participants, participants relied heavily on the landmarks in the different-cause judgment when they sensed conflicts between the cues. When participants were made explicitly aware of the landmark instability, their reliance on landmarks diminished yet remained consistent with the MLE rule of optimal cue weighting.

These findings reinforce the privileged status landmark-based navigation holds in spatial navigation, compared to path integration at least when it is based on visual optic flow. Despite appearing counterintuitive, this strong reliance on landmarks aligns with previous studies. For example, Zhao & Warren found that a majority of participants used landmarks exclusively to determine the walking direction back to the target location when the landmarks had been rotated by up to 90° (Zhao & Warren, 2015b). Similarly, in another study, when a landmark was unexpectedly shifted by as much as 115°, participants continued to rely on the landmark to localize the target (Zhao & Warren, 2015a).

This predominant reliance on landmarks likely stems from humans' habitual use of landmark-based navigation in daily life, because humans are endowed with high visual acuity to discern fine-grained spatial differences in landmarks (e.g., distances and angles to landmarks; Caves et al., 2018). In contrast, humans are generally poor at path integration (Loomis et al., 1993). Furthermore, path integration is error-prone and only works well in small-scale environments (Anastasiou et al., 2023), but human daily navigation primarily occurs in large-scale environments. This feature of human navigation system is in stark contrast to rodents, who are better at path integration than landmark-based navigation due to their very poor visual acuity – human eyesight is estimated to be 40 to 60 times sharper than that of a rat (Caves et al., 2018).

#### *7.1.3. Prior belief about causal structure*

Unlike sensory noise and cue weighting, we cannot infer anything about the prior belief about causal structure from behavioral results, as it is a cognitive construct. Our modeling results indicate that, in most cases, the prior belief was not affected by cue conflict, even when participants were explicitly aware of landmark instability. Interestingly, in one case when self-motion-defined feedback was provided in double-cue conditions, prior belief in a common cause even increased with heightened cue conflict. These findings challenge the common intuition that prior belief should update in alignment with new stimulus statistics (Roy et al., 2023).

Previous research has examined perceptual interaction between visual and auditory stimuli in the presence of cue conflict. Two classic phenomena have been investigated: the ventriloquism effect, wherein the observer's localization of an auditory stimulus is biased towards the location of a concurrent visual stimulus (Bertelson & Radeau, 1981); and the McGurk illusion, which occurs when auditory information clashes with visual information, leading to a fusion of the two senses and altering our perception of speech sounds (McGurk & MacDonald, 1976).

Prior behavioral studies have yielded consistent results: compared to incongruent pairs of auditory-visual stimuli, exposure to congruent pairs of auditory and visual stimuli increased the ventriloquism effect (Tong et al., 2020; Van Wanrooij et al., 2010) and the McGurk effect (Gau & Noppeney, 2016; Nahorna et al., 2012, 2015), signaling stronger tendency to bind the auditory stimulus and the visual stimulus (i.e., prior belief in a common cause). However, behavioral studies infer, rather than directly estimate, changes in prior belief, leaving room for alternative explanations, such as the prior knowledge on stimulus distribution and sensory noise levels of individual cues. Note that Tong et al. (2020) actually considered sensory noise levels of individual cues, as they compared behavioral performance with individual cues before and after cue conflict exposure. However, they did not consider the prior knowledge about stimulus distribution, which also influences causal structure judgments.

Modeling studies have yielded mixed results. Odegaard and colleagues found that exposure to spatiotemporally congruent audiovisual stimulus pairs did not alter prior belief in a common cause (Odegaard et al., 2017), whereas exposure to spatially incongruent but temporally congruent stimulus pairs unexpectedly increased it. Using a similar but improved paradigm, Hong and colleagues found that although prior belief in a common cause did not change at the group-level, individual differences existed: prior belief got updated in the same direction of the stimulus statistics for some participants, but in the opposite direction for others (Hong et al., 2022). Their simulations suggest that sensory noise levels of individual cues may change following bimodal stimulus exposure, influencing the likelihood of common-cause judgment. However, as they did not measure pre- and post-exposure performance in single-cue conditions, this interpretation remains unverified.

Our findings are broadly consistent with these modeling studies (Hong et al., 2022; Odegaard et al., 2017): with heightened cue conflict attributed to landmark instability, prior belief in a common cause either remained unchanged or increased. The Bayesian belief updating hypothesis suggests that prior belief in a common cause should be updated in the direction of new stimulus statistics. For example, it should decrease every time a new piece of evidence for cue conflict is sensed (Glasauer, 2019). Accordingly, our findings might be explained in two possible ways.

The first possible explanation is that the total amount of new evidence pointing to different causes actually did not increase in the presence of the large conflict condition. The increased sensory noise level of landmarks, as found in the LC-informed, LC-informed-additional, and LC-feedback groups, could increase the common-cause likelihood in the conflict conditions, as spatially distant distributions would have larger overlap with each other when they are more widespread (Hong et al., 2022). We tested this hypothesis by simulating the likelihood of different causes aggregated across trials and across decision strategies, which represents the total amount of new evidence for different causes experienced by the participant throughout the experiment. The three groups in the main experiment significantly differed on this variable ($F(2,138) = 17.818$, $p < 0.001$, $\eta_p^2 = 0.206$). Specifically, new evidence for different causes was less in the LC-absent group than the LC-present (mean = 0.420 vs. 0.524, $t = 5.182$, $p < 0.001$) and LC-informed groups (mean = 0.420 vs. 0.517, $t = 4.816$, $p < 0.001$), but did not differ between the LC-present and LC-informed groups (mean = 0.524 vs. 0.517, $t = 0.638$, $p = 0.763$). These results indicate that different-cause evidence did increase with heightened cue conflict regardless of awareness of landmark instability, ruling out the first explanation.

The second possible explanation resides in the computational process of belief-updating itself. The cognitive system's reluctance to update prior beliefs may be driven either by an inherent inclination to maintain a stable internal model of the external world or the computational cost incurred by the belief updating process. The belief updating process may only occur when people are explicitly prompted to do so, such as through specific feedback providing detailed information on the frequency of landmark relocation (Cavalan et al., 2023). This detailed information was lacking in our study. On the contrary, adjusting lower-level parameters like cue weighting and sensory noise may be more beneficial, as they allow for localized effects without disrupting the broader internal model. The primary principle of this hypothesis is to maintain the stability of the broader internal model by not incorporating new sensory evidence into causal structure beliefs. This hypothesis explains our finding of unchanged prior beliefs in causal structure despite new evidence indicative of heightened cue conflict in the majority of instances.

However, it remains puzzling why, under one circumstance (LC-feedback group), prior belief in common cause increased despite increased evidence suggestive of separate causes. This perplexing phenomenon can be explained by the predictive coding hypothesis (Odegaard et al., 2017; Talsma, 2015). This hypothesis proposes that discrepancies between the internal model and new sensory evidence are reconciled to cause modifications to the internal model itself, aiming to maintain a stable ultimate perception of the external world. For example, suppose a navigator initially assumes a 40 % cue congruency rate (i.e., prior belief in common cause). If a new trial indicates a 30 % likelihood of cue congruency (i.e., likelihood of common cause), to maintain the original hypothesis of 40 % cue congruency (i.e., posterior belief in common cause), the navigator has to adjust prior belief in common cause by increasing it to 60.87 %.[3] This hypothesis can explain our finding of increased prior belief in common cause despite heightened cue conflict in the LC-feedback group. Which goal the observer prioritizes may depend on the specific experimental context. Future investigation is needed to elucidate the prior belief updating process.

Notably, our findings revealed a weak relationship between prior belief about causal structure and cue-weighting behavior. In most instances (LC-informed and LC-informed-additional groups), reliance on landmarks was reduced but prior beliefs remained unchanged. In one instance (LC-feedback group), reduced reliance on landmarks accompanied increased prior belief in common cause, but this had minimal influence on landmark reliance, as sensory weights did not differ between common-cause and different-cause judgments ($w_{opt,sen} = w_{dc}$). Variations in prior belief, which directly impact the relative frequencies of different cause judgments, should not affect landmark reliance observed in behavior. While reduced reliance on unstable landmarks is a common behavioral finding in spatial navigation studies (Burgess et al., 2004; Chen et al., 2017; Knierim et al., 1995; Lenck-Santini et al., 2002; Zhao & Warren, 2015a), attributing this phenomenon to navigators learning landmark instability statistics and updating the internal model accordingly may be an oversimplification (Roy et al., 2023).

### 7.2. Contributions of self-rated confidence to cue weighting

In the BCI model, same-cause judgments employ a bottom-up cue-weighting strategy, where cue weights are determined by the relative sensory noise levels ($w_{opt,sen}$). In contrast, for different-cause judgments, we incorporated the parameter $w_{dc}$ to model the alternation rate between different cues (de Winkel et al., 2017, 2018). A key question is whether this cue-weighting strategy differs fundamentally from the bottom-up strategy assumed in same-cause judgments.

We found that cue relative confidence ($rc$) was significantly and positively correlated with $w_{dc}$, even after accounting for the sensory optimal weight ($w_{opt,sen}$). This finding means that participants relied more heavily on the cue type they felt more confident with when perceiving sensory conflicts, reflecting a top-down influence from metacognition. This finding is broadly consistent with our previous finding (Chen et al., 2017), but further implies that the top-down influences on cue weighting from metacognition may manifest when navigators sensed cue conflict. In contrast, $w_{dc}$ was generally not positively correlated with $w_{opt,sen}$ in the majority of instances, except in the LC-absent group. We speculate that in this group, minimal cue conflicts may make it difficult to disentangle the cue integration and cue alternation strategies. In one instance, $w_{opt,sen}$ negatively predicted $w_{dc}$ after accounting for $rc$. Together, these findings suggest that cue weighting differs fundamentally between same-cause and different-cause judgments.

Nevertheless, in the current study, the overall effect of self-rated confidence on cue weighting appeared relatively modest compared to our previous study (Chen et al., 2017). This discrepancy may be due to individuals' varying metacognitive abilities concerning different types of path integration cues. Unlike our previous study, which tested body-based self-motion cues for path integration, the current study employed visual optic flow. In everyday situations, people accumulate significantly more experience using body-based self-motion cues than optic flow for navigation. Additionally, the desktop VR environment used here is less immersive than the head-mounted-display in our previous study, potentially limiting participants' metacognitive ability to assess their own performance. These factors may have attenuated the top-down influence of self-confidence on cue weighting.

### 7.3. Explaining cue combination suboptimality within BCI framework

Cue combination suboptimality is frequently observed in spatial navigation (Newman et al., 2023) and other domains (Rahnev & Denison, 2018). The application of the BCI model provide insights into the underlying causes of cue combination suboptimality.

First, we found that the less precise prior knowledge about target location ($\sigma_{pr}$) ameliorated cue combination suboptimality, consistent with previous studies (Aston et al., 2022). Based on the internal mechanisms of the BCI model, $\sigma_{pr}$ contributes to cue combination suboptimality via two routes. In the first route, the target location prior constitutes a common information source shared by different cues, creating correlated error that diminishes the cue integration benefit in behavior (Oruç et al., 2003). A larger $\sigma_{pr}$ lowers correlated error, reducing the influence of the prior knowledge about target location and facilitating the detection of the cue integration benefit in behavior. In the second route, the prior knowledge about target location affects the common-cause likelihood: a larger $\sigma_{pr}$ increases the common-cause judgment ($\rho = 0.257$, $p = 0.002$, across all participants in the main experiment), resulting in higher common-cause posterior probability and more frequent adoption of the cue integration strategy, which in turn ameliorates cue combination suboptimality. While the first route is well-recognized in the literature (Aston et al., 2022; Oruç et al, 2003), the second route has not been mentioned.

Additionally, we found that a stronger prior belief in common cause ($p_{cc,pr}$) corresponded to less severe cue combination sub-

[3] Based on the equation $p(C=1|x_l,x_m) = \frac{p(x_l,x_m|C=1)\times p(C=1)}{p(x_l,x_m|C=1)\times p(C=1)+p(x_l,x_m|C=2)\times p(C=2)}$, we get $0.4 = \frac{0.3\times p(C=1)}{0.3\times p(C=1)+0.7\times(1-p(C=1))}$. Hence, $p\ (C=1) = 60.87\%$.

optimality. Within the BCI model, $p_{cc,pr}$ influences cue combination suboptimality by affecting the common-cause posterior probability ($p_{cc,post}$). Specifically, a stronger prior belief in common cause ($p_{cc,pr}$) increased the common-cause posterior probability and the number of trials in which the cue integration strategy is adopted, which in turn lessens cue combination suboptimality.

These findings may explain one notable discrepancy between the current study and previous studies on spatial cue combination. Although the current study revealed cue integration effect in the combination and small conflict conditions, this effect was suboptimal. In contrast, previous studies demonstrated optimal or near-optimal integration effect (Chen et al., 2017; Nardini et al., 2008; Sjolund et al., 2018; Zhao & Warren, 2015b). The prior knowledge about target location might be more precise in the current study than previous studies due to the substantially narrower range of target locations in the current study, which would cause cue combination suboptimality. However, it is challenging to determine whether prior belief about causal structure was lower in the current study than previous studies, as factors influencing this variable are not well understood.

Beyond the parameters included in the current BCI model, shared sensory noise between the single-cue conditions must also be considered. Although not incorporated in the current model, shared sensory noise contributes to cue combination suboptimality by introducing correlated errors between cues (Oruç et al., 2003), similar to the influence of prior knowledge about target location (Aston et al., 2022). Shared sensory noise might have been more pronounced in the current study, which contrasted visual optic flow and visual landmarks that belong to the same visual modality. In contrast, previous studies contrasted body-based self-motion cues and landmarks, which belong to distinct sensory modalities with minimal overlap along the sensory processing pathway (Chen et al., 2017; Nardini et al., 2008; Sjolund et al., 2018; Zhao & Warren, 2015b). Note that while both optic flow and body-based self-motion cues can contribute to path integration, they operate through distinct sensory pathways, and there is no direct visual overlap between them. Future work should consider incorporating shared sensory noise into the BCI model. However, doing so might introduce redundancy with existing parameters, such as motor noise and prior knowledge about target location. All these factors contribute to correlated errors between cues, albeit at different processing stages.

## 8. Conclusions and future directions

This study investigated how cue conflict influences the cognitive processes underlying spatial navigation, focusing on the interplay between conflicting landmark cues and optic-flow-based path integration. Using the BCI model, which outperformed an alternative model lacking prior information, we gained novel insights into the psychological processes affected by cue conflict. Our findings advance understanding of the cognitive mechanisms governing navigation in conflicting-cue scenarios and underscore the BCI model's potential as a framework for exploring spatial cue interactions.

Despite these insights, several questions remain for future research. First, one critical area is characterizing the prior knowledge of target location. For example, is this knowledge cue-independent or cue-specific in a multi-cue environment? While the current BCI model employs a global prior of target location, accumulating across trials, other types of spatial priors may also be at play. How would these priors affect navigation behavior?

Second, our findings suggest that cue conflict attributed to unstable landmarks reduces the weight assigned to them, yet the mechanisms behind this devaluation require further exploration. Given our finding that unstable landmarks needed to be devalued to have lowered weight assigned to them, integrating loss functions associating lower benefit with unstable landmarks into the BCI framework might address this gap (McNamara & Chen, 2022).

Third, while the present study focused on simple navigation with one-dimensional distance estimation, future work can extend the BCI model to more complex navigation tasks, such as two-dimensional navigation (McNamara & Chen, 2022, Appendix B) or paradigms involving discrete stimulus and response distributions (Cheng, 1986; Lenck-Santini et al., 2001). Such extension could deepen our understanding of the basic principles of spatial navigation.

Finally, consistent with many previous studies on spatial navigation (Wolbers & Hegarty, 2010), the current work also revealed substantial individual differences within the context of BCI model, including sensory noise levels, weighting strategy in the different cause judgment, precision of the prior knowledge of target location, and prior belief about causal structure. We leveraged these individual differences to understand relationships between various cognitive constructs and behavioral phenomena. For instance, cue relative confidence was correlated with cue weighting in the different-cause judgment, suggesting that *meta*-cognitive processes influenced cue weighting once conflict is detected. As another example, people with more precise prior knowledge about target location exhibited greater cue combination suboptimality. Future work should explore the determinants of these differences, such as variations in memory capacity, preferred spatial learning strategies, daily navigation experiences, and etc.

## CRediT authorship contribution statement

**Xiaoli Chen:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Yingyan Chen:** Methodology, Investigation, Data curation. **Timothy P. McNamara:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

## Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 3.5 in order to improve the language and readability of the

manuscript, with caution. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This research was supported by research grants from National Natural Science Foundation of China (NSFC, #32100839), STI 2030 – Major Projects (#2021ZD0200409), Cao Guangbiao High Science and Technology Foundation, Zhejiang University (ZJU, 2020QN002), and National Science Foundation Grant (US, #2217889). We thank Yang Zhang (张阳), Chen Zhang (张琛), and Xialv Ye (叶夏绿) for their help in data collection.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2025.101734.

## Data availability

Data will be made available on request.

## References

Anastasiou, C., Baumann, O., & Yamamoto, N. (2023). Does path integration contribute to human navigation in large-scale space? *Psychonomic Bulletin & Review, 30*(3), 822–842. https://doi.org/10.3758/s13423-022-02216-8

Aston, S., Negen, J., Nardini, M., & Beierholm, U. (2022). Central tendency biases must be accounted for to consistently capture Bayesian cue combination in continuous response data. *Behavior Research Methods, 54*, 508–521. https://doi.org/10.1101/2021.03.12.434970

Auger, S. D., Zeidman, P., & Maguire, E. A. (2015). A central role for the retrosplenial cortex in de novo environmental learning. *eLife*, *4*, e09031.

Badde, S., Navarro, K. T., & Landy, M. S. (2020). Modality-specific attention attenuates visual-tactile integration and recalibration effects by reducing prior expectations of a common source for vision and touch. *Cognition, 197*, Article 104170. https://doi.org/10.1016/j.cognition.2019.104170

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & Psychophysics, 29*(6), 578–584. https://doi.org/10.3758/bf03207374

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

Bromiley, P. A. (2013). *Products and convolutions of Gaussian probability density functions*. https://api.semanticscholar.org/CorpusID:18045887.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology, 44*(1), 108–132. https://doi.org/10.1006/jmps.1999.1279

Bruns, P., Dinse, H. R., & Röder, B. (2020). Differential effects of the temporal and spatial distribution of audiovisual stimuli on cross-modal spatial recalibration. *The European Journal of Neuroscience, 52*(7), 3763–3775. https://doi.org/10.1111/ejn.14779

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.

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. *Journal of Neurophysiology, 103*(6), 3054–3069.

Campbell, M. G., Attinger, A., Ocko, S. A., Ganguli, S., & Giocomo, L. M. (2021). Distance-tuned neurons drive specialized path integration calculations in medial entorhinal cortex. *Cell Reports, 36*(10). https://doi.org/10.1016/j.celrep.2021.109669

Campbell, M. G., Ocko, S. A., Mallory, C. S., Low, I. I. C., Ganguli, S., & Giocomo, L. M. (2018). Principles governing the integration of landmark and self-motion cues in entorhinal cortical codes for navigation. *Nature Neuroscience, 21*(8), 1096–1106. https://doi.org/10.1038/s41593-018-0189-y

Cavalan, Q., de Gardelle, V., & Vergnaud, J.-C. (2023). No evidence of biased updating in beliefs about absolute performance: A replication and generalization of Grossman and Owens (2012). *Journal of Economic Behavior & Organization, 211*, 530–548. https://doi.org/10.1016/j.jebo.2023.05.010

Caves, E. M., Brandley, N. C., & Johnsen, S. (2018). Visual acuity and the evolution of signals. *Trends in Ecology & Evolution, 33*(5), 358–372. https://doi.org/10.1016/j.tree.2018.03.001

Chen, G. F., King, J. A., Burgess, N., & O'Keefe, J. (2013). How vision and movement combine in the hippocampal place code. *Proceedings of the National Academy of Sciences of the United States of America, 110*(1), 378–383.

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

Chen, X., Vieweg, P., & Wolbers, T. (2019). Computing distance information from landmarks and self-motion cues—Differential contributions of anterior-lateral vs Posterior-medial entorhinal cortex in humans. *NeuroImage, 202*, Article 116074. https://doi.org/10.1016/j.neuroimage.2019.116074

Chen, X., Wei, Z., & Wolbers, T. (2022). Coexistence of cue-specific and cue-independent spatial representations for landmarks and self-motion cues in human retrosplenial cortex. *bioRxiv*, 2022.05.16.491990. doi: 10.1101/2022.05.16.491990.

Chen, X., Wei, Z., & Wolbers, T. (2024). Repetition suppression reveals cue-specific spatial representations for landmarks and self-motion cues in the human retrosplenial cortex. *eNeuro, 11*(4), Article ENEURO.0294-23.2024. https://doi.org/10.1523/ENEURO.0294-23.2024

Chen, X., Wei, Z., & Wolbers, T. (2025). Representational similarity analysis reveals cue-independent spatial representations for landmarks and self-motion cues in human retrosplenial cortex. *Imaging Neuroscience, 3*, Article imag_a_00516. https://doi.org/10.1162/imag_a_00516

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

Cheng, K. (2008). Whither geometry? Troubles of the geometric module. *Trends in Cognitive Sciences, 12*(9), 355–361.

Cheng, K., Huttenlocher, J., & Newcombe, N. S. (2013). 25 years of research on the use of geometry in spatial reorientation: A current theoretical perspective. *Psychonomic Bulletin & Review, 20*(6), 1033–1054.

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

Cheung, A., Stürzl, W., Zeil, J., & Cheng, K. (2008). The information content of panoramic images II: View-based navigation in nonrectangular experimental arenas. *Journal of Experimental Psychology: Animal Behavior Processes, 34*(1), 15–30.

Chrastil, E. R., Nicora, G. L., & Huang, A. (2019). Vision and proprioception make equal contributions to path integration in a novel homing task. *Cognition, 192*, Article 103998.
de Winkel, K. N., Katliar, M., & Bülthoff, H. H. (2017). Causal inference in multisensory heading estimation. *PLOS ONE, 12*(1), Article e0169676. https://doi.org/10.1371/journal.pone.0169676
de Winkel, K. N., Katliar, M., Diers, D., & Bülthoff, H. H. (2018). Causal inference in the perception of verticality. *Scientific Reports, 8*(1), 5483. https://doi.org/10.1038/s41598-018-23838-w
Ericson, J. D., & Warren, W. H. (2020). Probing the invariant structure of spatial knowledge: Support for the cognitive graph hypothesis. *Cognition, 200*, Article 104276. https://doi.org/10.1016/j.cognition.2020.104276
Fischer, L. F., Soto-Albors, R. M., Buck, F., & Harnett, M. T. (2020). Representation of visual landmarks in retrosplenial cortex. *eLife, 9*, e51458.
French, R. L., & DeAngelis, G. C. (2020). Multisensory neural processing: From cue integration to causal inference. *Vision Physiology, 16*, 8–13. https://doi.org/10.1016/j.cophys.2020.04.004
Gallistel, C. R. (1990). *The organization of learning*. The MIT Press.
Gau, R., & Noppeney, U. (2016). How prior expectations shape multisensory perception. *NeuroImage, 124*, 876–886. https://doi.org/10.1016/j.neuroimage.2015.09.045
Glasauer, S. (2019). Chapter 1—Sequential Bayesian updating as a model for human perception. In S. Ramat & A. G. Shaikh (Eds.), *Progress in Brain Research* (Vol. 249, pp. 3–18). Elsevier. doi: 10.1016/bs.pbr.2019.04.025.
Gothard, K. M., Skaggs, W. E., & McNaughton, B. L. (1996). Dynamics of mismatch correction in the hippocampal ensemble code for space: Interaction between path integration and environmental cues. *Journal of Neuroscience, 16*(24), 8027–8040.
Harootonian, S. K., Ekstrom, A. D., & Wilson, R. C. (2022). Combination and competition between path integration and landmark navigation in the estimation of heading direction. *PLOS Computational Biology, 18*(2), Article e1009222. https://doi.org/10.1371/journal.pcbi.1009222
Hastie, T. (2009). *The elements of statistical learning: Data mining, inference, and prediction*.
Hollingworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology & Scientific Methods, 7*, 461–469. https://doi.org/10.2307/2012819
Hong, F., Badde, S., & Landy, M. S. (2022). Repeated exposure to either consistently spatiotemporally congruent or consistently incongruent audiovisual stimuli modulates the audiovisual common-cause prior. *Scientific Reports, 12*(1), 15532. https://doi.org/10.1038/s41598-022-19041-7
*JASP Team (2023). JASP (Version 0.17.1)*. (n.d.). [Computer software].
Jetzschke, S., Ernst, M. O., Froehlich, J., & Boeddeker, N. (2017). Finding home: Landmark ambiguity in human navigation. *Frontiers in Behavioral Neuroscience, 11*. https://doi.org/10.3389/fnbeh.2017.00132
Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. *Current Opinion in Neurobiology, 13*(2), 150–158. https://doi.org/10.1016/s0959-4388(03)00042-4
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.
Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLOS ONE, 2*(9), e943.
Kuehn, E., Chen, X., Geise, P., Oltmer, J., & Wolbers, T. (2018). Social targets improve body-based and environment-based strategies during spatial navigation. *Experimental Brain Research, 236*(3), 755–764.
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.
Lenck-Santini, P. P., Save, E., & Poucet, B. (2001). Place-cell firing does not depend on the direction of turn in a Y-maze alternation task. *The European Journal of Neuroscience, 13*(5), 1055–1058. https://doi.org/10.1046/j.0953-816x.2001.01481.x
Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods, 49*(2), 513–537. https://doi.org/10.3758/s13428-016-0740-2
Lew, A. R. (2011). Looking beyond the boundaries: Time to put landmarks back on the cognitive map? *Psychological Bulletin, 137*(3), 484–507.
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), 73–91.
Mao, D., Kandler, S., McNaughton, B. L., & Bonin, V. (2017). Sparse orthogonal population representation of spatial context in the retrosplenial cortex. *Nature Communications, 8*(1), 243. https://doi.org/10.1038/s41467-017-00180-9
Rizzo, M. L., & Székely, G. J. (2016). Energy distance. *Wiley Interdisciplinary Reviews: Computational Statistics, 8*(1), 27–38.
McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746–748. https://doi.org/10.1038/264746a0
McNamara, T. P., & Chen, X. (2022). Bayesian decision theory and navigation. *Psychonomic Bulletin & Review, 29*(3), 721–752. https://doi.org/10.3758/s13423-021-01988-9
Miller, N. Y., & Shettleworth, S. J. (2007). Learning about environmental geometry: An associative model. *Journal of Experimental Psychology. Animal Behavior Processes, 33*(3), 191–212. https://doi.org/10.1037/0097-7403.33.3.191
Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691–692. https://doi.org/10.1093/biomet/78.3.691
Nahorna, O., Berthommier, F., & Schwartz, J.-L. (2012). Binding and unbinding the auditory and visual streams in the McGurk effect. *The Journal of the Acoustical Society of America, 132*(2), 1061–1077. https://doi.org/10.1121/1.4728187
Nahorna, O., Berthommier, F., & Schwartz, J.-L. (2015). Audio-visual speech scene analysis: Characterization of the dynamics of unbinding and rebinding the McGurk effect. *The Journal of the Acoustical Society of America, 137*(1), 362–377. https://doi.org/10.1121/1.4904536
Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology, 18*(9), 689–693.
Newcombe, N. S. (2023). What have we learned from research on the "geometric module"? *Learning & Behavior, 52*(1), 14–18. https://doi.org/10.3758/s13420-023-00617-w
Newman, P. M., & McNamara, T. P. (2022). Integration of visual landmark cues in spatial memory. *Psychological Research, 86*(5), 1636–1654. https://doi.org/10.1007/s00426-021-01581-8
Newman, P. M., Qi, Y., Mou, W., & McNamara, T. P. (2023). Statistically optimal cue integration during human spatial navigation. *Psychonomic Bulletin & Review, 30*(5), 1621–1642. https://doi.org/10.3758/s13423-023-02254-w
Odegaard, B., Wozny, D. R., & Shams, L. (2017). A simple and efficient method to enhance audiovisual binding tendencies. *PeerJ, 5*, e3143.
O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford University Press.
Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*(23), 2451–2468.
Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *The Journal of Neuroscience, 31*(47), 17220. https://doi.org/10.1523/JNEUROSCI.2028-11.2011
Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences, 19*(5), 285–293. https://doi.org/10.1016/j.tics.2015.03.002
Qi, Y., & Mou, W. (2024). Relative cue precision and prior knowledge contribute to the preference of proximal and distal landmarks in human orientation. *Cognition, 247*, Article 105772. https://doi.org/10.1016/j.cognition.2024.105772
Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163. JSTOR. doi: 10.2307/271063.
Raftery, A. E., Madigan, D., & Volinsky, C. T. (1996). Accounting for model uncertainty in survival analysis improves predictive performance. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 5: Proceedings of the Fifth Valencia International Meeting* (p. 0). Oxford University Press. doi: 10.1093/oso/9780198523567.003.0017.
Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences, 41*, e223.

Ratliff, K. R., & Newcombe, N. S. (2008). Reorienting when cues conflict: Evidence for an adaptive-combination view. *Psychological Science, 19*(12), 1301–1307. https://doi.org/10.1111/j.1467-9280.2008.02239.x
Reifman, A., & Keyton, K. (2010). Winsorize. In N. J. Salkind (Ed.), *Encyclopedia of Research Design* (pp. 1636–1638). Thousand Oaks, CA: Sage.
Riemer, M., Achtzehn, J., Kuehn, E., & Wolbers, T. (2022). Cross-dimensional interference between time and distance during spatial navigation is mediated by speed representations in intraparietal sulcus and area hMT+. *NeuroImage, 257*, 1–12. https://doi.org/10.1016/j.neuroimage.2022.119336
Rohde, M., van Dam, L. C. J., & Ernst, M. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research, 29*(4–5), 279–317. https://doi.org/10.1163/22134808-00002510
Rohe, T., & Noppeney, U. (2015). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision, 15*(5), 22. https://doi.org/10.1167/15.5.22
Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review, 21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4
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.
Roy, C., Wiebusch, D., Botsch, M., & Ernst, M. O. (2023). Did it move? Humans use spatio-temporal landmark permanency efficiently for navigation. *Journal of Experimental Psychology: General, 152*(2), 448–463.
Saleem, A. B., Diamanti, E. M., Fournier, J., Harris, K. D., & Carandini, M. (2018). Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature, 562*(7725), 124–127. https://doi.org/10.1038/s41586-018-0516-1
Shapiro, M. L., Tanila, H., & Eichenbaum, H. (1997). Cues that hippocampal place cells encode: Dynamic and hierarchical representation of local and distal stimuli. *Hippocampus, 7*(6), 624–642.
Shettleworth, S. J., & Sutton, J. E. (2005). Multiple systems for spatial learning: Dead reckoning and beacon homing in rats. *Journal of Experimental Psychology: Animal Behavior Processes, 31*(2), 125–141.
Siegel, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. *Advances in Child Development and Behavior, 10*, 9–55.
Sjolund, L. A., Kelly, J. W., & McNamara, T. P. (2018). Optimal combination of environmental cues and path integration during navigation. *Memory & Cognition, 46*(1), 89–99.
Talsma, D. (2015). Predictive coding and multisensory integration: An attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience, 9*. https://www.frontiersin.org/articles/10.3389/fnint.2015.00019.
Tanila, H., Shapiro, M. L., & Eichenbaum, H. (1997). Discordance of spatial representation in ensembles of hippocampal place cells. *Hippocampus, 7*(6), 613–623.
Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189–208.
Tong, J., Li, L., Bruns, P., & Röder, B. (2020). Crossmodal associations modulate multisensory spatial integration. *Attention, Perception & Psychophysics, 82*(7), 3490–3506. https://doi.org/10.3758/s13414-020-02083-2
Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank & I. Campari (Eds.), *Spatial Information Theory: A Theoretical Basis for GIS* (pp. 14–24). Springer Berlin Heidelberg.
Umbach, G., Kantak, P., Jacobs, J., Kahana, M., Pfeiffer, B. E., Sperling, M., & Lega, B. (2020). Time cells in the human hippocampus and entorhinal cortex support episodic memory. *Proceedings of the National Academy of Sciences of the United States of America, 117*(45), 28463–28474. https://doi.org/10.1073/pnas.2013250117
Van Wanrooij, M. M., Bremen, P., & John Van Opstal, A. (2010). Acquired prior knowledge modulates audiovisual integration. *European Journal of Neuroscience, 31*(10), 1763–1771. https://doi.org/10.1111/j.1460-9568.2010.07198.x
Wang, L., Mou, W., & Dixon, P. (2018). Cue interaction between buildings and street configurations during reorientation in familiar and unfamiliar outdoor environments. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 44*(4), 631–644. https://doi.org/10.1037/xlm0000478
Wolbers, T., & Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences, 14*(3), 138–146.
Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLoS Computational Biology, 6*(8), Article e1000871. https://doi.org/10.1371/journal.pcbi.1000871
Xu, Y., Regier, T., & Newcombe, N. S. (2017). An adaptive cue combination model of human spatial reorientation. *Cognition, 163*, 56–66. https://doi.org/10.1016/j.cognition.2017.02.016
Yoganarasimha, D., Yu, X., & Knierim, J. J. (2006). Head direction cell representations maintain internal coherence during conflicting proximal and distal cue rotations: Comparison with hippocampal place cells. *Journal of Neuroscience, 26*(2), 622–631. https://doi.org/10.1523/JNEUROSCI.3885-05.2006
Zhao, M., & Warren, W. H. (2015a). Environmental stability modulates the role of path integration in human navigation. *Cognition, 142*, 96–109.
Zhao, M., & Warren, W. H. (2015b). How you get there from here: Interaction of visual landmarks and path integration in human navigation. *Psychological Science, 26*(6), 915–924.