The source of systematic errors in human path integration

Article *in* Journal of Experimental Psychology Human Perception & Performance · September 2022

DOI: 10.7939/r3-ftnc-

Yafei Qi
University of Alberta

Manuscript

Accepted by *Journal of Experimental Psychology: Human Perception and Performance*

Sources of systematic errors in human path integration

Yafei Qi, Weimin Mou

Department of Psychology, University of Alberta

This study is supported by a grant from NSERC to Weimin Mou.

Correspondence concerning this article should be addressed to Yafei Qi or Weimin Mou, P217 Biological Sciences Building, University of Alberta, Edmonton AB T6G 2E9, Canada. E-mail: yyqi@ualberta.ca or wmou@ualberta.ca.

The Matlab code for 5 × 2 cross-validation is placed at https://doi.org/10.7939/r3-ftnc-8m22.

Word Count: 18038

Abstract

Triangle completion is a task widely used to study human path integration, an important navigation method relying on idiothetic cues. Systematic biases (compression patterns in the inbound responses) have been well documented in human triangle completion. However, the sources of systematic biases remain controversial. We used cross-validation modeling to compare three plausible theoretical models that assume that systematic errors occur in the encoding outbound path solely (encoding-error model), executing the inbound responses solely (execution-error model), and both (bi-component model), respectively. The data for cross-validation modeling are from a previous study (Qi et al., 2021), in which participants learned three objects' locations (one at the path origin, that is, home) very well before walking each outbound path and then pointed to the objects' original locations after walking the outbound path. The modeling algorithm used one inbound response (i.e., response to the home) or multiple inbound responses (i.e., responses to two non-home locations and the home) for each outbound path. The algorithm of using multiple inbound responses demonstrated that the bi-component model outperformed the other models in accounting for the systematic errors. This finding suggests that both encoding the outbound path and executing the inbound responses contribute to the systematic biases in human path integration. In addition, the results showed that the algorithm using only the home response could not distinguish among these three models, suggesting that the typical triangle-completion task with only the home response for each outbound path cannot determine the sources of the systematic biases.

Keywords: path integration; encoding-error model; execution-error model, cross-validation; triangle completion

Public Significance Statements

The cross-validation modeling of this study demonstrated that human systematic errors in returning to the path origin after walking an outbound path came from both encoding the outbound path and executing the return path, which unified two opposite models in the literature, the encoding-error model attributing the errors to encoding the outbound path solely and the execution-error model attributing the errors to executing the return path solely.

Demonstrating that cross-validation algorithm using multiple responses but not that using home response only for each outbound path could determine the bi-component model, this study also provides important contributions to the research methods to study human path integration.

1. Introduction

Path integration is the navigation process that employs idiothetic cues (i.e., proprioception, vestibular, and optic flow) and integrates the distances traveled and angles turned during motion so that navigators can continuously update their position and heading with respect to fixed reference locations in space (Etienne et al., 1996; Mittelstaedt & Mittelstaedt, 1982). The fixed locations can be the origin of the path traveled (e.g., the nest for an animal who is out for foraging) or remembered important locations in the environment (e.g., the grocery store for a human individual who will visit later after traveling from home to office). Thus, path integration plays an important role in navigation, especially when allothetic cues (e.g., visual landmarks) are scarce or navigation occurs in darkness (Klatzky et al., 1998).

Path integration is ubiquitous among mobile animals, including ants (Müller & Wehner, 1988), bees (Collett & Collett, 2000), rodents (Etienne & Jeffery, 2004), birds (Saint Paul, 1982), mammals (Mittelstaedt & Mittelstaedt, 1980), and humans (Loomis et al., 1999). Critically, path integration has been suggested as one important means of constructing spatial knowledge of the environment (Gallistel, 1990). By tracking the path lengths and turn angles, and linking routes between known places, path integration enables one to acquire a labeled graph that incorporates local metric information (Chrastil & Warren, 2014; Warren et al., 2017) or a cognitive map that includes globally consistent metric information (Jacobs & Schenk, 2003; Wang, 2016).

Path integration is not an error-free process. Errors in path integration can be accumulated quickly with the increase of the complexity of the path, for example with the increase of the number of legs in a path (Kelly et al., 2008; Rieser & Rider, 1991). Previous studies using triangle-completion tasks have found that the human participants' homebound behavior exhibits systematic distortion (Kearns et al., 2002; Klatzky et al., 1999; Loomis et al., 1993). In the triangle-completion task, participants walked an outbound path, which consists of two linear segments and a turn angle between them, and then returned to or pointed to the origin of the outbound path (Klatzky et al., 1998; Loomis et al., 1993). Participants' responses of the inbound path (i.e., homing vector) include the turn angle and path length. Participants usually overshot small values, and conversely, undershot large values, showing a compression pattern relative to the correct values of both turn angle and path length. This systematic distortion was distinguished from random errors (Chrastil & Warren, 2017; Harootonian et al., 2020).

A compression pattern relative to the correct values has been widely and long reported in magnitude judgments of various types of stimuli including size, weight, brightness, loudness, and duration (Stevens & Greenbaum, 1966). Stevens and Greenbaum (1966) referred to the compression pattern as the regression effect and attributed the effect primarily to participants' tendency to shrink the judgment range under their control. Other researchers attributed the compression pattern to the stimulus range controlled by experimenters (e.g., Teghtsoonian & Teghtsoonian, 1978). Petzschner and Glasauer (2011) proposed a Bayesian model to explain the compression pattern in reproducing a previewed distance or angle. Participants in their study walked a distance to approach a visible target or turned an angle to face a visible target. They then reproduced the distance or angle without the presence of the target. The results showed that participants biased their reproduced magnitudes towards the mean of the previewed magnitudes. Hence, participants not only used the perceived magnitudes in the specific trial but also used the prior distribution of the magnitudes (Harootonian et al., 2022; McNamara & Chen, 2021). The prior knowledge could be learned from past trials (see also Harootonian et al., 2020). Note that other studies suggested that prior knowledge could be primarily determined by experiences outside the experiment (e.g., categorical information, Huttenlocher et al., 1991).

A strict Bayesian approach assumes that separate estimates of the true value (prior or perceived magnitude) are combined in judgment but do not change the representation of the perceived magnitude (Zhang & Mou, 2017). Hence, the representation of the perceived magnitude should be free of compression. However, to explain the compression pattern reported in the triangle-completion task, researchers proposed that compression could occur both in executing the inbound path (Chrastil & Warren, 2021) and in encoding the outbound path (Fujita et al., 1993; Harootonian et al., 2020). The latter proposal implies that participants might use the Bayesian inference in encoding rather than in response. Thus, examining the sources of the compression pattern reported in triangle completion is not only theoretically important in human navigation but also in broad fields of experimental psychology.

Performing the triangle-completion task requires three cognitive stages (Fujita et al., 1993). The initial stage involves sensing the traversed path and forming internal representations of leg lengths and turn angles, referred to as the encoding process. In the second stage, the internalized representations are employed to compute the desired inbound responses (i.e., inbound path length and turn angle), referred to as the integration process. Ultimately, the desired inbound response is executed, referred to as the execution process. The important yet inconclusive theoretical question is which stage or stages the systematic errors originate from (Chrastil & Warren, 2021; Fujita et al., 1993; Harootonian et al., 2020). Answering this question is important to advance our understanding of the nature of human path integration.

One intuitive answer is that systematic errors in the inbound path length and turn angle originated from the execution process. However, Klatzky, Loomis, and their colleagues (Fujita et al., 1993; Klatzky et al., 1999; Loomis et al., 1999) provided innovative insights that systematic errors in encoding the outbound path can also well explain the systematic errors appearing in the inbound path length and turn angle. Their influential model, the encoding-error model, assumes that while the systematic errors originate from encoding the outbound path, the subsequent processes, i.e., computing the desired inbound responses via cognitive trigonometry and executing it, are free of systematic errors (Fujita et al., 1993).

There are three important theoretical contributions of the encoding-error model. First, it indicates that *counter-intuitively* the systematic errors appearing in response measures may not originate from execution and instead from encoding. Second, it suggests that human path integration may significantly differ from animal path integration. Animals may only represent and update the homing vector but do not encode the outbound path in memory (e.g., Benhamou & Séguinot, 1995; Etienne & Jeffery, 2004; Wehner et al., 1996). This type of spatial updating is referred to as continuous updating. In contrast, spatial updating with encoding of the outbound path in memory is referred to as configural updating (He & McNamara, 2018; Loomis et al., 1999; Wiener et al., 2011). Hence, while researchers hypothesize that animal path integration is continuous updating (Wiener et al., 2011, p. 62), the encoding-error model suggests that human path integration is configural updating. Last, the encoding-error model suggests that humans can develop configural knowledge of the outbound path. This configural knowledge is different from route knowledge because the configural knowledge can support a novel short-cut between two points on the outbound path and thus is more like a survey (map-like) knowledge. Therefore, path integration can be a means to develop map-like knowledge (Gallistel, 1990).

More specifically, the encoding-error model stipulates that there are two linear functions, the encoding function of leg lengths and the encoding function of turn angles, which determine the encoded values from the actual values of the outbound path. Each encoding function has two parameters, the slope, and the intercept. Therefore, for each given outbound path, the corresponding internal representation of the path can be described by the encoding functions. As a result, the desired inbound response can be calculated from the encoding functions assuming no systematic bias in the integration process. Given that the desired inbound response is executed without systematic bias, the encoding-error model can predict the participants' inbound response, at least on average. Fujita et al. (1993) fit the encoding-error model with empirical data of triangle completion. They estimated the parameters of the encoding functions by minimizing the discrepancy between the model's predictions and participants' actual responses to the path origins. For both functions, the slope tended to be smaller than 1 and the intercept tended to be larger than 0, showing a compression pattern of the encoded values relative to the correct values. Moreover, the modeling results showed that the encoding-error model fit the data very well. The performance of the encoding-error model was still impressive when data from other studies were applied, suggesting that encoding distortion captured the path integration errors under a variety of situations (Klatzky et al., 1999; May & Klatzky, 2000; Péruch et al., 1997; Wartenberg et al., 1998).

However, the demonstration that systematic distortion can be attributed to the encoding component (Fujita et al., 1993) does not exclude the possibility that systematic distortion can also be attributed to the execution component alone (referred to as the execution-error model). Intuitively, an execution-error model stipulating that execution errors follow a compression pattern (a linear function to predict the response values from the correct values with a slope less than 1 and an intercept larger than 0) can readily explain the observed compression pattern of the response values relative to the correct values. Thus, it is challenging to dissociate the encoding-error model from the execution-error model empirically. We speculate that due to this challenge, Fujita et al. (1993) did not contrast the encoding-error model with the execution-error model to prove the relative superiority of the encoding-error model. Although testing the encoding-error model is theoretically critical, no other modeling work had been conducted to further test the encoding-error model until two recent studies reported by Harootonian et al. (2020) and Chrastil and Warren (2021).

Harootonian et al. (2020) still assumed that systematic errors occur in the encoding process rather than in the integration or execution process, similar to the original encoding-error model. However, they proposed that the systematic errors primarily occur in encoding the leg lengths but not in encoding the turn angles whereas the original encoding-error model claimed systematic errors in both leg lengths and turn angles of the outbound path. Furthermore, different encoding functions were used for the lengths of the first and the second legs whereas the original encoding-error model used one common function for both legs. They fit their model and the original encoding-error model to data in a triangle-completion task in which participants returned home after walking an outbound path on an omnidirectional treadmill. The model comparison results showed superior performance of their model over the original encoding-error model. However, as designed to examine variants of the encoding-error model, this study still cannot distinguish between the encoding-error model and the execution-error model.

More relevantly, Chrastil and Warren (2021) tested models of encoding errors solely, execution errors solely, and both types of errors. In their study, participants did both simple tasks (e.g., distance or angle reproduction tasks) and triangle-completion tasks. They used data of *reproduction tasks* to estimate the parameters of the encoding and execution functions for triangle-completion tasks. Then the three models, using the corresponding functions (e.g., an encoding-error model used the encoding functions), generated the predictions for the inbound response errors in the triangle-completion task. The results of the model comparison showed that the execution-error model performed better than the encoding-error model. Furthermore, the model including both types of errors did not perform better than the execution-error model. These results suggest that the observed systematic errors in inbound responses were sufficiently explained by the systematic errors in executing the inbound path, but not by the systematic errors in encoding the outbound path. The finding of Chrastil and Warren (2021) is theoretically important as it is the first modeling work clearly indicating that systematic errors in the human triangle-completion task are not solely contributed to the encoding errors, undermining the key argument of the encoding-error model (Fujita et al., 1993).

However, the finding of Chrastil and Warren (2021) could not decisively lead to the conclusion that systematic errors in inbound responses are primarily attributed to systematic execution errors either. One critical concern is whether the reproduction tasks that Chrastil and Warren (2021) employed could truly measure parameters for the *pure* encoding and execution functions. In particular, in their reproduction tasks, participants walked a distance or turned an angle (encoding path). After being stopped by a sound, they reproduced the distance or the angle (response path). By assuming that there were only systematic encoding errors in the encoding path or only systematic execution errors in the response path, Chrastil and Warren separately estimated the parameters of the encoding and execution functions from the reproduction tasks. However, their assumption may be inaccurate because there could be both systematic errors in encoding and execution (Chrastil & Warren, 2014).

Chrastil and Warren (2021) also measured the distance error in a blind-walking task. They then subtracted the errors in the blind-walking task from the errors in the reproduction task to get the pure encoding function. Specifically, in blind-walking, participants perceived an egocentric distance visually and then walked an equivalent distance while being blindfolded

(Chrastil & Warren, 2014). Assuming that there were no systematic encoding errors in perceiving an egocentric distance visually and considering that the response path was the same in the blind walking and the reproduction task, Chrastil and Warren attributed the difference of the errors in these two tasks to the pure encoding errors. Nevertheless, visual perceiving distance may introduce systematic encoding errors. Previous research suggested that there is a calibration/recoupling between locomotor displacement and the visually perceived distance (Rieser et al., 1990; 1995), hence systematic encoding errors in locomotion may also occur in visual perceiving distance. Consequently, these methods were not perfect to estimate either encoding or execution functions if there were indeed both systematic encoding and execution errors. In addition, one may be also wondering whether the functions derived from the reproduction tasks are the same as those used in a much more complicated triangle-completion task.

Therefore, the sources of systematic biases in the inbound responses of the triangle-completion task are still not clear. The primary purpose of the current study was to further test the sources of systematic biases. Adopting a model cross-validation approach (Arlot & Celisse, 2010; Refaeilzadeh et al., 2009), we tested three models: the encoding-error model, the execution-error model, and a bi-component model with both encoding and execution biases. We used the data of the triangle-completion task from Qi et al. (2021) for both model fitting and model validation. In the step of model fitting, we used half data to estimate the parameters of different models (i.e., encoding functions for the encoding-error model, execution functions for the execution-error model, and both functions for the bi-component model). In the step of model validation, we compared the performance of the three models in explaining the other half data. Because we estimated the parameters of encoding/execution functions directly using the data of the triangle-completion task, we avoided the issues of estimating encoding/execution functions from other independent tasks (e.g., reproduction tasks) discussed above.

Note that in a typical triangle-completion task, participants had one inbound response (i.e., homing vector) for each outbound path. Mou and Zhang (2014) indicated that from only one inbound response, researchers cannot correctly recover (or calculate) participants' representations of their positions and orientations that guide their inbound responses at the end of the outbound path. They argued that many possible pairs of position and orientation representations at the end of the outbound path could lead to the same homing vector. Because position and orientation representations at the end of the outbound path are not only the outcome of the represented outbound path but also determine the desired inbound responses, we conjectured that from one inbound response, we could not determine the represented outbound path and desired inbound responses. Mou and Zhang (2014) further demonstrated that from multiple inbound responses, they could calculate participants' representations of their position and orientation at the end of the outbound path (see also Qi et al., 2021; Zhang & Mou, 2017; Zhang et al., 2020). Following this result, we conjectured that from multiple inbound responses for one single outbound path, we could determine the represented outbound path and the desired inbound responses and then could estimate the encoding and execution functions. Unlike the typical triangle-completion task in which participants only need to make a single response (i.e., the homing vector), participants in Qi et al. (2021) were required to indicate multiple locations (including home location) that they had learned before walking a two-segment path. Thus, using the data from Qi et al. (2021), the current study validated models using multiple inbound responses for each outbound path.

2. Current study

2.1 Description of the data

The data used for model fitting and model validation in the current study came from the path integration conditions of the four experiments in Qi et al. (2021)[1]. Figure 1 illustrates the path configurations and object arrays used in the four experiments of Qi et al. (2021). The experimental task was conducted in an immersive virtual environment. Participants in Qi et al. (2021) learned the locations of three objects (i.e., A, B, and C in Figure 1) while standing at the origin O (i.e., the home location). O overlapped with either B or C across experiments. After learning, the objects disappeared. Participants traveled along the two outbound legs, i.e., OT and TP. At the endpoint of the outbound path (i.e., P), participants reported the three objects' locations (including home location) by pinpointing the locations individually on the floor using a virtual stick in different cue conditions. Relevant to the current study, participants in the path integration condition only had iditothetic cues. There were 28 participants in each of the four experiments (112 participants in total). Each participant completed 8 outbound paths (three responses for each path) in the path integration condition.

As depicted in Figure 1, the length of the outbound path can be 0.9 m or 1.8 m. And the turn angle on the outbound path can be -20º, ±50º, -70º, ±80º, ±100º, 110º, ±130º, or 160º relative to the direction along the first outbound leg OT (reference direction). Clockwise is positive.

---

[1] The primary purpose of Qi et al. (2021) was to investigate how people combine self-motion and landmark cues to find home and non-home goal locations. Qi et al. (2021) did not examine the sources of systematic errors of path integration.

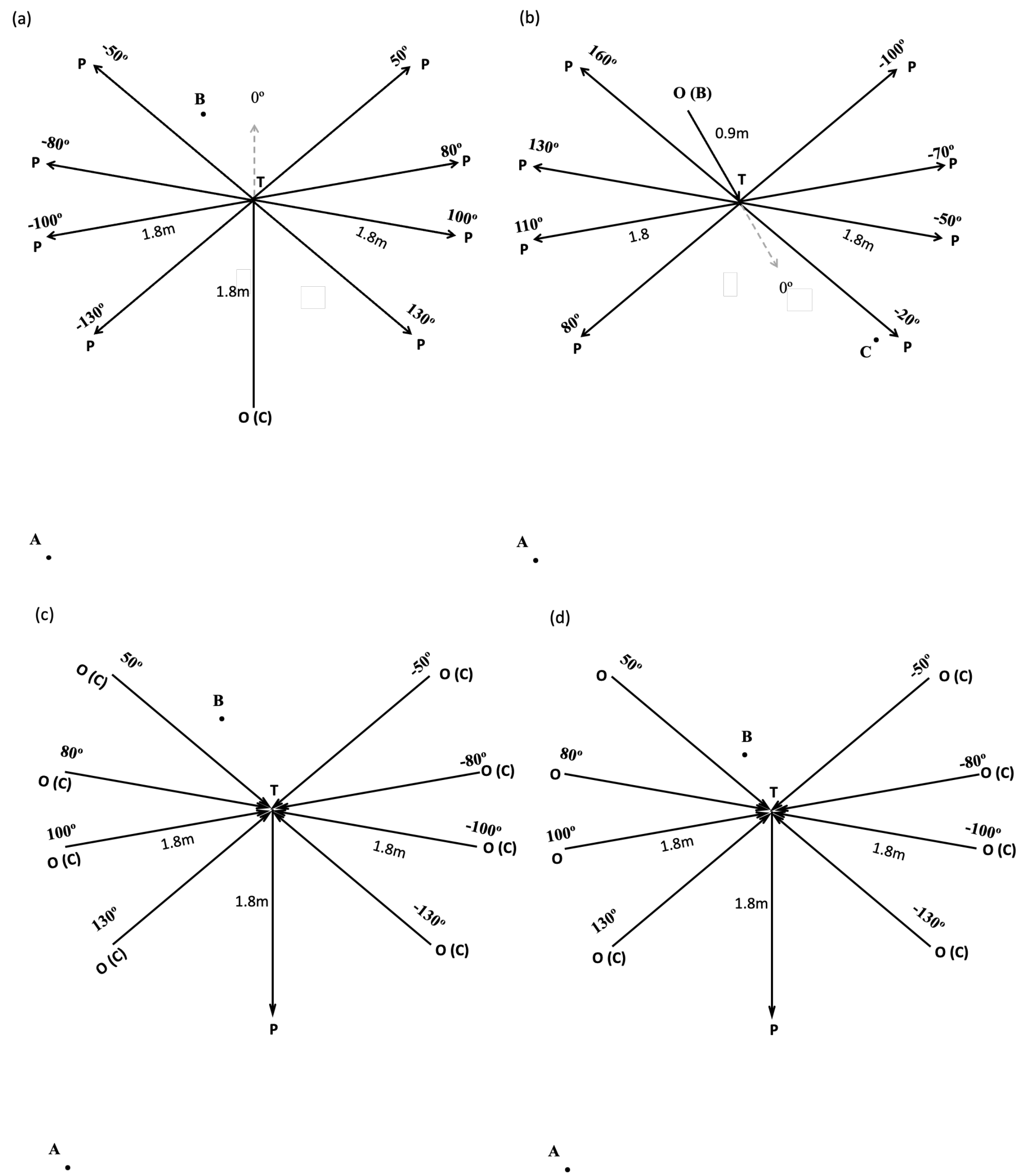


Figure 1. The schematic of outbound path configurations and locations of target objects in four experiments (a, b, c, and d corresponding to experiments 1, 2, 3, and 4 respectively) of Qi et al. (2021). O is the learning location and A, B, and C are the three target locations. An outbound path is comprised of origin O, turning point T, and end point P. The values of turn angles (positive if participants turned right from the direction of OT) and leg lengths are superimposed on each outbound path.

2.2 The compression pattern of the response inbound path length and turn angle

The response measures include the inbound path length and the inbound turn angle for each target location (A, B, and C in Figure 1). Figure 2 depicts examples of the response measures for a target location overlapping with the origin (home target, O) and for a non-home target (A).

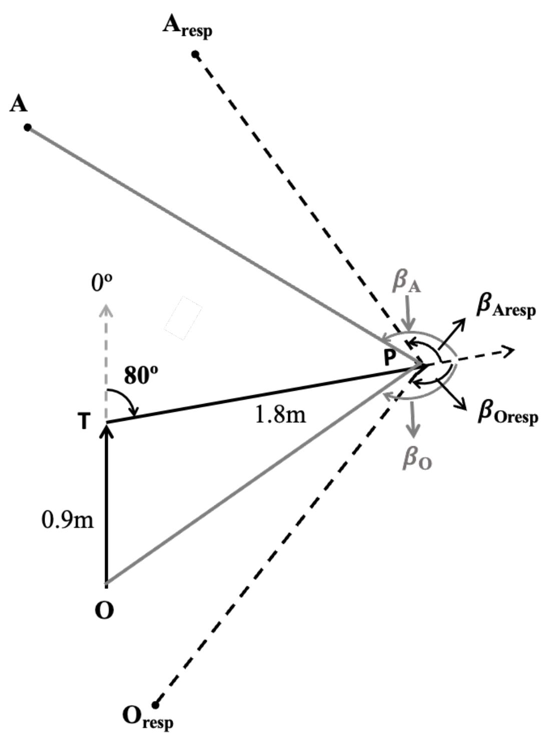


Figure 2. Illustrating the response measures of the current study. O and A are the correct locations of two targets whereas O_{resp} and A_{resp} are the response locations of two targets (O is the home, A is a non-home target). β_O and β_A are the correct inbound turn angles for the targets O and A. β_{Oresp} and β_{Aresp} are the response inbound turn angles for the targets O and A.

The response inbound path length (e.g., PO_{resp}) is the length between the end of the outbound path (i.e., P) and the response target location that the participant pinpointed using a virtual stick (e.g., O_{resp}). The response inbound turn angle (e.g., β_{Oresp}) is the angular difference between the participant's heading at P (i.e., the direction of TP) and the direction from P to the response target location (e.g., O_{resp}). The correct inbound path length (e.g., PO) is the length between the end of the outbound path (i.e., P) and the correct target location (e.g., O). The correct inbound turn angle (e.g., β_O) is the angular difference between the participant's heading at P (i.e., the direction of TP) and the direction from P to the correct target location (e.g., O). In the rest of this paper, we will only use O to represent all target locations regardless of whether it is the home location or non-home location.

Figure 3A plots the response inbound path length (including all three target objects for each outbound path) as a function of correct inbound path length, yielding a linear regression line (the yellow line with markers in Figure 3A) with a slope less than 1 and a positive intercept ($y = 0.633x + 2.201$, $r = .243$). That is, participants tended to overshoot the small distances that they were supposed to produce and reversely, tended to undershoot the large distances. Figure 3B plots the response inbound turn angle as a function of the correct inbound turn angle, yielding a linear regression line with a slope less than 1 and a positive intercept ($y = 0.864x + 28.257$, $r = .632$). That is, participants overturned small angles and underturned large angles. Overall, consistent with previous research (Klatzky et al., 1990; Loomis et al., 1993), the current study confirmed a compression pattern relative to correct values of the inbound responses in triangle completion. Note that the regression line did not cross with the diagonal line (y = x) at the mean of x, referred to as *bias to the mean*, for either length (mean = 2.5m) or angle (mean = 129º). Instead, participants overestimated all correct lengths and angles (referred to as *bias to the upper* *extreme*). Findings of *bias to the extremes* rather than *bias to the mean* were reported in previous studies (e.g., Chrastil & Warren, 2020, Figure 7A for length; Harootonian et al., 2020, for angle and length; Klatzky et al., 1999, Figure 3 for angle; also see Stevens & Greenbaum, 1966 for a variety of different stimuli). The results of *biases to the extreme* could occur because participants might use the prior distribution of the encoding values and response values from their experiences prior to the experiment (Klatzky et al., 1999) as well as from their experiences in the prior trials (Harootonian et al., 2020; Petzschner & Glasauer, 2011). Specifically, participants in the current study might have the overall bias to point to their back (categorical information about the prior, Huttenlocher et al., 1991) because 80% of the correct angles (2156/2688) were larger than 90º (see Figure S1). In addition, Mou and Zhang (2014) suggested that participants might overall overestimate the inbound lengths using a virtual stick for pointing responses because the length of the virtual stick might be underestimated in virtual environments, which might partially explain the *bias to the upper extreme* for length.

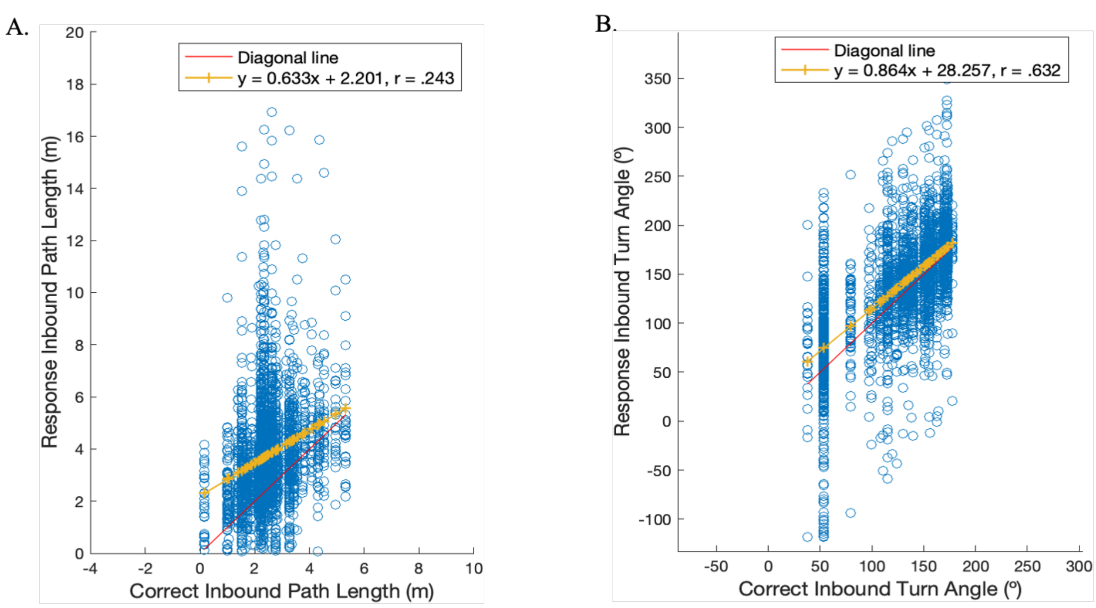


Figure 3. (A) The response inbound length as a function of the correct inbound length. (B) The response inbound turn angle as a function of the correct inbound turn angle. The diagonal lines in red (y = x) indicate the perfect inbound response. The yellow lines indicate the regression lines. Each dot indicates one individual pair of predicted and response values from all three targets and all 896 outbound paths (2688 dots in total).

2.3 Specifications of individual models

To examine the sources of the compression patterns of inbound responses relative to the correct values, we formulated three theoretically plausible models (i.e., the encoding-error model, the execution-error model, and the bi-component model). In addition, we also included a baseline model that assumes no systematic bias and used the correct values as the predicted values for the inbound responses.

2.3.1 The encoding-error model

The encoding functions of the outbound path length and the outbound turn angle comprise a set of 4 parameters, 2 for each function. $\theta_{L\_s}^{enc}$, $\theta_{L\_i}^{enc}$ are the slope and the intercept of the linear function for encoding the outbound path length whereas $\theta_{A\_s}^{enc}$, $\theta_{A\_i}^{enc}$ are the slope and the intercept of the linear function for encoding the outbound turn angle. Same as the original encoding-error model, $\theta_{L\_s}^{enc}$, $\theta_{L\_i}^{enc}$ are used for both the first and second legs of the outbound path. Thus, the encoded values of leg length L_e and turn angle α_e can be represented with these parameters,

$$L_e = \theta_{L\_s}^{enc} \times L + \theta_{L\_i}^{enc}, \tag{1}$$

$$\alpha_e = \theta_{A\_s}^{enc} \times \alpha + \theta_{A\_i}^{enc}, \tag{2}$$

where L and α are the correct length and turn angle of the outbound path, respectively (see values in Figure 1).

As depicted in Figure 4A, hypothetical participants encode outbound segment L1, L2, and turn angle α as L_{1e}, L_{2e}, and α_e. According to Formulas 1 and 2, $L_{1e} = \theta_{L\_s}^{enc} \times L1 + \theta_{L\_i}^{enc}$, $L_{2e} = \theta_{L\_s}^{enc} \times L2 + \theta_{L\_i}^{enc}$, $\alpha_e = \theta_{A\_s}^{enc} \times \alpha + \theta_{A\_i}^{enc}$.

In a Cartesian coordinate system, by means of theorems of trigonometry, the encoded outbound path can be represented in terms of vectors, $\overrightarrow{OT_e} = \frac{L_{1e}}{L1} \times \overrightarrow{OT}$, and $\overrightarrow{T_eP_e} = L_{2e} \times \frac{\overrightarrow{T_eP_e}}{||\overrightarrow{T_eP_e}||}$.

Where the $\left\|\overrightarrow{T_eP_e}\right\|$ is the length of the vector of $\overrightarrow{T_eP_e}$. $\frac{\overrightarrow{T_eP_e}}{||\overrightarrow{T_eP_e}||}$ equals to the unit vector (a vector with the length of 1) with the direction of the vector $\overrightarrow{OT}$ being rotated by the angle of α_e.

Accordingly, the participants consider themselves standing at P_e and facing the direction of h_e, same as the direction of $\overrightarrow{T_eP_e}$. To pinpoint the target location, they intend to produce the desired inbound vector $\overrightarrow{P_eO}$, which consists of the desired inbound turn angle β_e and the desired inbound path length L_{3e}:

$$\overrightarrow{P_eO} = -(\overrightarrow{OT_e} + \overrightarrow{T_eP_e}), \quad (3)$$

$$\beta_e = \text{dir}(\overrightarrow{P_eO}) - (\text{dir}(\overrightarrow{OT}) + \alpha_e), \quad (4)$$

$$L_{3e} = \left\|\overrightarrow{P_eO}\right\|. \quad (5)$$

Where the dir $(\overrightarrow{P_eO})$ is the direction of $\overrightarrow{P_eO}$ and dir $(\overrightarrow{OT})$ is the direction of $\overrightarrow{OT}$. The direction of a vector is specified by the angular distance from a fixed reference direction in the virtual environment (e.g., the UP direction in Figure 1) to the vector. Where the $\left\|\overrightarrow{P_eO}\right\|$ is the length of the vector of $\overrightarrow{P_eO}$.

As there is no systematic bias in executing the inbound path based on the assumptions of the encoding-error model, the participants are able to implement the desired inbound path length and turn angle without bias (e.g., $L_r = L_{3e}$, $\beta_r = \beta_e$ in Figure 4A) while standing at P and facing the direction of h actually. Thus, the predicted response vector $\overrightarrow{PO_{pred}}$ can be given by

$$\overrightarrow{PO_{pred}} = L_{3e} \times \frac{\overrightarrow{PO_{pred}}}{||\overrightarrow{PO_{pred}}||}, \quad (6)$$

where $\frac{\overrightarrow{PO_{pred}}}{||\overrightarrow{PO_{pred}}||}$ equals to the unit vector with the direction of the vector $\overrightarrow{OT}$ being rotated by the angle of $(\alpha + \beta_e)$.

We then get the predicted response location O_{pred}.

$$O_{pred} = \mathrm{P} + \overrightarrow{PO_{pred}}. \quad (7)$$

Where O_{pred} and P represent the coordinates in the Cartesian coordinate system used in Qi et al. (2021), where the direction of UP in Figure 1 is y positive and the direction of RIGHT in Figure 1 is x positive.

Thus, following Formula 1-7, the coordinates of the predicted response location O_{pred} can be expressed in terms of parameters $\theta_{L\_s}^{enc}$, $\theta_{L\_i}^{enc}$, $\theta_{A\_s}^{enc}$, and $\theta_{A\_i}^{enc}$, and several constants (e.g., L1, L2, and α) for each path.

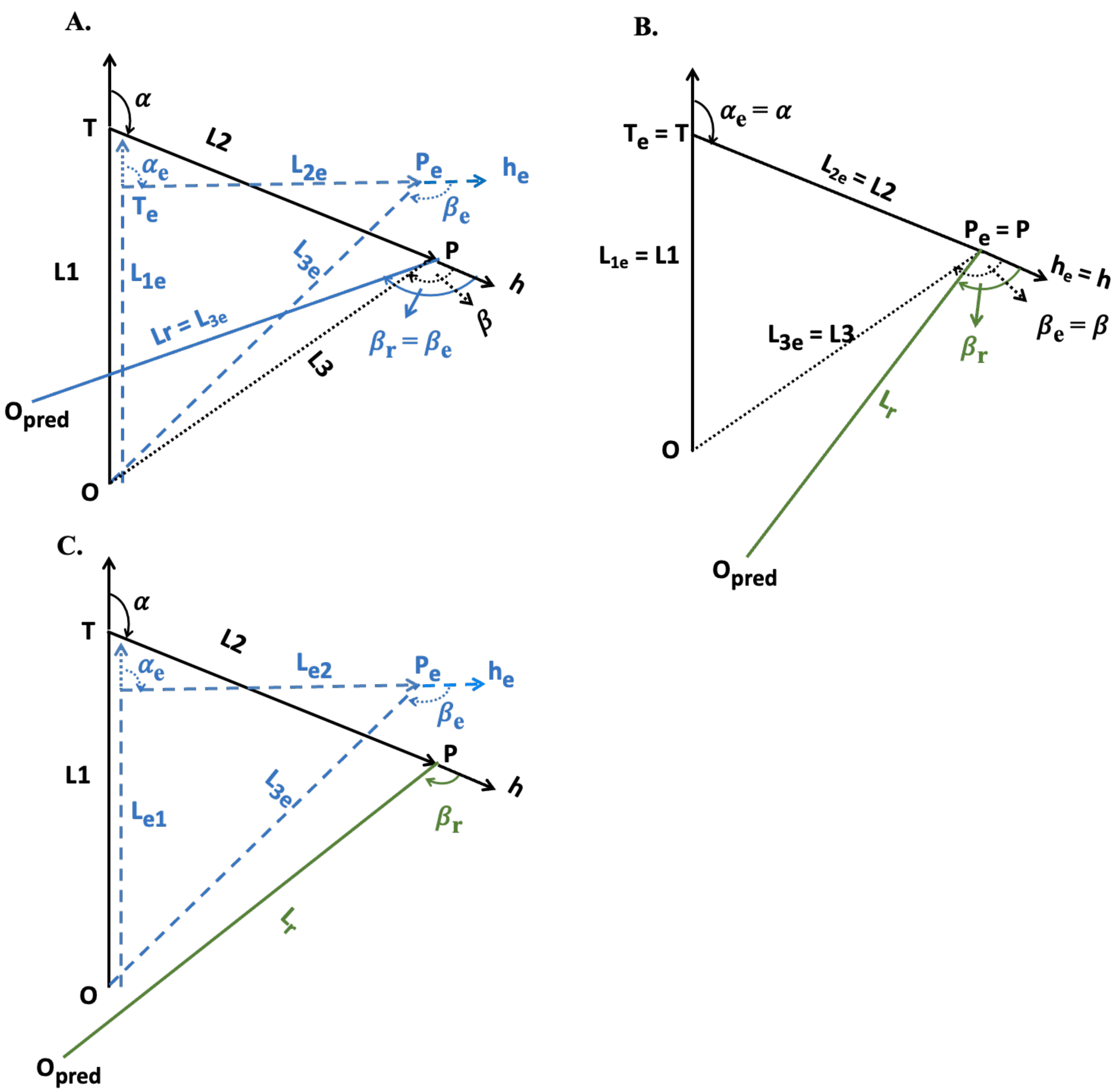


Figure 4. Illustration of predictions of different models. In each panel, the outbound path of a participant, O-T-P (solid black), consists of lengths L1 and L2 and turn angle α. H is the participant's heading at the end of the outbound path. The prediction of the participants' inbound path, PO$_{\text{pred}}$ (solid blue indicating inbound responses without systematic errors or solid green indicating inbound responses with systematic execution errors), consists of length L_r and inbound turn angle β_r. O$_{\text{pred}}$ is the predicted location of O. (A) the encoding-error model. The encoded outbound path, O-Te-Pe (blue dotted), consists of lengths L1e and L2e and turn angle

αe, which are determined by the encoding functions. h_e is the encoded heading at the end of the outbound path. The desired inbound responses are free of execution errors (i.e., L_r = L_{3e} and β_r = β_e). (B) the execution-error model. The outbound path is free of encoding errors (α_e = α and P_e = P). The inbound responses (L_r and β_r) are solely determined by the execution functions. (C) the bi-component model. The inbound responses (L_r and β_r) are determined by the systematic errors in encoding (blue dots) according to the encoding functions and in execution (green solid) according to execution functions.

2.3.2 The execution-error model

The execution-error model assumes that the process of encoding is independent of the systematic bias and the navigators estimate their self-localization (i.e., T_e = T and P_e = P in Figure 4B) accurately.

The execution functions for inbound path length and angle have 2 parameters, respectively. While $\theta_{L\_s}^{exe}$ and $\theta_{L\_i}^{exe}$ are the slope and intercept for the inbound path length, $\theta_{A\_s}^{exe}$ and $\theta_{A\_i}^{exe}$ are the slope and intercept for inbound turn angle.

The executed values of inbound length L_r and turn angle β_r (see Figure 4B) can be represented as:

$$L_r = \theta_{L\_s}^{exe} \times L_{3e} + \theta_{L\_i}^{exe}, \tag{8}$$

$$\beta_r = \theta_{A\_s}^{exe} \times \beta_e + \theta_{A\_i}^{exe}, \tag{9}$$

where L_{3e} and β_e equal to the correct length L3 and turn angle β for the inbound path, respectively, because there is no systematic error in encoding the outbound path.

Therefore, the predicted response vector $\overrightarrow{PO_{pred}}$ can be calculated according to Formula 10:

$$\overrightarrow{PO_{pred}} = L_r \times \frac{\overrightarrow{PO_{pred}}}{||\overrightarrow{PO_{pred}}||}. \tag{10}$$

Where $\frac{\overrightarrow{PO_{pred}}}{||\overrightarrow{PO_{pred}}||}$ equals to the unit vector with the direction of the vector $\overrightarrow{OT}$ being rotated by the angle of $(\alpha + \beta_r)$.

As a result, the predicted location O_{pred} can be calculated by Formula 7.

2.3.3 The bi-component model

Since the bi-component model presumes that both the encoding and execution processes contribute to systematic errors, it incorporates the previously described encoding functions for the outbound path and execution functions for the inbound path (see Figure 4C).

Specifically, Formula 1 through 5 still holds in encoding the outbound path and estimating the desired inbound response, i.e., L_{3e} and β_e, for the current model. Formula 8-10 still holds when executing the desired inbound response through the execution functions. As a result, Formula 7 can be used to calculate the model's predicted response location O_{pred}.

2.3.4 The baseline model

The baseline model presumes no systematic bias in both encoding and execution stages, i.e., the slopes are one and the intercepts are zero for all the encoding functions and the execution functions. Thus, the baseline model directly used the correct values of the target locations to predict participants' response locations (O_{pred} = O).

Note that Harootonian et al. (2020) showed the influence of the immediately preceding trial. Participants tended to bias the encoded distance of the current trial towards the encoded distance of the previous trial (e.g., a larger distance in the previous trial would lead to overestimation of a short distance in the current trial), which indicates that the Bayesian prior of the true value assimilates the information of the immediately preceding trial. According to the three models interested in the current study (encoding-error model, execution-error model, and bi-component model), a Bayesian prior could be considered in encoding the outbound path, executing the inbound path, or in both, predicting history effects in different processes. To simply the model comparison, we did not add parameters of the history effect to the models in the current study.

2.4 Cross-validation for models without considering participant variable

We conducted cross-validation for models without considering participants' differences in their compression patterns in either encoding or response functions. Therefore, one value of each parameter (e.g., eight free parameters, $\theta_{L\_s}^{enc}$, $\theta_{L\_i}^{enc}$, $\theta_{A\_s}^{enc}$, $\theta_{A\_i}^{enc}$, $\theta_{L\_s}^{exe}$, $\theta_{L\_i}^{exe}$, $\theta_{A\_s}^{exe}$, and $\theta_{A\_i}^{exe}$ for the bi-component model) was estimated for all participants.

For each model, the technique of 5 times of 2-fold (5 × 2) cross-validation (Alpaydm, 1999; Dietterich, 1998) was employed for the computational modeling of the response locations. To be specific, the original dataset (all 896 outbound paths, 8 paths × 4 experiments × 28 participants for each experiment) was partitioned randomly into two equal subsamples, S1 and S2, with 448 outbound paths each. One subsample (e.g., S1) was assigned to the model training to estimate the model parameters, and the other (e.g., S2) was used for the model validation. Then, the two subsamples were swapped, that is, S2 was used for model training and S1 was the subsample to test the model performance. The above random subsampling and cross-validation operations were repeated 5 rounds. Each half of the dataset was applied to both model fitting and validation in each round. Afterward, model performance in model validation can be averaged across the ten folds (5 × 2 folds) to obtain a more robust estimation of the model performance by reducing the impact of sampling (partitioning) errors.

The process of modeling was carried out using two different algorithms. One only used the data of the home response location for every outbound path, as in the previous typical triangle-completion studies, whereas the other used all three response locations for every outbound path. As we speculated above, only using the response to the home for every outbound path, cross-validation modeling may not distinguish the three models (single-component models and the bi-component model). In contrast, using the responses to three locations for every outbound path, cross-validation modeling may distinguish the three interested models.

2.4.1 Model fitting

The functions of each model were determined (i.e., the parameters of θs were estimated) by making the models' predictions (O_{pred}) as closely as possible to the participants' responses (O_{resp}). The discrepancy was measured by the mean squared error (MSE) between the predicted and response locations across all outbound paths and all targets (3 for the algorithms using multiple response locations and 1 for the algorithms using home response locations only) in training subsamples (the data used for model fitting):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}[\left(O_{xi}^{pred} - O_{xi}^{resp}\right)^2 + \left(O_{yi}^{pred} - O_{yi}^{resp}\right)^2], \quad (11)$$

where the $(O_{xi}^{pred}, O_{yi}^{pred})$ is the predicted location based on the model, $(O_{xi}^{resp}, O_{yi}^{resp})$ is the response location, and n is the number of data points.

Then using Matlab's fminsearch function, we found the value of parameters that minimize the MSE for each model. The fminsearch function can detect the minimal value of an objective function (e.g., MSE) by means of various optimization algorithms. To boost the possibility of locating a global minimum rather than a local one for the objective function, the search ran 500 iterations and each time started with random initial values of parameters. After

500 iterations, the fitting procedure located the minimum of MSE at an optimal solver, and this solver was the set of best-fitting parameters.

Table1 summarizes the averaged ten-fold results of fitting different models to response data, including parameters and fitting performance, using two distinct algorithms (see Supplementary Materials and Table S1 for results of individual folds). These parameters would be held for the subsequent model validation.

For brevity, the encoding-error model is referred to as Model 1, the execution-error model as Model 2, the bi-component model as Model 3, and the baseline model as Model 0 (abbreviated as M1, M2, M3, and M0, respectively in the following sections).

The fitting performance of a specific model *M* is evaluated by the squared root of the MSE (RMSE), the percentage of the variance of the baseline model explained by the individual model (Partial $R^2 = 1- \frac{MSE\ of\ M}{MSE\ of\ M0}$), and the maximum log-likelihood (MaxLogL).

To calculate the maximum log-likelihood, we assumed that the deviations of the predicted locations from the response locations ($O_{xi}^{pred} - O_{xi}^{resp}, O_{yi}^{pred} - O_{yi}^{resp}$), referred to as the locational residuals, were from a bivariate normal distribution with zero means ($\mu = (0,0)$) and undetermined covariance matrix (Σ). The maximum log-likelihood of the locational residuals were calculated by Formula 12 (Jordan, 2003; Taboga, 2021):

$$\text{MaxLogL} = \log\left[\left(\frac{1}{\sqrt{2\pi}}\right)^{cn} \times e^{-\frac{cn}{2}} \times \left|\hat{\Sigma}\right|^{-\frac{n}{2}}\right]. \quad (12)$$

Where c is the dimension of the data (c = 2 for the locational residuals), and n refers to the number of the data points (n = 498 × 3 for the algorithms of using multiple locations and n = 498 for the algorithms of using the home response locations only). $\hat{\Sigma}$ is

$$\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}\left(O_{xi}^{pred}-O_{xi}^{resp}\right)^2 & \frac{1}{n}\sum_{i=1}^{n}\left(O_{xi}^{pred}-O_{xi}^{resp}\right)\left(O_{yi}^{pred}-O_{yi}^{resp}\right) \\ \frac{1}{n}\sum_{i=1}^{n}\left(O_{xi}^{pred}-O_{xi}^{resp}\right)\left(O_{yi}^{pred}-O_{yi}^{resp}\right) & \frac{1}{n}\sum_{i=1}^{n}\left(O_{yi}^{pred}-O_{yi}^{resp}\right)^2 \end{bmatrix}$$ from each individual models. $|\hat{\Sigma}|$ is the determinant of the matrix.

Table 1

Model fitting performance using multiple locations (upper) or only home response locations (lower). Parameters are estimated slopes and intercepts of encoding functions ($\theta_{L\_s}^{enc}$ and $\theta_{L\_i}^{enc}$ for length, $\theta_{A\_s}^{enc}$ and $\theta_{A\_i}^{enc}$ for angle) and execution functions ($\theta_{L\_s}^{exe}$ and $\theta_{L\_i}^{exe}$ for length, $\theta_{A\_s}^{exe}$ and $\theta_{A\_i}^{exe}$ for angle) for all four models in the model fitting. The RMSE, maximum log-likelihood, and partial r-squared are goodness-of-fit measures. M0 = the baseline model, M1=the encoding-error model, M2 = the execution-error model, M3 = the bi-component model.

	Multiple response locations										
	Parameters								5×2 Fitting		
Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	3.178	-5961.9	0
M1	1.04	0.48	0.79	18.38	1	0	1	0	3.076	-5882.2	0.063
M2	1	0	1	0	0.70	1.29	0.78	41.11	3.054	-5865.2	0.077
M3	0.82	0.78	0.84	20.42	0.69	1.10	0.82	34.21	3.017	-5831.5	0.099
	Home response locations only										
	Parameters								5×2 Fitting		
Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	2.805	-1867.1	0
M1	0.68	0.67	0.45	23.43	1	0	1	0	2.620	-1815.5	0.128
M2	1	0	1	0	0.42	2.10	0.47	84.21	2.625	-1816.8	0.124
M3	2.53	3.94	0.48	26.20	0.73	0.11	1.18	12.55	2.618	-1815.0	0.129

Table 1 shows that the bi-component model (M3) is the best model according to the three goodness-of-fit measures numerically when all three response locations were included in the model fitting. In contrast, although the three interested models (M1-M3) are better than the base model (M0), they could not distinguish from each other when only the home response locations were included in the model fitting. However, the superiority of the bi-component model (M3) using all three response locations might be attributed to the fact that the bi-component model (M3) has more free parameters than the encoding-error model and the execution-error model (M1 and M2). This issue could be addressed by some model selection criteria (e.g., AIC, Akaike, 1973 or BIC, Schwarz, 1978) that penalize free parameters to be estimated. This issue could also be addressed by cross-validation which applied the estimated parameters to independent data (i.e., test subsamples) so that there is no free parameter in any models. The current study used the second approach. We still conducted AIC and BIC analyses for the training subsamples as some readers might be interested (see Supplementary Materials and Table S3).

2.4.2 Model validation

In each round of cross-validation (five rounds in total), after fitting models to each training subsample (S1 or S2), we evaluated the generalizability of models using the corresponding test subsample (S2 or S1). Table 2 shows the averaged validation performance over ten test subsamples after performing the cross-validation five times for all four models (see Supplementary Materials and Table S2 for results of individual folds).

More specifically, for each model, the estimated parameters derived from each training subsample were applied to predict the response locations for the corresponding test subsample that were not involved in estimating the parameters. The residuals between the predicted and response locations were used to calculate the RMSE, maximum log-likelihood, and partial r-squared.

Table 2

Model validation performance using multiple locations (upper) or only home response locations (lower). Parameters are the same as in Table 1 from model fitting. The RMSE, maximum log-likelihood, and partial r-squared are generalizability measures, which were calculated by applying the parameters to the test subsamples.

	Multiple response locations										
	Parameters								5×2 Validation		
Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	3.178	-5961.9	0
M1	1.04	0.48	0.79	18.38	1	0	1	0	3.085	-5889.9	0.058
M2	1	0	1	0	0.70	1.29	0.78	41.11	3.060	-5868.9	0.073
M3	0.82	0.78	0.84	20.42	0.69	1.10	0.82	34.21	3.031	-5843.6	0.090
	Home response locations only										
	Parameters								5×2 Validation		
Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	2.805	-1867.1	0
M1	0.68	0.67	0.45	23.43	1	0	1	0	2.632	-1819.2	0.120
M2	1	0	1	0	0.42	2.10	0.47	84.21	2.633	-1819.1	0.119
M3	2.53	3.94	0.48	26.20	0.73	0.11	1.18	12.55	2.634	-1819.8	0.118

Table 2 indicates that the bi-component model (M3) is the best model according to the three generalizability measures when all three response locations were included in the model evaluation. In contrast, although the encoding-error model, execution-error model, and the bi-component model (M1, M2, and M3) are better than the baseline model (M0), they could not distinguish from each other when only the home response locations were included in the model evaluation.

These conclusions were quantified by the maximum likelihood ratios (LRs) analysis. Because all models have the same number of free parameters for the test subsamples, LRs can be directly calculated from the MaxLogLs without adjustment due to difference in parameter numbers. Table 3 summarizes the results.

Table 3

Maximum likelihood ratio (LR) between models (row model over column model) in model validation using multiple locations (left) or only home response locations (right).

	Multiple response locations				Home response locations only			
LR	M0	M1	M2	M3	M0	M1	M2	M3
M1	$1.86 \times 10^{31**}$				$6.70 \times 10^{20**}$			
M2	$2.31 \times 10^{40**}$	$1.25 \times 10^{9**}$			$7.55 \times 10^{20**}$	$1.13^{\text{—}}$		
M3	$2.37 \times 10^{51**}$	$1.28 \times 10^{20**}$	$1.02 \times 10^{11**}$		$3.66 \times 10^{20**}$	$0.55^{\text{—}}$	$0.49^{\text{—}}$	

Note: * indicates clear evidence, i.e., LR > 3 or LR <1/3, and ** indicates strong evidence, i.e., LR > 10 or LR <1/10. — indicates no evidence (Glover & Dixon, 2004).

The results of the maximum likelihood ratio shown in Table 3 demonstrate that there is strong evidence in favor of the bi-component model (M3) over the encoding-error model (M1) and the execution-error model (M2) when the cross-validation included multiple response locations, whereas there was no clear evidence favoring any models when the cross-validation included only home response locations.

Furthermore, we adopted Alpaydin's 5×2cv combined F test to examine the differences in models' performance (Alpaydm, 1999, see also Raschka, 2018). To compare the results of two competing models, the difference in the value of RMSE (dRMSE) between them was calculated, generating 5 × 2 difference matrices (RMSEs of ten-folds in validation of each model are listed in Table S2). d_i^j was used to denote the dRMSE value on the jth (j = 1, 2) fold of the ith (i = 1, …, 5) round in a difference matrix and d_i^{avg} denotes the averaged RMSE difference in the ith round, $d_i^{avg} = (d_i^1 + d_i^2) / 2$.

Then the estimated variance of the difference for the ith round is given by

$$s_i^2 = \left(d_i^1 - d_i^{avg}\right)^2 + \left(d_i^2 - d_i^{avg}\right)^2. \tag{13}$$

The F statistic is calculated as:

$$f = \frac{\sum_{i=1}^{5}\sum_{j=1}^{2}\left(d_i^j\right)^2}{2\sum_{i=1}^{5} s_i^2}, \tag{14}$$

which approximately follows an F distribution with (10, 5) degrees of freedom.

Table 4 summarizes the mean dRMSE of all pairs of the models and the corresponding significance of Alpaydin's F-test. Consistent with the results indicated by Table 3, when three locations' data were included (left panel), the results show that the bi-component model ($RMSE_{M3} = 3.031$) significantly outperforms the encoding-error model ($RMSE_{M1} = 3.085$, $p < .001$) and the execution-error model ($RMSE_{M2} = 3.060$, $p = .02$) in predicting the actual responses. The execution-error model presents significantly better performance than the encoding-error model ($p < .01$). All the three interested models have substantially better predictive performance than the baseline model ($RMSE_{M0} = 3.178$, all p values $< .001$).

By contrast, when only the data of home response locations were used in the cross-validation (right panel), there was no significant difference in RMSE among M1, M2, and M3 although RMSEs in these three models, approximately 2.63, were significantly smaller than that of the baseline model (M0) ($RMSE_{M0} = 2.805$, all p values $< .01$).

Table 4

Alpaydin's F-test examining the differences in RMSE (dRMSE) between models (the row model minus the column model) when using multiple locations (left) or only home response locations (right).

Multiple response locations					Home response locations only				
dRMSE	M0	M1	M2	M3	dRMSE	M0	M1	M2	M3
M1	-.093**				M1	-.174***			
M2	-.118***	-.025**			M2	-.172***	.001¯		
M3	-.147**	-.056***	-.029*		M3	-.171***	.002¯	.001¯	

Note: Asterisks denote significant dRMSE (***p < .001; **p < .01; *p < .05) and a dash (¯) indicates non-significant dRMSE.

Figure 5 visually presents locational residuals of model validation. We calculated the *mean predicted locations* of each target (three for multiple response locations or one for home only) in each outbound path (32 in total) across the ten folds of the test subsamples based on different models. We also calculated the *mean response location* of the target across participants who replaced this target. The *locational residual* of one target for one model is the difference between the *mean predicted location* based on this model and the *mean response location* of the target across participants (mean predicted location – mean response location).

Figure 5A, employing multiple response locations, reveals clear differences in predictive performance among all these models. In particular, the bi-component model achieves more centric dots and a smaller area of 95% density contours of the residual distributions compared with other competing models, indicating that it is capable to predict the actual responses of the participants more accurately. By contrast, Figure 5B, employing only the home response locations, shows that apart from the baseline model, the performance of the other three models is not distinguishable (the dots of various colors are mixed up and the ellipses overlap).

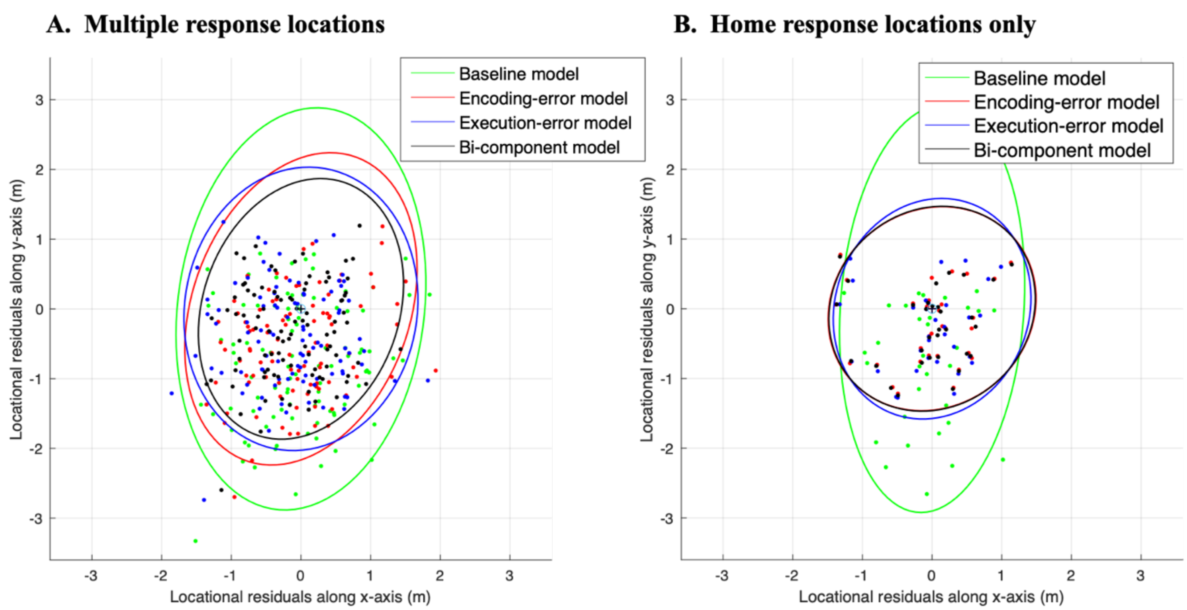


Figure 5. Visualizing the differences (locational residuals) between mean response locations and mean predicted locations from different models using (A) multiple response locations or (B) only home response locations. The open circle with a cross at (0, 0) indicates the response location, the coordinate of which varied in real experiments but is set to (0, 0) as a reference. Individual dots represent coordinates of the locational residuals for all targets (96 targets in A and 32 in B). Ellipses indicate the 95% density contours of the bivariate normal distributions with zero means ($\mu = (0,0)$) and covariance matrix (Σ) of the locational residuals according to the baseline model (green), encoding-error model (red), execution-error model (blue), and bi-component model (black), respectively.

2.4.3 Model recovery

The results of 5 × 2 cross-validation indicated that the bi-component model was the best model to predict the response locations. Furthermore, although the algorithm of using all three objects can dissociate the bi-component model from the encoding-error and execution-error models, the algorithm of using only home response locations cannot. Because both these conclusions are dependent on the cross-validation methods used in the current project, these conclusions will be significantly strengthened if the cross-validation methods used in the current project can be shown to distinguish the true model from other models using the *simulated* response locations produced by each of the three models (the encoding-error, execution-error, and bi-component models).

For each model (i.e., the true model), we generated simulated response locations for all ten subsamples (5 × 2 folds). Using the corresponding parameters derived from model fitting using multiple objects (e.g., the values for M1, M2, and M3 in the upper table of Table 1), we calculated the predicted locations for all three targets for each of the 448 outbound paths in each subsample. Using the corresponding RMSE in the upper table of Table 1, we generated random noises for both dimensions (x and y) of all predicted locations from a normal distribution ($\mu = 0$, $\sigma = \frac{\mathrm{RMSE}}{\sqrt{2}}$). Each simulated response location is then the sum of the predicted location and the noise. We applied both algorithms of 5 × 2 cross-validation (using multiple response locations or using only home response locations) to the simulated response locations and examined whether the generalizability measure (i.e., LR) in the model validation could distinguish the true model from other models. We created 100 sets of simulated response locations and conducted 5 × 2 cross-validation for all of them[2].

The frequency of successfully distinguishing the true model from other models could also indicate the discriminability of the cross-validation methods. For each true model, we calculated the likelihood ratio between any two models for each of the 100 simulations and classified the likelihood ratios into different categories (see details in Supplementary Materials and Figure S2).

[2] Note that it takes about 3.5 hours to finish 5 × 2 cross-validation for each simulation subsample using all three response locations of each outbound path.

Figure 6 presents the confusion matrix in model recovery. The best model was determined only when it had likelihood three times higher than both other two models. The results showed that the algorithm of using multiple response locations can successfully distinguish the true model from other models. Occasionally the algorithm could not find the best model (i.e., no model had likelihood three times than both other two models) (e.g., for true model M2, 18% chance of failure to find the best model). However, in the most time, the algorithm recovered the true model (98% for true model M1, 82% for true model M2, and 100% for true model M3) and never recovered any distracting models. By contrast, the algorithm of using home response locations cannot clearly distinguish the true model from other models. In most cases, the algorithm could not find the best model (with a rate larger than 49%). Consequently, the algorithm could recover the true model at a low rate (23% for true model M1, 50% for true model M2, and 7% for true model M3). The algorithm also at times recovered distracting models.

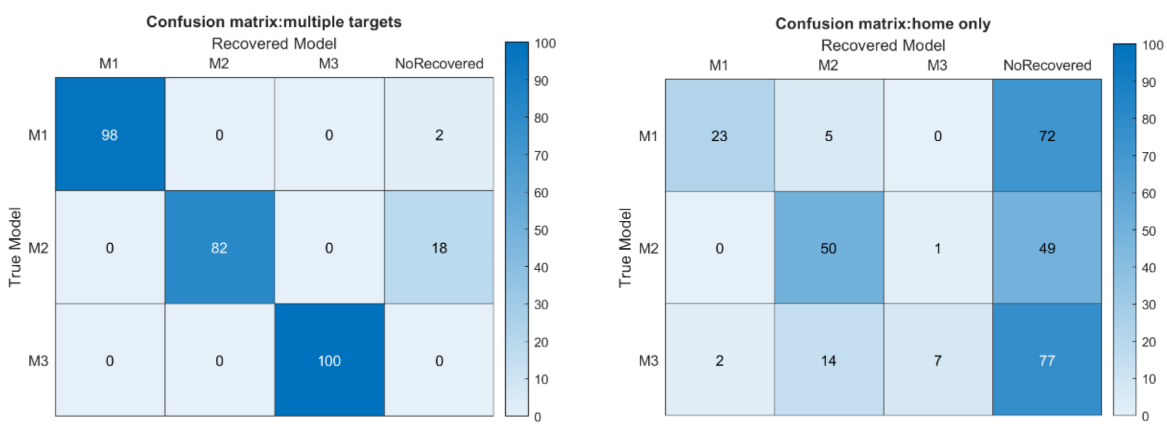


Figure 6. Confusion matrices in model recovery using multiple response locations (left) or home response locations only (right). The number in each cell indicates the frequency of the recovered model being the best model. NoRecoved means that no best model was recovered by the algorithm.

2.4.4 Similarity of parameters' values estimated from real and simulated response locations

The algorithm using multiple response locations estimated 16 parameters (four parameters for M1, four for M2, and eight for M3, see Table 1) based on participants' response locations. Similarly, this algorithm could also estimate 16 parameters based on simulated locations produced by each true model. The similarity between the estimated parameters based on real and simulated response locations should reflect the similarity between real and simulated response locations, thus indicating the closeness between the true model that produced the real response locations and each model. The model closest to the true model should be the best model. The similarity between parameters based on real response locations and simulated response locations from different models were illustrated by Figure 7 (see exact parameters in Table S4. The parameter distance was shortest when the simulated locations were produced by M3 (RMSE = 9.44, 6.8, and 1.5 for M1, M2, and M3 respectively). The parameters based on simulated locations from M3 explained the largest proportion of the total variance of the 16 parameters based on real response locations ($r^2 = 1 - \frac{MSE}{Var}$, r^2 = .46, .72, and .99 for M1, M2, and M3 respectively). The rates of likelihood of M3 over other models were larger than 3.33×10^{10} (logL = -58.64, -53.38, and -29.15 for M1, M2, and M3 respectively). Therefore, the similarity between real and simulated response locations from M3 was largest, indicating M3 was the best model.

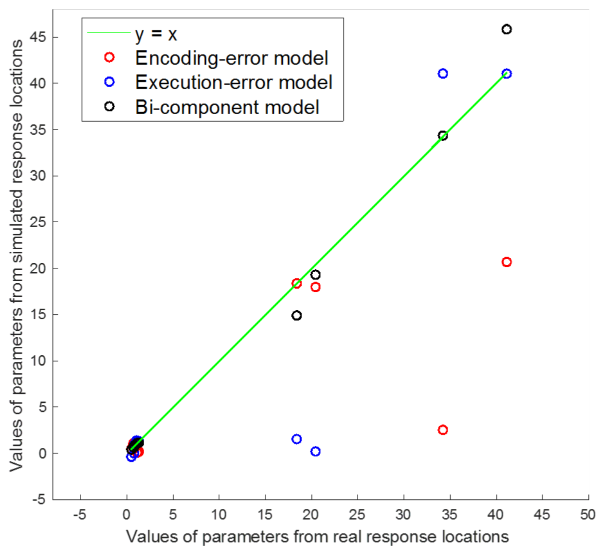


Figure 7. Illustrating the similarity of estimated parameters based on real data and simulated data from different models. The diagonal lines in green (y = x) indicate the ideal outcome that the parameters derived from real data are perfectly recovered from simulated data. Open dots depict the individual pairs of values of parameters based on real and simulated response locations for each model.

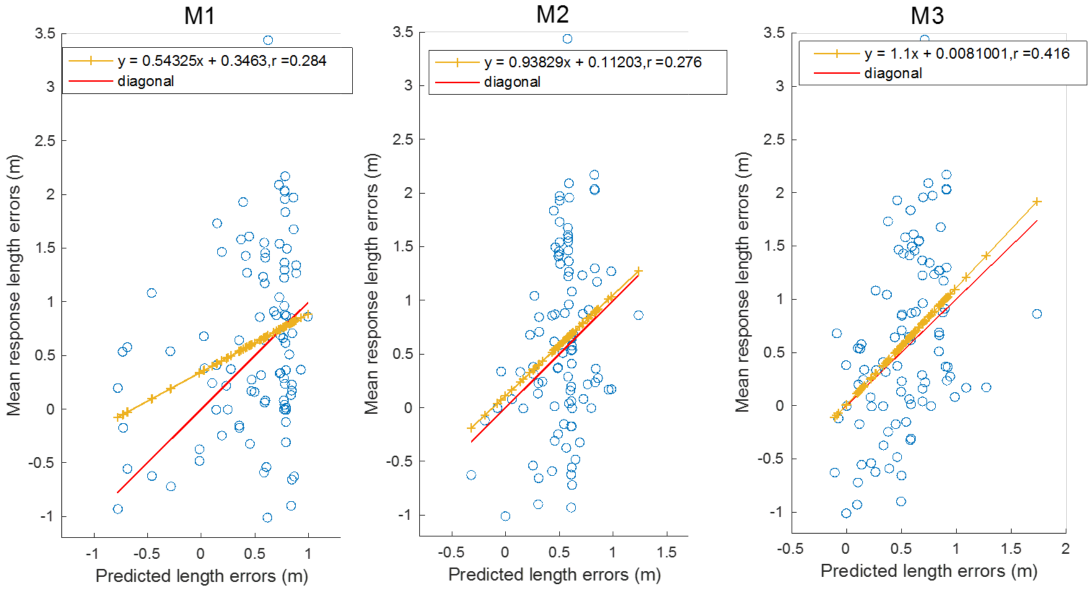
2.4.5 Predictive performance on the response error of participants

In addition, we compared the predictive performance of different models in terms of participants' response error (inbound path length or turn angle), consistent with previous studies (Chrastil & Warren, 2021; Fujita et al., 1993). We conducted the following analyses of the mean predicted locations of targets across the ten-fold test subsamples, which were used in model validation. The predicted inbound path ($\overrightarrow{PO_{Pred}}$) was calculated from the testing position (P) to the predicted location (O_{pred}) based on each model. The predicted error (inbound path length or turn angle) was defined as the difference between the predicted and correct values for each target and each unique outbound path (32 different types of paths, 8 in each of the four experiments).

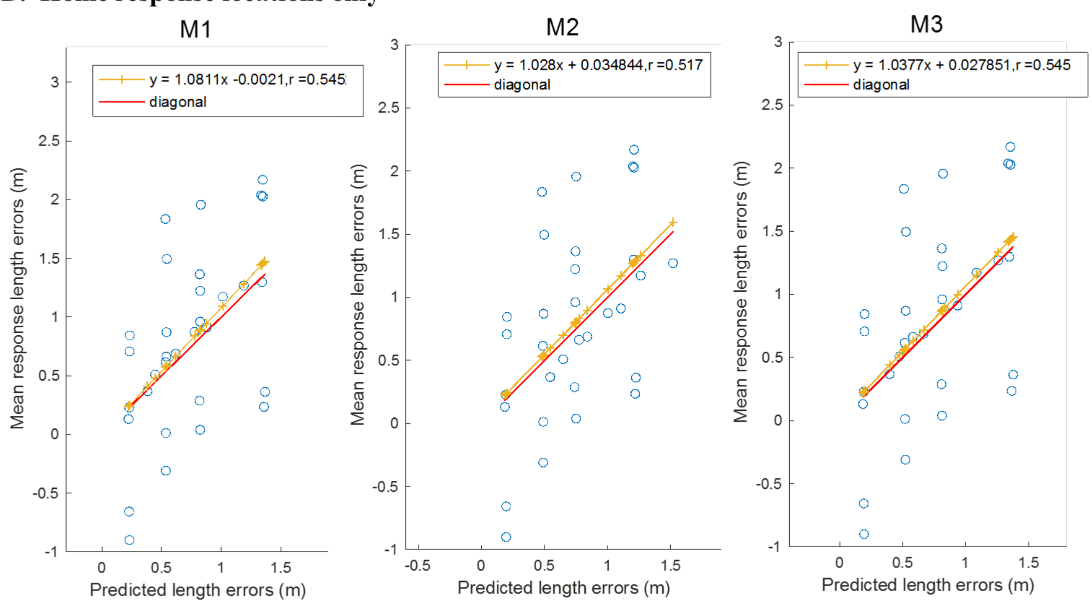
The individual response error (inbound path length or turn angle) was defined as the difference between the response and correct values. The mean response error for each target and each unique outbound path was the average of the individual response errors across participants for the specific target and the specific outbound path.

Figure 8 illustrates the mean predictive performance of different models in terms of inbound length error and angle error. It shows that the bi-component model (M3) had the highest correlation coefficients for both inbound length (see *r*s in Figure 8A) and angle errors (Figure 8C) when the cross-validation included multiple response locations of each outbound path. Nevertheless, the correlation coefficients of the three models were comparable when the cross-validation only included the home response location of each outbound path (see *r*s in Figure 8B and Figure 8D).

A. Multiple response locations



B. Home response locations only



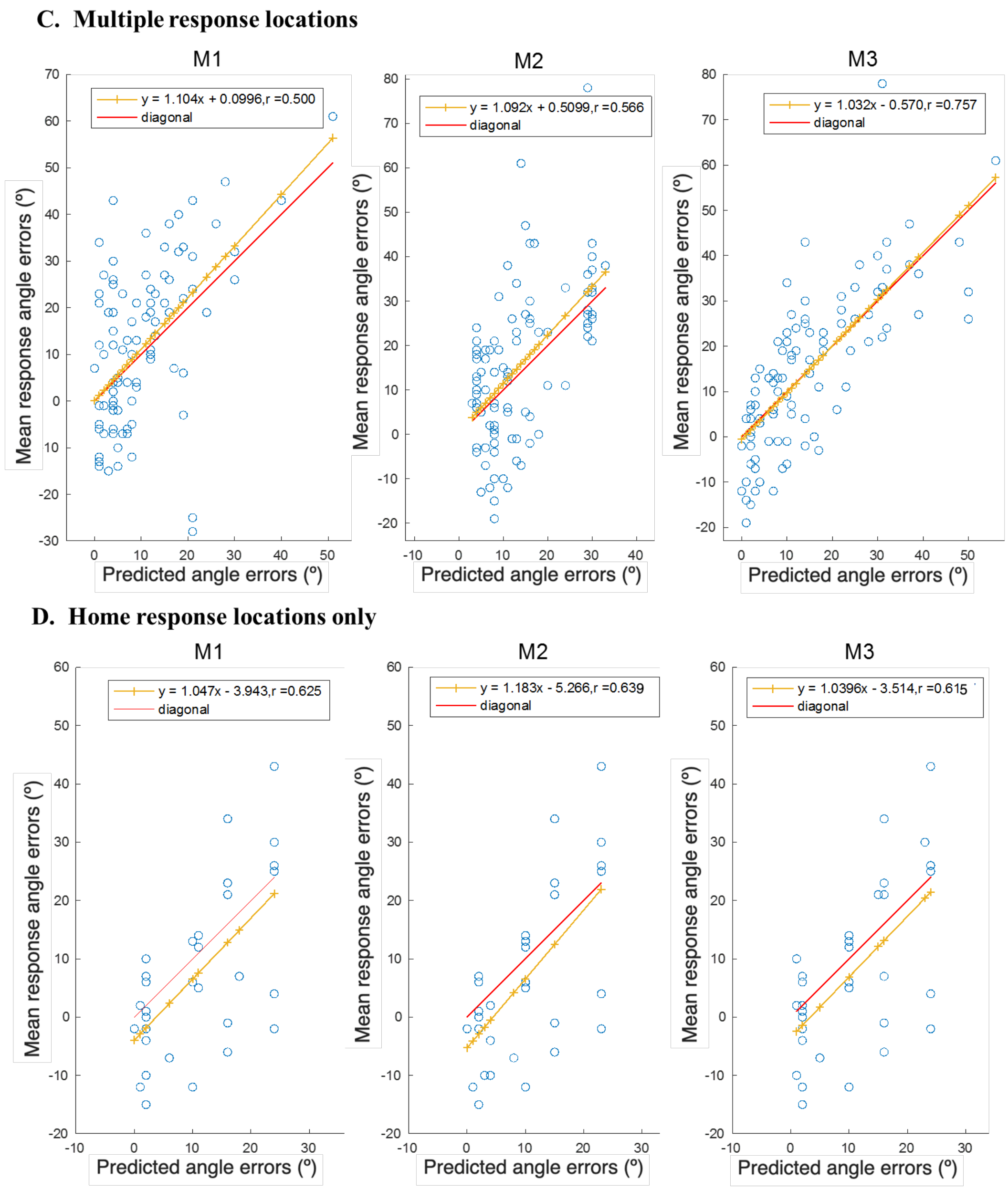


Figure 8. Illustrating the predicted errors in inbound path length (panels A and B) and turn angle (panels C and D) as a function of the mean response errors using multiple response locations or only home response locations. The diagonal lines in red (y = x) indicate the ideal outcome that the response errors are perfectly predicted. The yellow lines indicate the regression lines. Open dots depict the individual pairs of predicted errors and mean response errors across participants,

for each object, and for each path (32 paths in total), according to the encoding-error model (M1), execution-error model (M2), and bi-component model (M3), respectively.

The likelihood ratios were computed to compare the models' performance in predicting inbound length errors and angle errors. Following Glover and Dixon (2004), the likelihood ratio of favoring Model$_i$ over Model$_j$ (i.e., λ_{ij}) can be computed as

$$\lambda_{ij} = \left(\frac{1-r_j^2}{1-r_i^2}\right)^{\frac{n}{2}}, \tag{15}$$

where the r_i^2 and r_j^2 are squared mean correlation coefficients from Model$_i$ and Model$_j$ in Figure 8, indicating the variance that is explained by Model$_i$ and Model$_j$, respectively, and *n* is the number of data points. In the current example, *n* equals 96 (i.e., 32 paths × 3 response locations) for taking multiple response locations or equals 32 (i.e., 32 paths × 1 response location) for taking only home response locations into the cross-validation.

The results of likelihood ratios for the three competing models are reported in Table 5. For both length and angle errors, the method of employing multiple response locations demonstrates compelling evidence (i.e., five out of six likelihood ratios of over 100) that the bi-component model is superior to the encoding-error and execution-error models in describing mean response errors. However, no clear evidence (i.e., no likelihood ratios of over 2) is presented by employing only home response locations, showing that it cannot distinguish between models in terms of predictive power.

Table 5

Maximum likelihood ratios (λ) for competing models (row model over column model) in predicting inbound path length errors (left) and turn angle errors (right) using multiple locations or only home response locations.

	Length errors						Angle errors					
	Multiple response locations			Home response locations only			Multiple response locations			Home response locations only		
λ	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
M1												
M2	0.8			0.5—			113.1**			1.6—		
M3	161.5**	203.9**		1.0—	1.9—		5.6×10^{11}**	4.9×10^{9}**		0.7—	0.5—	

Note: * indicates clear evidence, i.e., LR > 3 or LR <1/3, and ** indicates strong evidence, i.e., LR > 10 or LR <1/10. — indicates no evidence (Glover & Dixon, 2004).

2.5 Groups of participants differing in compression pattern of the response

In the cross-validation described above, we did not consider the participant variable. For each model, we estimated the best model parameters being applied to all participants. However, participants might differ in the compression pattern (i.e., some had a strong compression pattern whereas others had a weak compression pattern), so the best model parameters for each group might be significantly different from each other. Therefore, the conclusions on a model comparison based on the best model parameters for all participants and based on the best model parameters for each group of participants might not be consistent. We considered the variability of participants' responses in their triangle completion and classified participants into two groups based on the compression pattern of the inbound responses.

As illustrated in Figure 9, the participants showed variations in their compression pattern (e.g., the slopes of the regression lines) of the inbound responses. The dots inside the blue box in Figure 9C-D represent the participants who showed a compression pattern (i.e., with a slope between 0 and 1, and intercept larger than 0) or had strong compression whereas the dots outside the blue box represent the participants who did not show compression pattern or had weak compression. Considering compression patterns in both length and angle, we could also divide participants into four groups based on both (47 in strong for both, 13 in weak for both, 22 in strong for angle and weak for length, 30 in weak for angle and strong for length). However, we might not be able to conduct meaningful 5 × 2 cross-validations for all four groups, especially the group with only 13 participants. Hence, we divided participants into two groups instead of four so that we had enough participants in each group for 5 × 2 cross-validations.

Across the regression lines of individual participants, the correlation coefficient (r) was significantly higher in the inbound turn angle (Figure 9D) than in the inbound path length (Figure 9C) (mean $r = 0.65$ for angle and mean $r = 0.39$ for length), $t\,(111) = 6.36$, $p < .001$, Cohen's $dz = .60$. Moreover, the number of participants showing significant correlations ($p \leq .05$) was significantly larger in the regression for inbound turn angle (Figure 9D) than for the inbound path length (Figure 9C) (61 participants for angle and 22 participants for length, sharing 5 participants with significant correlations in both), McNemar's $\chi^2\,(1) = 16.01$, $p < .001$. Hence, the compression patterns of individual participants in terms of inbound turn angle were much more reliable than in terms of inbound path length. Consequently, we classified the participants into two groups according to their compression on the inbound turn angle: the strong compression group (69 participants showing compression) and the weak compression group (43 participants showing no compression). Moreover, the distribution of participants in compression groups in terms of length was independent of in terms of angle ($\chi^2\,(1) = .03$, $p = .86$), indicating that the strong and weak compression groups only based on angle had similar proportions of participants with strong and weak compression in length. Therefore, the strong compression group had strong compression in angle and average compression in length whereas the weak compression group had weak compression in angle and average compression in length.

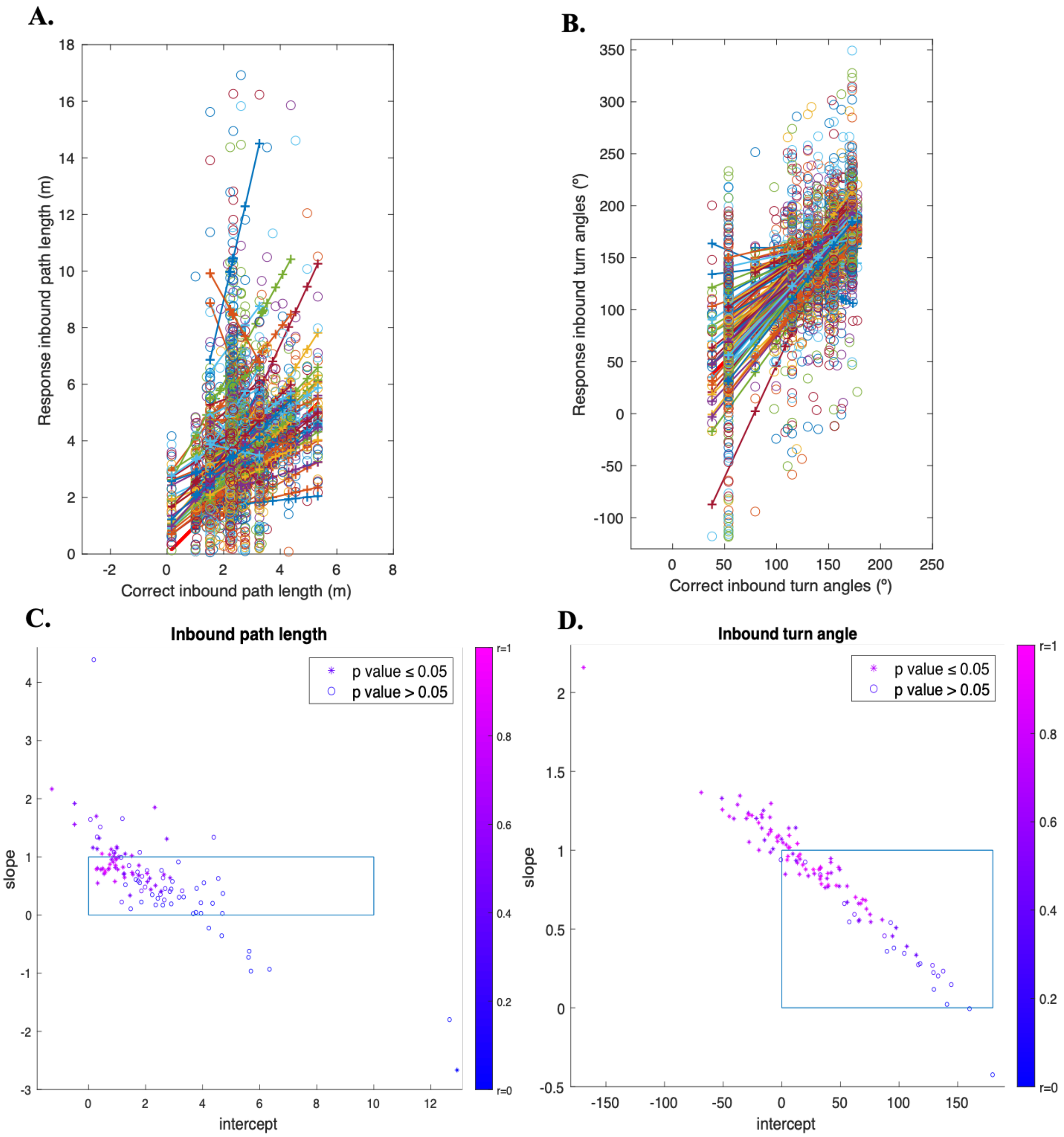


Figure 9. Each line indicates the linear regression of response values on the correct values for one participant in terms of inbound path length **(**A) and turn angle (B), respectively. (C-D) illustrate the slope-intercept, correlation coefficient (i.e., *r*-value), and its significance (i.e., *p*-

value) of the linear regression relationship in terms of inbound path length (C) and turn angle (D), respectively.

We conducted model validation for each group using the model parameters estimated in the model fitting described in section 2.4.1 (see details in Supplementary Materials and Tables S5-S8). Model validation based on the parameters from the algorithm using multiple locations showed that all the three models (M1-M3) even performed worse than the baseline model (M0) for the weak compression group (negative Partial R^2 in Table S5) although the bi-component model (M3) was still the best model for the strong compression group. These findings suggest that the best model parameters for all participants might not be appropriate for the weak compression group. Therefore, it is important to conduct cross-validation for each group and then calculate the overall model performance.

2.6 Cross-validation for different groups

We conducted 5 × 2 cross-validations for each group of compression. As we primarily used model validation performance in model comparison, we did not report the fitting results of two compression groups for the interest of brevity (see Supplementary Materials Table S9 for the averaged fitting performance across ten folds).

2.6.1 Model validation

As illustrated in Tables 6, 7, and 8, the algorithm using home response locations only could not differentiate the three models (M1-M3) regardless of the compression group.

The algorithm using multiple response locations showed different model comparison results for the strong and weak compression groups. For the strong compression group, generalizability measures in Table 6, likelihood ratios in Table 7, and the results of *Alpaydin's F-*

test on dRMSE in Table 8 (also see Table S10 for RMSEs of individual folds) all suggest that the bi-component model (M3) was the best. By contrast, for the weak compression group, none of the generalizability measures, likelihood ratios, or *Alpaydin's F-test* on dRMSE could differentiate the four models including the baseline model.

Table 6

Model validation performance for the strong (upper) and weak (lower) compression groups. Parameters are estimated from model fitting for each corresponding group. The RMSE, maximum log-likelihood, and partial r-squared are generalizability measures, which were calculated by applying the parameters to the test subsamples.

	Strong compression group										
	Multiple response locations										
	Parameters								5×2 Validation		
Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	3.382	-3770.4	0
M1	1.14	0.49	0.79	15.45	1	0	1	0	3.214	-3692.0	0.096
M2	1	0	1	0	0.60	1.86	0.68	58.50	3.125	-3645.5	0.146
M3	0.64	1.19	0.88	18.69	0.57	1.76	0.72	52.07	3.084	-3623.2	0.168
	Home response locations only										
	Parameters								5×2 Validation		
Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	3.037	-1190.9	0
M1	0.84	0.58	0.46	18.64	1	0	1	0	2.738	-1139.7	0.186
M2	1	0	1	0	0.44	2.42	0.50	81.24	2.745	-1140.7	0.182
M3	2.88	0.84	0.43	12.46	0.72	0.86	1.51	12.50	2.743	-1140.5	0.183
	Weak compression group										
	Multiple response locations										
	Parameters								5×2 Validation		
Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	2.817	-2169.2	0
M1	0.86	0.46	0.80	21.2	1	0	1	0	2.814	-2169.3	0.002
M2	1	0	1	0	0.85	0.46	1.00	2.96	2.816	-2170.0	5.48E-04
M3	0.81	0.56	0.80	21.9	0.92	0.17	1.04	-1.73	2.810	-2168.3	0.005

Model	Home response locations only										
	Parameters								5×2 Validation		
	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	2.379	-662.8	0
M1	0.51	0.69	0.44	33.0	1	0	1	0	2.329	-657.0	0.041
M2	1	0	1	0	0.44	1.43	0.44	84.44	2.328	-656.9	0.042
M3	1.54	11.7	2.25	17.8	0.58	0.06	2.24	17.13	2.342	-659.1	0.029

Table 7

Maximum likelihood ratio (LR) between models (row model over column model) in model validation for the strong (upper) and weak (lower) compression groups using multiple locations (left) or only home response locations (right).

	Strong compression group							
	Multiple response locations				Home response locations only			
LR	M0	M1	M2	M3	M0	M1	M2	M3
M1	$1.06 \times 10^{34**}$				$1.64 \times 10^{22**}$			
M2	$1.72 \times 10^{54**}$	$1.62 \times 10^{20**}$			$6.03 \times 10^{21**}$	0.37—		
M3	$8.03 \times 10^{63**}$	$7.58 \times 10^{29**}$	$4.66 \times 10^{9**}$		$7.84 \times 10^{21**}$	0.48—	1.30—	
	Weak compression group							
	Multiple response locations				Home response locations only			
LR	M0	M1	M2	M3	M0	M1	M2	M3
M1	0.88—				322.58**			
M2	0.43—	0.49—			370.37**	1.15—		
M3	2.50—	2.84—	5.81*		40.32**	0.13*	0.11*	

Note: * indicates clear evidence, i.e., LR > 3 or LR <1/3, and ** indicates strong evidence, i.e., LR > 10 or LR <1/10. — indicates no evidence (Glover & Dixon, 2004).

Table 8

Alpaydin's F-test examining the differences in RMSE (dRMSE) between models (the row model minus the column model) for the group with strong (upper) and weak (lower) compression patterns when using multiple locations (left) or only home response locations (right).

Strong compression group									
Multiple response locations					Home response locations only				
dRMSE	M0	M1	M2	M3	dRMSE	M0	M1	M2	M3
M1	-.168**				M1	-.299**			
M2	-.257***	-.125**			M2	-.292**	.007—		
M3	-.298***	-.13***	-.041*		M3	-.294***	.005—	.002—	

Weak compression group									
Multiple response locations					Home response locations only				
dRMSE	M0	M1	M2	M3	dRMSE	M0	M1	M2	M3
M1	.003—				M1	.050—			
M2	.001—	-.002—			M2	.051—	-.001—		
M3	.007—	-.004—	-.006—		M3	.037—	.014—	.014—	

We also compared the overall performance of all models by combining the locational residuals of the two compression groups (see Tables 9-11 for generalizability measures, likelihood ratios, and the results of *Alpaydin's F-test*). Figure 10 visually illustrates the locational residuals of individual targets achieved by different models using the two algorithms. All results suggest that the bi-component model was the best based on the cross-validation using multiple response locations whereas there was no best model based on the cross-validation using home response.

Table 9

The overall performance of model validation of the two compression groups using multiple locations (upper) or only home response locations (lower). Parameters are the weighted average of the best parameters for each group (weighted by the numbers of participants in different groups). The RMSE, maximum log-likelihood, and partial r-squared are generalizability measures, which were based on the combined locational residuals of the two compression groups.

Multiple response locations	
Parameters	5×2 Validation

Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	3.179	-5964.5	0
M1	1.03	0.48	0.79	17.64	1	0	1	0	3.069	-5878.0	0.067
M2	1	0	1	0	0.70	1.32	0.81	37.18	3.012	-5828.4	0.102
M3	0.71	0.95	0.85	19.91	0.70	1.15	0.84	31.41	2.984	-5803.2	0.118

	Home response locations only										
	Parameters								5×2 Validation		
Model	$\theta_{L\_s}^{enc}$	$\theta_{L\_i}^{enc}$	$\theta_{A\_s}^{enc}$	$\theta_{A\_i}^{enc}$	$\theta_{L\_s}^{exe}$	$\theta_{L\_i}^{exe}$	$\theta_{A\_s}^{exe}$	$\theta_{A\_i}^{exe}$	RMSE	MaxLogL	Partial R^2
M0	1	0	1	0	1	0	1	0	2.805	-1868.3	0
M1	0.72	0.63	0.45	24.14	1	0	1	0	2.591	-1807.2	0.146
M2	1	0	1	0	0.44	2.04	0.48	82.47	2.595	-1807.9	0.143
M3	2.37	5.03	1.13	14.50	0.66	0.55	1.79	14.28	2.600	-1809.7	0.141

Table 10

The overall results of the maximum likelihood ratio (LR) between models (row model over column model) in model validation using multiple locations (left) or only home response locations (right).

	Multiple response locations				Home response locations only			
LR	M0	M1	M2	M3	M0	M1	M2	M3
M1	$3.86 \times 10^{37**}$				$3.47 \times 10^{26**}$			
M2	$1.29 \times 10^{59**}$	$3.35 \times 10^{21**}$			$1.69 \times 10^{26**}$	$0.49^{—}$		
M3	$1.08 \times 10^{70**}$	$2.80 \times 10^{32**}$	$8.36 \times 10^{10**}$		$2.74 \times 10^{25**}$	0.08^{**}	0.16^{*}	

Note: * indicates clear evidence, i.e., LR > 3 or LR <1/3, and ** indicates strong evidence, i.e., LR > 10 or LR <1/10. — indicates no evidence (Glover & Dixon, 2004).

Table 11

The overall results of Alpaydin's F-test examining the differences in RMSE (dRMSE) between models (the row model minus the column model) when using multiple locations (left) or only home response locations (right).

Multiple response locations					Home response locations only				
dRMSE	M0	M1	M2	M3	dRMSE	M0	M1	M2	M3
M1	-.110*				M1	-.214**			
M2	-.167***	-.057**			M2	-.210**	.004—		
M3	-.195***	-.085***	-.028*		M3	-.207**	.008—	.003—	

Note: Asterisks denote significant dRMSE (***p < .001; **p < .01; *p < .05) and a dash (—) indicates non-significant dRMSE.

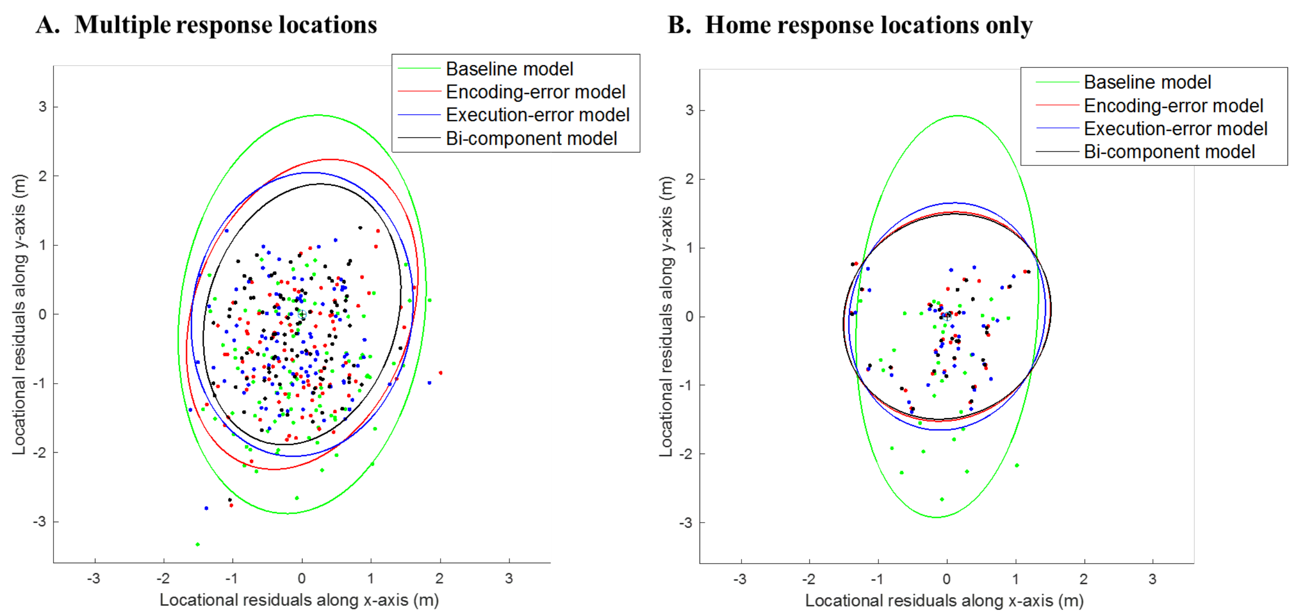


Figure 10. Visualizing the differences (locational residuals) between mean response locations and mean predicted locations from different models using (A) multiple response locations or (B) only home response locations. The open circle with a cross at (0, 0) indicates the response location, the coordinate of which varied in real experiments but is set to (0, 0) as a reference. Individual dots represent coordinates of the locational residuals for all targets (96 targets in A and 32 in B). Ellipses indicate the 95% density contours of the bivariate normal distributions with zero means ($\mu = (0,0)$) and covariance matrix (Σ) of the locational residuals according to the

baseline model (green), encoding-error model (red), execution-error model (blue), and bi-component model (black), respectively.

2.6.2 Model recovery using varied values of parameters across participants

In the model recovery described above (see section 2.4.3), we used the fixed values of model parameters for all participants (Table 1) to produce simulated locations based on each true model. The simulation results indicated that the algorithm using multiple objects could recover the true models very well whereas the algorithm using home locations could not recover the true models (see Figure 6 for the confusion matrix). As participants showed different compression patterns (Figure 9), it is important to examine whether the algorithms can still recover the true model when varied values of model parameters are used to create simulated locations (below we refer to it as *model recovery with varied parameter values* and refer to the previous one as *model recovery with fixed parameter values*). Note that we conducted 5 × 2 cross-validations for strong and weak compression groups to address the issue of participants' differences in the compression pattern. Unfortunately, 5 × 2 cross-validation is not feasible for each participant. Conducting model recovery with varied parameter values is especially important as it can further address the issue of participants' differences in compression patterns. If we demonstrate that 5 × 2 cross-validations using the multiple response locations can recover the true model in model recovery with varied parameter values, our conclusion based on 5 × 2 cross-validations using the multiple response locations should also be able to recover the true model using participants' response locations.

Same as the *model recovery with fixed parameter values*, we still created 100 sets of simulated response locations from each model and conducted 5 × 2 cross-validations for all of them in conducting *model recovery with varied parameter values*. Difference from the *model recovery with fixed parameter values,* we used varied values for each of the intercept and slope parameters. Specifically, we sampled each parameter from a uniform distribution with a mean same as the fixed value of the model parameters in *model recovery with fixed parameter values* (i.e., the parameters illustrated in Table 1). The range of the uniform distribution for slope parameters was twice the distance between the mean slope and 1 (i.e., the upper limit). The range of the uniform distribution for intercept parameters was twice the distance between the mean intercept and 0. For example, $\theta_{L\_s}^{enc}$ in M3 (a slope parameter in Table 1) was sampled from a uniform distribution U (0.82 - $|1 - 0.82|$, 0.82 + $|1 - 0.82|$). $\theta_{A\_i}^{exe}$ in M3 (an intercept parameter in Table 1) was sampled from a uniform distribution U (34.21 - $|0 - 34.21|$, 34.21 + $|0 - 34.21|$). As a result, we created 112 samples for each parameter of each model and then assigned them randomly to 112 participants. Using the outbound paths and target locations of each participant, we created the simulated response locations based on each model by applying the assigned values of model parameters.

Figure 11 presents the confusion matrix in model recovery (frequency in each category of likelihood ratio in model validation was reported in Supplementary Materials Figure S3.). The results showed that the algorithm of using multiple response locations upon most occasions can successfully distinguish the true model from other models (64% for true model M1, 84% for true model M2, and 100% for true model M3). By contrast, the algorithm of using home response locations cannot clearly distinguish the true model from other models. In most cases, the algorithm could not find the best model (with a rate larger than 55%). Consequently, the algorithm could recover the true model at a very low rate (12% for true model M1, 28% for true model M2, and 9% for true model M3). Moreover, the algorithm also at times recovered distracting models.

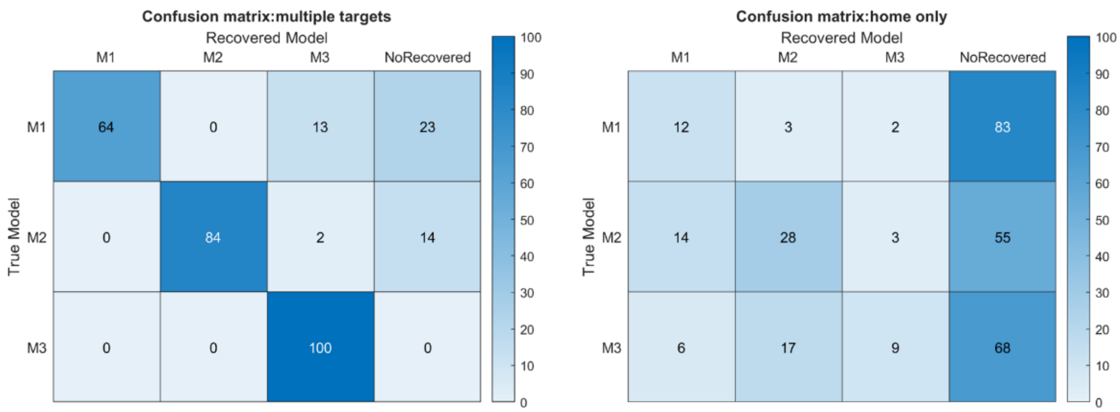


Figure 11. Confusion matrices in model recovery using multiple response locations (left) or home response locations only (right). The number in each cell indicates the frequency of the recovered model being the best model. NoRecoved means that no best model was recovered by the algorithm.

2.6.3 Similarity of parameters values estimated from real and simulated response locations

The similarity between parameters based on participants' response locations and based on simulated locations from different models was illustrated in Figure 12 (see exact parameters in Supplementary Materials and Table S11). The parameter distance was shortest when the simulated locations were produced by M3 (RMSE= 8.59, 6.82, and 1.13 for M1, M2, and M3 respectively). The parameters based on simulated locations from M3 explained the largest proportion of the total variance of the 16 parameters based on participants' response locations ($r^2 = 1 - \frac{MSE}{Var}$, r^2 = .56, .72, and .99 for M1, M2, and M3 respectively). The ratios of likelihood of M3 over other models were larger than 2.97×10^{12} (logL = -57.11, -53.41, and -24.69 for M1, M2, and M3 respectively).

Therefore, the similarity between participants' response locations and simulated locations from the bi-component model was the largest, suggesting the bi-component model was the best.

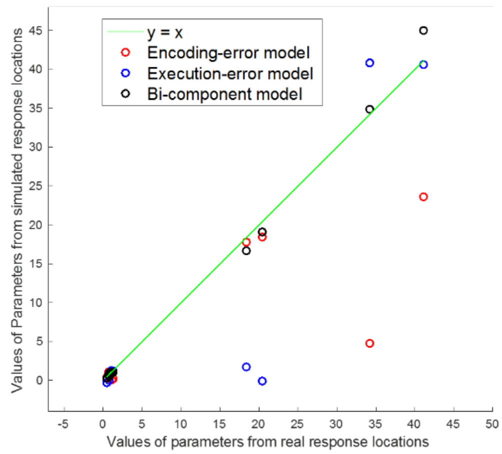


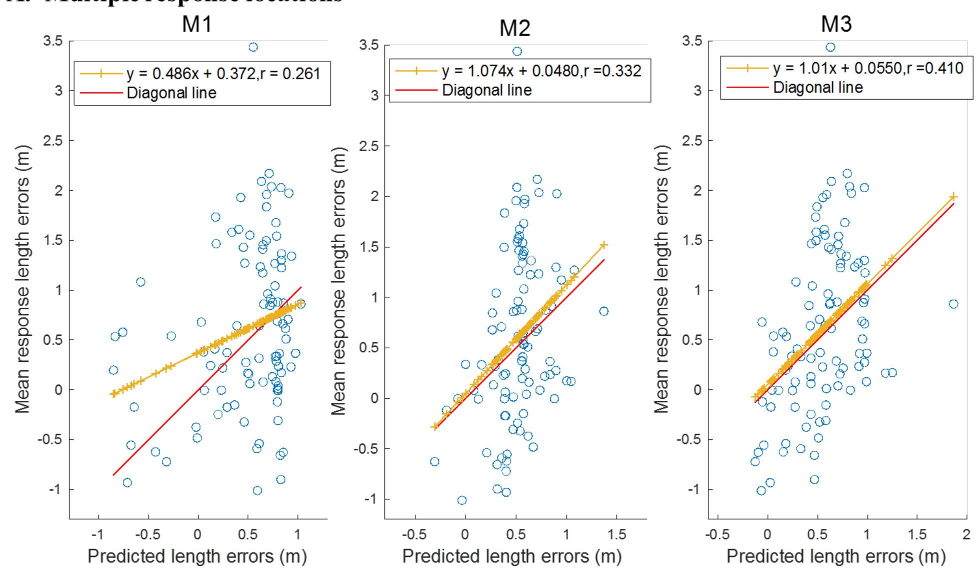
Figure 12. Illustrating the similarity of estimated parameters based on real data and simulated data from different models. The diagonal lines in green (y = x) indicate the ideal outcome that the parameters derived from real data are perfectly recovered from simulated data. Open dots depict the individual pairs of values of parameters based on real and simulated response locations for each model.

2.6.4 Predictive performance on the response error of participants based on best parameters for each group

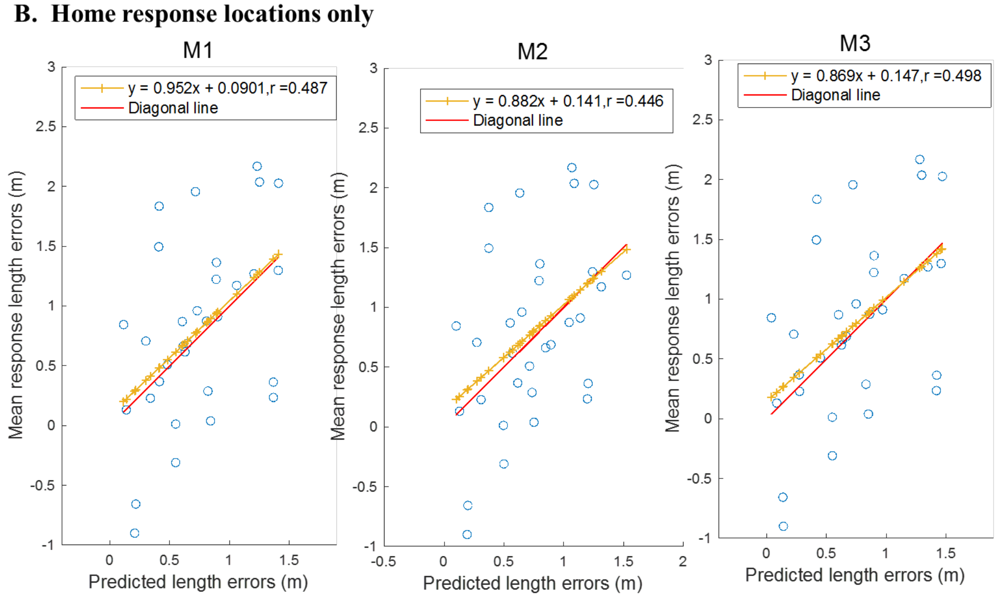
We compared the predictive performance of different models in terms of inbound length error and angle error, using the best parameters for each group. The predicted error and the mean response error (in terms of inbound path length or turn angle) for each target and each unique outbound path were defined and calculated in the same way mentioned above (2.4.5).

Figure 13 illustrates the mean predictive performance of different models in terms of inbound length error and angle error. Table 12 shows that the bi-component model (M3) had the highest correlation coefficients for both inbound length (Figure 13A) and angle errors (Figure 13C) when the cross-validation included multiple response locations of each outbound path. Nevertheless, the correlation coefficients of the three models (Figure 13B and Figure 13D) were comparable when the cross-validation only included the home response location of each outbound path.

A. Multiple response locations



B. Home response locations only



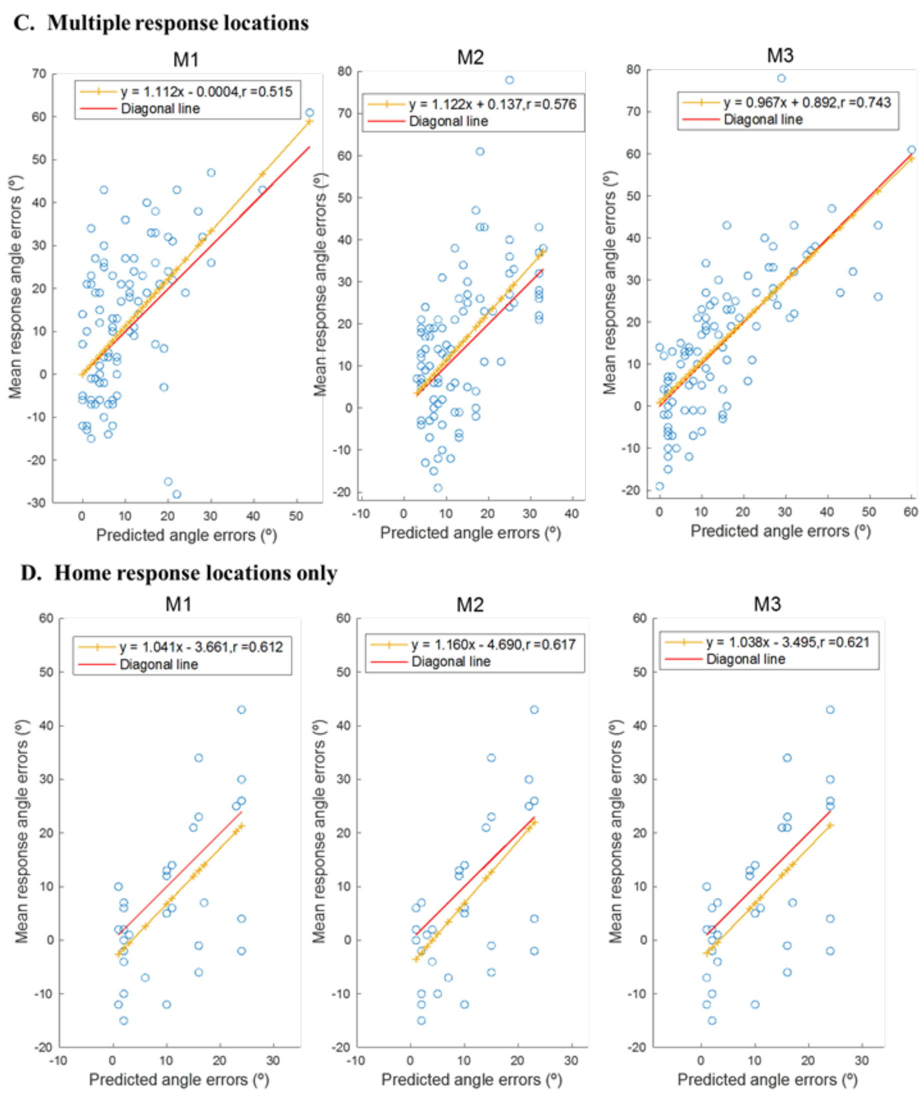


Figure 13. The overall performance of the predicted errors in inbound path length (panels A and B) and turn angle (panels C and D) as a function of the mean response errors using multiple response locations or only home response locations. The diagonal lines in red (y=x) indicate the ideal outcome that the response errors are perfectly predicted. The yellow lines indicate the regression lines. Open dots depict the individual pairs of predicted errors and mean response errors across participants, for each object and each path (32 paths in total), according to the

encoding-error model (M1), execution-error model (M2), and bi-component model (M3), respectively.

Table 12

Maximum likelihood ratios (λ) for competing models (row model over column model) in predicting inbound path length errors (left) and turn angle errors (right) using multiple locations or only home response locations.

	Length errors						Angle errors					
	Multiple response locations			Home response locations only			Multiple response locations			Home response locations only		
λ	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
M1												
M2	9.2^{*}			$0.5^{—}$			95.3^{**}			$1.2^{—}$		
M3	232.2^{**}	25.2^{**}		$1.3^{—}$	$2.8^{—}$		$2.1 \times 10^{10**}$	$2.2 \times 10^{8**}$		$1.3^{—}$	$1.1^{—}$	

Note: * indicates clear evidence, i.e., LR > 3 or LR <1/3, and ** indicates strong evidence, i.e., LR > 10 or LR <1/10. $^{—}$indicates no evidence (Glover & Dixon, 2004).

3. Discussion

The primary purpose of the current study was to identify the possible sources of the systematic biases in human path integration. We used model cross-validation to compare three plausible theoretical models (the encoding-error model, the execution-error model, and the bi-component model) in explaining the systematic errors of the inbound responses when participants only had idiothetic cues in the path integration conditions of Qi et al. (2021). There are two important findings. First, cross-validation modeling using all three inbound responses for each outbound path indicated that the bi-component model outperformed the encoding-error model (Fujita et al., 1993) and the execution-error model (Chrastil & Warren, 2021). This finding suggests that systematic biases in human path integration occurred in both encoding the outbound path and executing the desired inbound responses. Second, modeling using only the home response for each outbound path failed to distinguish among these three models.

To the best of our knowledge, the current study provided the first modeling evidence indicating that there are systematic biases in both encoding the outbound path (path lengths and turn angles) and in executing the desired inbound responses (path lengths and turn angles) in the triangle-completion task. The finding of both encoding and execution biases unified the encoding-error model (Fujita et al., 1993) and the execution-error model (Chrastil & Warren, 2021) into the bi-component model.

Although the finding of the current study appears to challenge the encoding-error model by undermining its assumption that there is no systematic bias in execution, it supports the key theoretical claims of the encoding-error model (Fujita et al., 1993; Klatzky et al., 1999; Loomis et al., 1993; Loomis et al., 1999). According to the encoding-error model (one version of the configural updating models), people encode the configuration of the outbound path by encoding the leg lengths and turn angles between legs. People calculate the inbound response based on the remembered outbound path. Therefore, the systematic biases (compression patterns) in encoding the outbound path should lead to the appearance of systematic biases in the inbound responses. The evidence of systematic encoding errors (i.e., the encoding functions of the bi-component model) provided by the current study is consistent with these claims. Note that although Fujita et al. (1993) showed that the encoding-error model well explained the compression patterns in the inbound responses, it could not remove the possibility that the compression patterns in the inbound responses were caused solely by the systematic biases in executing the inbound responses. Thus, we believe that the current study indeed provides clearer evidence for the encoding biases by separating the encoding biases from the execution biases.

The current evidence of systematic execution errors is in line with the past studies (Bakker et al., 1999, 2001), which demonstrated systematic inaccuracies in simply producing specific angles. Specifically, the participants in Bakker et al. (1999) were required to produce cardinal angles (e.g., 90º, 180º, 270º) around a point under different combinations of sensory feedback. Note that in this task participants did not need to encode the angles by locomotion or visually but were only informed of the angles verbally. The significant undershoot pattern in all conditions would reflect the systematic errors in execution.

Chrastil and Warren (2021) provided the first modeling evidence to indicate that there are systematic execution errors in the triangle-completion task. They separately estimated the encoding functions and the execution functions from reproduction tasks (the simple translation and rotation tasks) by assuming that there were only encoding biases or execution biases. They argued that if people only have systematic biases in encoding but not in execution, the encoding functions estimated from the reproduction task should well explain the systematic errors in the triangle-completion task. Their modeling results showed that the discrepancy between the predicted and observed inbound responses was greater when the predicted values were only based on the encoding functions than when the predicted values were only based on the execution functions. Thus, these results suggested that there were systematic execution errors. However, it is not clear whether the encoding functions or execution functions from the simple translation and rotation tasks are the same as those functions in the triangle-completion task. The current study, using cross-validation modeling, estimated encoding functions and execution functions in the triangle-completion task using half of the data measured in the triangle-completion task per se, instead of using other independent and simpler tasks (e.g., reproduction tasks in Chrastil & Warren, 2021). Therefore, the current study avoided the issues of assuming that the encoding functions or execution functions from the reproduction tasks are the same as those functions in the triangle-completion task. As the current study still showed that there are systematic biases in execution, separately from encoding biases, it provided clearer evidence for execution biases, one of the key claims of the execution-error model.

Chrastil and Warren (2021) also showed that the model with both encoding functions and execution functions did not outperform the model with only execution functions. In contrast, the current study indicated that both encoding biases and execution biases contributed to the biases in inbound responses. This discrepancy might occur because these two studies used different methods of estimating the encoding functions and execution functions. Chrastil and Warren (2021) estimated the encoding functions and the execution functions from reproduction tasks by assuming that there were only encoding biases or execution biases. They then used these encoding and execution functions in the model with both encoding and execution biases. However, the best parameters of encoding functions in the model with both biases may differ from the best parameters of encoding functions in the model with only encoding biases. Similarly, the best parameters of execution functions in the model with both biases may differ from the best parameters of execution functions in the model with only execution biases. By contrast, the current study estimated the encoding functions and the execution functions for the bi-component model independently rather than simply borrowing the encoding functions estimated for the encoding-error model and the execution functions estimated for the execution-error model. As shown in Tables 1 and 2, the parameters of encoding functions in the encoding-error model (M1) differ from the parameters of encoding functions in the bi-component model

(M3). The parameters of execution functions in the execution-error model (M2) also differ from the parameters of execution functions in the bi-component model (M3).

The finding that the bi-component model was the best is not attributed to more free parameters of the bi-component model than the other two models. In model validation, as the models were validated using the other halves of the data (test subsamples), the numbers of free parameters were the same for all three models. The likelihood ratio still showed the superiority of the bi-component model (see Table 3, left sub-table for multiple response locations). Furthermore, the findings of cross-validation modeling using the simulated response locations (multiple response locations) clearly indicated that if the true model was the encoding-error model (M1) or the execution-error model (M2), the bi-component model (M3) never outperformed the true model when the simulated locations were created using fixed values of parameters (Figure 6, upper panel) and seldom outperformed the true model when the simulated locations were created using varied values of parameters (Figure 11, upper panel).

In addition to using cross-validation, using multiple inbound responses for each outbound path is also critical to differentiate the bi-component model from the other two models. Different from the typical triangle-completion task with only one inbound response (i.e., the homing vector) for each outbound path, the triangle-completion task used in Qi et al. (2021) required participants to indicate three learned locations (including the home location) during the response phase. Previous studies indicated that one inbound response may not be able to recover participants' encoded positions and headings at the endpoint of the outbound path (e.g., Mou & Zhang, 2014). As one inbound response can be caused by many possible encoded positions and headings at the endpoint of the outbound path, this implies that the errors in the inbound response can be attributed to the encoding biases alone, the execution biases alone, or the combination of both. In contrast, multiple inbound responses (multiple target locations) for each outbound path can recover the participants' encoded positions and headings at the endpoint of the outbound path (e.g., Mou & Zhang, 2014; Qi et al., 2021; Zhang et al., 2020). Thus, we conjectured that the encoding functions and the execution functions can be separated by a cross-validation algorithm using multiple inbound responses (multiple target locations) for each outbound path. These insights were confirmed by the modeling results based on the empirical data of Qi et al. (2021) (see Tables 3 and 4) and based on the simulated data (see Figures 6 and 11 and also Tables S4 and S11).

One may argue that the different discrimination abilities of the algorithms using multiple response locations and using home response locations alone might be attributed to the number of data points. The number in the former was three times that in the latter. According to Formulas 12 and 15, the likelihood ratio is the proportion of x^n (x is the ratio of RMSE, n is the data number). To address this issue, we calculated $\sqrt[3]{LR}$ for the LRs of M3 over M1 (LR_{31}) and M2 (LR_{32}) in model validation using multiple response locations (see Table 3 left, $LR_{31} = 1.28 \times 10^{20}$ and $LR_{32} = 1.02 \times 10^{11}$). The results were 5.04×10^6 and 4672.33, which still showed strong evidence favoring M3. Therefore, the evidence of favoring M3 using multiple response locations and the lack of evidence of favoring M3 only using home response locations should not be attributed to the different number of data points.

The current study supported the bi-component model, which considers linear functions to represent the working mechanisms of both encoding and execution processes, on the basis of previous research (Chrastil & Warren, 2014; 2021; Fujita et al., 1993; Loomis et al., 1993). However, we do not claim that there would be an immutable set of parameters for the current model across all pathways and contexts. Klatzky et al. (1999) reflected that the parameters of the encoding functions based on the encoding-error model varied with the values of the outbound path (e.g., the path lengths of 1-3m or 4-6m). In addition, we admit that the encoding functions could also vary as Harootonian et al. (2020) showed that encoding functions of turn angles could be removed from their version of the encoding-error model when participants walked much longer paths.

Additional studies are needed to examine the applicability of the bi-component model under various conditions, such as path integration on more complex paths, since navigators may adopt different navigational strategies depending on the complexity of the path (Klatzky et al., 1990; Wiener et al., 2011; Wiener & Mallot, 2006). On simple pathways, navigators are more likely to remember the path configuration, and calculate the vector to go home only when needed (that is, an *offline* process), which is a *configural* strategy; On complex pathways, however, storing the presentation of the path configuration is challenging for navigators, and they tend to switch to continuously updating the homing vector (that is, an *online* process), which is a *continuous* strategy. Wiener and Mallot (2006) demonstrated that participants pointed homeward even faster and more accurately as path complexity increased while maintaining the overall path length, turn angle, and turning direction constant. In addition, an outbound path with path crossover might also be hard to encode the configuration (Fujita et al., 1993; Klatzky et al., 1990). However, Yamamoto et al. (2014) found that the presence of path crossover in traveled paths caused little impact on path integration performance. Future studies may test the bi-component model using outbound paths with more turns and path crossover.

We acknowledge that the current study examined the sources of systematic biases in homing when participants pointed to the targets including the home object. In other studies, which tackled similar research questions (Chrastil & Warren, 2021; Fujita et al., 1993;

Harootonian et al., 2020), participants physically walked back home. We do not believe that this method discrepancy should undermine the conclusion of the current study because of the following evidence. First of all, although not as often as walking to the origin, pointing to the origin was still often used in the history of studying human path integration. In a review chapter on human path integration, Loomis and his colleagues wrote “Other variants of path completion have had the subject indicate only the direction of the origin from the dropoff point, typically by pointing to it using a protractor (e.g., Able & Gergits, 1985; Adler & Pelkie, 1985; Baker, 1985; Gould, 1985; Klatzky et al., 1998; Rieser & Frymire, 1995; Sadalla & Montello, 1989; Sholl, 1989).” (Loomis et al., 1999, p. 134). Hence, pointing, in addition to walking, can be used to study path integration.

Second, to our best knowledge, there is no study showing that walking and pointing to the origin led to different conclusions about human navigation. Rather, studies using either pointing or walking showed the same results. Tcheang et al. (2011) showed that participants after adapting to a smaller vision-locomotion gain (i.e., visual cues indicated a smaller turn angle than did locomotion), overestimated the inbound turn angle in the following triangle completion task without vision. This result indicated that participants underestimated the turn angle in the outbound path because of the smaller gain. Du et al. (2020) replicated this result although participants in Tcheang et al. (2011) walked to the origin while participants in Du et al. (2020) pointed to the origin. Hence, underestimating the turn angle in the outbound path led to overestimating the inbound turn angle regardless of whether the response methods were walking or pointing. Thus, pointing, in addition to walking, can examine the biases of encoding the outbound path.

Can pointing, in addition to walking, be used to examine the biases of executing the desired inbound path? Walking (including walking forward and turning the body) and pointing appear to be two different kinds of actions. While walking is gradual (e.g., step by step), pointing seems more immediate. One may assume that execution biases occur in gradual actions but not in immediate actions. Following this assumption, one may speculate that pointing has very minimal execution errors. This speculation sounds reasonable but is inconsistent with the findings of the current study. The current study demonstrated the compression patterns (slope is smaller than 1 and intercept is larger than 0) in both inbound path length and inbound turn angle on the group level and individual levels (Figures 3 and 9). Furthermore, the best model (i.e. the bi-component model) clearly showed the compression pattern in the execution functions for both length ($\theta_{L\_s}^{exe}$ = 0.69 and $\theta_{L\_i}^{exe}$ = 1.10) and angle ($\theta_{A\_s}^{exe}$ = 0.82 and $\theta_{A\_i}^{exe}$ = 34.21) (see Table 1 for M3 using multiple locations). Therefore, pointing can reflect the execution biases. Hence, there is no reason to believe that the compression patterns in inbound pointing responses in the current study were caused by a mechanism different from that caused the compression patterns in inbound walking responses.

We speculated that one of the reasons why pointing to the origin was less used than walking to the origin in the research of human path integration is that in real environments, pointing may generally only indicate the direction of the origin whereas walking can indicate both direction and distance of the origin. However, nowadays in immersive virtual environments, participants could point to the exact location of the home with a virtual stick in a relatively small environment (e.g., up to 6m in Qi et al. (2021), see Figure S1 in the current paper). We argue that pointing is a more effective way to study human path integration. First, it is fast to collect participants' pointing responses than walking responses. Second, there are fewer safety issues or space requirements to collect participants' pointing responses than walking responses. Last, it is possible to collect several inbound pointing responses for a single outbound path, which is important as the current study showed that the algorithm using multiple responses could differentiate models but the algorithm using homing only could not differentiate models.

Participants in the current study pointed to three objects after each outbound path, which provided a unique opportunity to differentiate models. However, one may be wondering whether the task of pointing to multiple objects invokes spatial updating mechanisms different from that used in pointing to the home location only. When people keep track of three objects during locomotion, they might only be able to update self-to-object vectors and have no extra resources to update the path configuration at the same time. In contrast, when people only keep track of the home location, they might have enough resources to update both the self-to-object vector and path configuration. Hence, participants pointing to three objects in the current study might have been less likely to have configual updating than those who only had a homing response in the typical homing studies (Kearns et al., 2002; Klatzky et al., 1999). We appreciated this concern but argued that this concern had been addressed by the learning procedure in the paradigm of pointing to multiple objects used in the current study.

Mou and Zhang (2014), when originally introducing the paradigm of pointing to multiple objects in the inbound phase, acknowledged and addressed the issue of different memory loads in the paradigms of pointing to multiple objects and pointing to the origin only. They wrote "participants were allowed enough time to learn the directions of five objects accurately (see details in Experiment 1 for the evidence). When participants replaced the objects, they used a visible virtual stick to indicate the positions without any time pressure to ensure that they executed their responses as accurately as possible." (Mou & Zhang, 2014, p.557). Zhang et al.

(2020) directly compared the paradigm of pointing to multiple objects with the paradigm of pointing to the home location when they investigated whether the Bayesian cue combination occurred prior to or during homing. Their results in experiments 1 and 2 showed the same results, that is no Bayesian cue combination in homing when the second leg of the outbound path was much longer than the first leg of the outbound path. Furthermore, Lu et al. (2020) showed that online/offline spatial updating (analogue to continuous/configural updating) was not only determined by the number of objects to update during locomotion but also by the fidelity of spatial memory. When the same objects were placed at the same locations across all updating trials, participants appeared to use offline spatial updating regardless of the number of objects to update.

Therefore, as long as participants had well-learned target locations before walking the outbound path in the paradigm of pointing to multiple objects, they used the updating mechanisms similar to participants in the typical homing paradigm. Participants in the current study (i.e., Qi et al., 2021) had enough time to learn the three object locations. Furthermore, they saw the non-home objects at the same locations across all outbound paths so they should have learned the locations of objects very well. As a result, in addition to execution biases, the current study showed encoding biases, suggesting that participants in the current study still used configural updating.

One potential limitation of the current model is presuming minimal systematic integration errors, as with previously proposed models of path integration (Benhamou & Séguinot, 1995; Chrastil & Warren, 2021; Fujita et al., 1993; Harootonian et al., 2020). The integration errors emerge from computing the desired inbound responses based on the internalized representation of the traversed path. In addition to cognitive maps, humans also build labeled graphs (Warren, 2019; Warren et al., 2017), and the difference between these two may reflect the involvement of integration errors. One conjecture is that as the complexity of the outgoing path increases, the integration errors will subsequently surge (if one keeps using the *configural* navigation strategy). Future modeling studies may consider some possible systematic biases in the integration errors instead of assuming that there were random integration errors.

5. Conclusions

The results of modeling, using multiple inbound responses for each outbound path, support a bi-component model that incorporates both systematic biases in encoding the outbound path and executing the desired inbound responses to account for the systematic errors (regression to mean pattern) in the inbound responses. In addition, the results of modeling using only the home response for each outbound path could not dissociate the bi-component model from the encode-error model and the execution-error model. Our findings reconcile the execution-error model with the encoding-error model of human path integration. Furthermore, the current study demonstrates that cross-validation modeling using multiple inbound responses for each outbound path can be a powerful tool to understand human path integration.

Acknowledgments

This work was funded by the Natural Sciences and Engineering Research Council of Canada to Weimin Mou.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Caski (Eds.), *Second international symposium on information theory* (pp. 267-281). Budapest: Akademiai Kiado.

Alpaydm, E. (1999). Combined 5× 2 cv F test for comparing supervised classification learning algorithms. *Neural Computation*, *11*(8), *1885-1892*.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40-79.

Bakker, N. H., Werkhoven, P. J., & Passenier, P. O. (1999). The effects of proprioceptive and visual feedback on geographical orientation in virtual environments. *Presence: Teleoperators & Virtual Environments, 8*(1), 36-53.

Bakker, N. H., Werkhoven, P. J., & Passenier, P. O. (2001). Calibrating visual path integration in VEs. *Presence, 10*(2), 216-224.

Benhamou, S., & Séguinot, V. (1995). How to find one's way in the labyrinth of path integration models. *Journal of Theoretical Biology, 174*(4), 463-466.

Chrastil, E. R., & Warren, W. H. (2014). Does the human odometer use an extrinsic or intrinsic metric?. *Attention, Perception, & Psychophysics, 76*(1), 230-246.

Chrastil, E. R., & Warren, W. H. (2017). Rotational error in path integration: Encoding and execution errors in angle reproduction. *Experimental Brain Research, 235*(6), 1885-1897.

Chrastil, E. R., & Warren, W. H. (2021). Executing the homebound path is a major source of error in homing by path integration. *Journal of Experimental Psychology: Human Perception and Performance, 47*(1), 13-35.

Collett, M., & Collett, T. S. (2000). How do insects use path integration for their navigation?. *Biological Cybernetics, 83*(3), 245-259.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation, 10*(7), 1895-1923.

Du, Y., Mou, W., & Zhang, L. (2020). Unidirectional influence of vision on locomotion in multimodal spatial representations acquired from navigation. *Psychological Research, 84*(5), 1284-1303.

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

Etienne, A. S., Maurer, R., & Séguinot, V. (1996). Path integration in mammals and its interaction with visual landmarks. *The Journal of Experimental Biology, 199*(1), 201-209.

Fujita, N., Klatzky, R. L., Loomis, J. M., & Golledge, R. G. (1993). The encoding-error model of pathway completion without vision. *Geographical Analysis, 25*(4), 295-314.

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

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple and flexible statistic for empirical psychologists. *Psychonomic Bulletin & Review, 11*(5), 791-806.

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), e1009222.

Harootonian, S. K., Wilson, R. C., Hejtmánek, L., Ziskin, E. M., & Ekstrom, A. D. (2020). Path integration in large-scale space and with novel geometries: Comparing vector addition and encoding-error models. *PLoS Computational Biology, 16*(5), e1007489.

He, Q., & McNamara, T. P. (2018). Spatial updating strategy affects the reference frame in path integration. *Psychonomic Bulletin & Review, 25*(3), 1073-1079.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98*(3), 352-376.

Jacobs, L. F., & Schenk, F. (2003). Unpacking the cognitive map: The parallel map theory of hippocampal function. *Psychological Review, 110*(2), 285-315.

Jordan, M. I. (2003). *An Introduction to Probabilistic Graphical Models.* Chapter 13, Unpublished manuscript. Department of Statistics, University of California. https://people.eecs.berkeley.edu/~jordan/prelims/chapter13.pdf

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

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

Klatzky, R. L., Beall, A. C., Loomis, J. M., Golledge, R. G., & Philbeck, J. W. (1999). Human navigation ability: Tests of the encoding-error model of path integration. *Spatial Cognition and Computation, 1*(1), 31-65.

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science, 9*(4), 293-298.

Klatzky, R. L., Loomis, J. M., Golledge, R. G., Cicinelli, J. G., Doherty, S., & Pellegrino, J. W. (1990). Acquisition of route and survey knowledge in the absence of vision. *Journal of Motor Behavior, 22*(1), 19-43.

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.

Loomis, J. M., Klatzky, R. L., Golledge, R. G., & Philbeck, J. W. (1999). Human navigation by path integration. *Wayfinding behavior: Cognitive Mapping and Other Spatial Processes*, 125-151.

Lu, R., Yu, C., Li, Z., Mou, W., & Li, Z. (2020). Set size effects in spatial updating are independent of the online/offline updating strategy. *Journal of Experimental Psychology: Human Perception and Performance, 46*(9), 901-911.

May, M., & Klatzky, R. L. (2000). Path integration while ignoring irrelevant movement. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 169–186.

McNamara, T. P., & Chen, X. (2021). Bayesian decision theory and navigation. *Psychonomic Bulletin & Review*, 1-32.

Mittelstaedt, H., & Mittelstaedt, M. L. (1982). Homing by path integration In F. Papi & H. G. Wallraff (Eds.), *Avian navigation* (pp. 290–297). Berlin, Germany: Springer-Verlag.

Mittelstaedt, M. L., & Mittelstaedt, H. (1980). Homing by path integration in a mammal. *Die Naturwissenschaften, 67*(11), 566-567.

Mou, W., & Zhang, L. (2014). Dissociating position and heading estimations: Rotated visual orientation cues perceived after walking reset headings but not positions. *Cognition, 133*(3), 553-571.

Müller, M., & Wehner, R. (1988). Path integration in desert ants, Cataglyphis fortis. *Proceedings of the National Academy of Sciences, 85*(14), 5287-5290.

Péruch, P., May, M., & Wartenberg, F. (1997). Homing in virtual environments: Effects of field of view and path layout. *Perception, 26*(3), 301-311.

Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *Journal of Neuroscience, 31*(47), 17220-17229.

Qi, Y., Mou, W., & Lei, X. (2021). Cue combination in goal-oriented navigation. *Quarterly Journal of Experimental Psychology, 74*(11), 1981-2001.

Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software, 3*(24), 638.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. *Encyclopedia of Database Systems, 5*, 532-538.

Rieser, J. J., Ashmead, D. H., Talor, C. R., & Youngquist, G. A. (1990). Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception, 19*(5), 675-689.

Rieser, J. J., Pick, H. L., Ashmead, D. H., & Garing, A. E. (1995). Calibration of human locomotion and models of perceptual-motor organization. *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 480-497.

Rieser, J. J., & Rider, E. A. (1991). Young children's spatial orientation with respect to multiple targets when walking without vision. *Developmental Psychology*, *27*(1), 97-107.

Saint Paul, U. V. (1982). Do geese use path integration for walking home? In F. Papi, & H. G. Wallraff (Eds.), *Avian navigation* (pp. 298–307). New York: Springer.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461-464.

Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Perception & Psychophysics, 1*(5), 439-446.

Taboga, Marco (n.d.). *Multivariate normal distribution - Maximum Likelihood Estimation*. StatLect. https://www.statlect.com/fundamentals-of-statistics/multivariate-normal-distribution-maximum-likelihood.

Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Perception & Psychophysics, 24*(4), 305-314.

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

Wang, R. F. (2016). Building a cognitive map by assembling multiple path integration systems. *Psychonomic Bulletin & Review, 23*(3), 692-702.

Warren, W. H. (2019). Non-euclidean navigation. *Journal of Experimental Biology, 222*(jeb187917), 1–10.

Warren, W. H., Rothman, D. B., Schnapp, B. H., & Ericson, J. D. (2017). Wormholes in virtual space: From cognitive maps to cognitive graphs. *Cognition, 166*, 152-163.

Wartenberg, F., May, M., & Péruch, P. (1998). Spatial orientation in virtual environments: Background considerations and experiments. In C. Freska, C. Habel & K. F. Wender (Eds.), *Spatial cognition* (pp. 469-489). Berlin: Springer.

Wehner, R., Michel, B., & Antonsen, P. (1996). Visual navigation in insects: Coupling of egocentric and geocentric information. *The Journal of Experimental Biology*, *199*(1), 129-140.

Wiener, J. M., Berthoz, A., & Wolbers, T. (2011). Dissociable cognitive mechanisms underlying human path integration. *Experimental Brain Research, 208*(1), 61-71.

Wiener, J. M., & Mallot, H. A. (2006). Path complexity does not impair visual path integration. *Spatial Cognition and Computation, 6*(4), 333-346.

Yamamoto, N., Meléndez, J. A., & Menzies, D. T. (2014). Homing by path integration when a locomotion trajectory crosses itself. *Perception, 43*(10), 1049-1060.

Zhang, L., & Mou, W. (2017). Piloting systems reset path integration systems during position estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(3), 472-491.

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