THEORETICAL/REVIEW

Bayesian decision theory and navigation

Timothy P. McNamara[1] · Xiaoli Chen[2]

Abstract

Spatial navigation is a complex cognitive activity that depends on perception, action, memory, reasoning, and problem-solving. Effective navigation depends on the ability to combine information from multiple spatial cues to estimate one's position and the locations of goals. Spatial cues include landmarks, and other visible features of the environment, and body-based cues generated by self-motion (vestibular, proprioceptive, and efferent information). A number of projects have investigated the extent to which visual cues and body-based cues are combined optimally according to statistical principles. Possible limitations of these investigations are that they have not accounted for navigators' prior experiences with or assumptions about the task environment and have not tested complete decision models. We examine cue combination in spatial navigation from a Bayesian perspective and present the fundamental principles of Bayesian decision theory. We show that a complete Bayesian decision model with an explicit loss function can explain a discrepancy between optimal cue weights and empirical cues weights observed by (Chen et al. *Cognitive Psychology, 95*, 105–144, 2017) and that the use of informative priors to represent cue bias can explain the incongruity between heading variability and heading direction observed by (Zhao and Warren 2015b, *Psychological Science, 26*[6], 915–924). We also discuss (Petzschner and Glasauer's , *Journal of Neuroscience, 31*(47), 17220–17229, 2011) use of priors to explain biases in estimates of linear displacements during visual path integration. We conclude that Bayesian decision theory offers a productive theoretical framework for investigating human spatial navigation and believe that it will lead to a deeper understanding of navigational behaviors.

Keywords Bayesian decision theory · Spatial navigation · Cue integration · Cue combination · Path integration

In 1707, an English fleet under the command of Sir Clowdesley Shovell foundered after running aground on the Isles of Scilly, almost within eyesight of the southwest coast of England. Four ships went down and as many as 2,000 sailors lost their lives. The cause of the wreck, according to experts of the time, was that the seamen did not know their location because of errors in estimating longitude. Although the measurement of latitude had been solved since the time of the ancient Greeks, the measurement of longitude pestered maritime navigators well into the 18th century. Latitude can be determined from the altitude of the sun over the horizon, a measurement readily made with two rods hinged at one end. The measurement of longitude at sea, however, is a problem of measuring time, and it did not admit to practical solution until 1761 with the invention by John Harrison of a portable and reliable clock with a balance spring regulator (Boorstin, 1983).

Spatial navigation is a complex cognitive activity that depends on perception, action, memory, reasoning, and problem-solving (Golledge, 1999). Although people are not often faced with navigational problems as challenging and consequential as Admiral Shovell's, they rely on their navigational skills every day, from activities as common as getting from home to work and back again, to less routine endeavors, such as traveling across town in search of a new, highly rated restaurant. Effective navigation requires combining information from multiple sources or cues to estimate locations. The navigator needs to know their position in space and the locations of goals. One important category of spatial cues in body-powered navigation is body-based cues generated by self-motion, such as signals generated by the vestibular system, the perception of muscle contractions and limb position

✉ Timothy P. McNamara
  t.mcnamara@vanderbilt.edu

[1] Vanderbilt University, PMB 407817, 2301 Vanderbilt Place, Nashville, TN 37240, USA

[2] Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China

Table 1   Standard Model Equations

Mean of optimal combination	$\mu_{Opt} = w_1\mu_1 + w_2\mu_2$ μ_1 and μ_2 are the means of the two single-cue distributions.
Variance of optimal combination	$\sigma^2_{Opt} = \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1} = w_1^2\sigma_1^2 + w_2^2\sigma_2^2$ σ_1^2 and σ_2^2 are the variances of the two single-cue distributions.
Optimal weights on single cues	$w_1 = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2} = \frac{\sigma^2_{Opt}}{\sigma_1^2}$ $w_2 = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2} = \frac{\sigma^2_{Opt}}{\sigma_2^2}$ Note that $w_1 + w_2 = 1$.
Empirical weights on single cues	$\widehat{w}_1 = \frac{d_2}{d_1+d_2}$ $\widehat{w}_2 = \frac{d_1}{d_1+d_2}$ Empirical weights, as defined here, are computed from the centroids of the observed distributions of performance in the single-cue and double-cue conditions. d_i is the Euclidean distance between the mean of the Cue i distribution (μ_i) and the mean of the distribution of observed performance when both cues are available (i.e., Cue 1 + Cue 2 in Figure 1). Empirical weights should be equal to optimal weights in the standard model.

(proprioception), and internal copies of efferent movement signals. Navigation based solely on such body-based cues is often referred to as *path integration*.[1] Path integration is fundamental to maintaining orientation in space and may be crucial to forming a cognitive map of the environment (Chen et al., 2015; Gallistel, 1990; R. F. Wang, 2016). Humans rely heavily on their perceptions of the external world (especially the visual system) to navigate (Foo et al., 2005; Mou & Zhang, 2014; Zhao & Warren, 2015a), and these constitute a second category of spatial cues. These cues differ from those in the first category in that the stimuli are external, in the environment, not internal to the navigator.

Even a quotidian activity such as walking across campus for a meeting in a distant building may depend on combining information from multiple mental representations of the layout of the campus (e.g., Meilinger, 2008), the perception of objects and places along the path (e.g., R. F. Wang & Brockmole, 2003), and the perception of bodily motion during locomotion (e.g., Lindberg & Gärling, 1981). Spatial cues differ in their accuracy and precision, and a given cue's accuracy and precision may vary over time and space. A salient and significant building, for instance, may provide accurate and precise information about location for long periods of time, whereas body-based cues to self-motion can be error-prone, especially as locomotion distance increases (Loomis et al., 1993; Souman et al., 2009). Even GPS signals lose fidelity under some environmental conditions (e.g., dense foliage, urban canyons).

People also differ greatly in their abilities to use various sources of spatial information to learn new environments and to navigate (e.g., Allen, 1999; Gagnon et al., 2018; Hegarty et al., 2006; Ishikawa & Montello, 2006; Montello et al., 1999; Shelton et al., 2013; Weisberg et al., 2014).

Research on cue combination in navigation was stimulated at least in part by Cheng et al.'s (2007) ground-breaking review of work on cue interactions in spatial actions and judgments. Since the publication of that article, a number of projects have examined the extent to which multiple spatial cues are combined in human locomotion, reorientation, or navigation to improve performance relative to the use of single cues alone (Bates & Wolbers, 2014; Chen et al., 2017; Chrastil et al., 2019; Frissen et al., 2011; Nardini et al., 2008; Petrini et al., 2016; Petzschner et al., 2012; Sjolund et al., 2018; Twyman et al., 2018; Xu et al., 2017; L. Zhang et al., 2019; Zhao & Warren, 2015b). The most commonly used experimental paradigm in studies designed to investigate navigation requires participants to walk an outbound path consisting of several legs and then to return to a "home" location using their memories of the waypoints and the path. The paradigm is often referred to as "homing," or when the outbound path has two legs, as "triangle completion" (Loomis et al., 1993). This task is appealing because it is quite natural for people to execute and readily allows spatial cues to be manipulated during the outbound path, the return path, or both. The key questions in experiments that have used this paradigm have been whether the accuracy and precision of participants' performance in returning to the home location is better when multiple spatial cues are available (e.g., visual landmarks and body-based information from self-motion) than when only one is available (e.g., body-based information) and whether any improvements are statistically optimal (defined subsequently). These studies have included experimental conditions in which multiple cues are available but are

---

[1] Historically, path integration referred to navigation using only internal body-based cues (e.g., Mittelstaedt & Mittelstaedt, 1980). In contemporary taxonomies, path integration refers to the process whereby an organism estimates its position in an environment from spatial cues, both internal (e.g., proprioception) and external (e.g., optic flow), that by themselves do not uniquely specify position in that environment (Loomis et al., 1999). This is the way the term is used in this article.

in conflict (e.g., landmarks indicate one return direction whereas body-based cues indicate a different return direction). The purpose of the conflict condition is to assess the relative importance of each type of cue to the navigator.

Past research on cue combination in human navigation has used a model often referred to as the maximum-likelihood estimation (MLE) model to interpret the findings (e.g., Ernst & Banks, 2002; Rohde et al., 2016). We will refer to this model as the "standard model" so as not to confuse this particular application of MLE with the general method of parameter estimation. According to the standard model of cue combination, performance with multiple cues is a weighted average of performance with single cues. Each cue is weighted by its normalized reliability, such that more reliable cues receive greater weight than do less reliable cues. The reliability of a cue is typically estimated by the inverse of the variability of performance using the cue alone. The standard model is statistically optimal in that it minimizes the variance of the estimate in the multiple-cue condition (Cochran, 1937; Oruç et al., 2003). A graphical illustration is presented in Fig. 1. One can interpret the distributions in this figure as distributions of performance (using Cue 1 alone, Cue 2 alone, and both Cue 1 and Cue 2) and the predicted distribution for two cues using the standard model. In this example, Cue 2 is more reliable (less variable) than Cue 1. The predicted distribution for the optimal combination of the two cues is therefore closer to Cue 2 than to Cue 1 because the former is weighted more than the latter. The reliabilities of cues are typically interpreted as measures of sensory-perceptual noise (higher reliability = lower noise), although we prefer to think of them as measures of confidence or certainty (Halberda, 2016). The equations for predicting performance in multiple-cue conditions from single-cue conditions are contained in

Table 1 (the assumptions underpinning these equations are examined subsequently). For an excellent tutorial on the standard model of cue combination and sensory integration, see Rohde et al. (2016). Although the standard model of cue combination is commonly referred to as "Bayesian," it can be derived without invoking Bayes's theorem (Oruç et al., 2003).

As an example, consider Chen et al.'s (2017) project, which investigated people's abilities to combine visual spatial cues and body-based spatial cues in a homing task. Participants walked from a fixed starting location to three successive waypoints in an immersive virtual environment (outbound path) and then walked back to the remembered location of the first waypoint (return path). Stopping points on the return paths constituted participants' estimates of the first waypoint's location. On the outbound path, participants could see the entire environment (e.g., Fig. 2) and walk and turn normally, and hence, they had full access to visual cues and to body-based cues, such as proprioceptive, vestibular, and efferent information. The experimental conditions were distinguished by the events that occurred at the end of the outbound path: In the vision condition, participants were disoriented after reaching the final waypoint, so that when executing the return path, body-based cues were lost and only visual cues were available; in the body-based condition, the visual world was rendered invisible when participants reached the final waypoint, so that when executing the return path, only body-based cues were available; in the combination condition, participants were not disoriented and the world remained visible, so that both visual cues and body-based cues were available; and in the conflict condition, the landmark configuration was surreptitiously rotated by 15° so that the correct location defined by landmarks was different from the one defined by body-based cues. The principal dependent measures were the centroids and the variances of the distributions of stopping points. The equations in Table 1 were used to predict performance in the combination and the conflict conditions (double-cue conditions) from performance in the vision condition and the body-based condition (single-cue
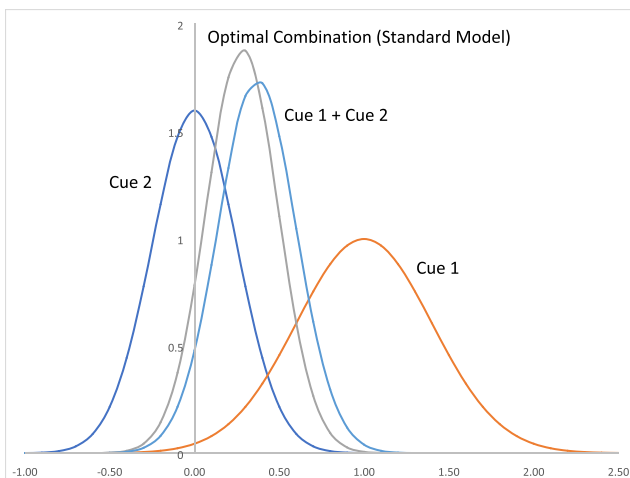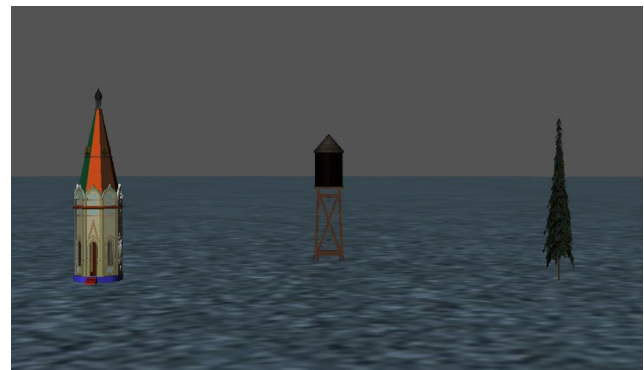

Fig. 1 Illustration of the standard model of cue combination. The figure shows distributions of performance in three cue conditions (Cue 1 alone, Cue 2 alone, both Cue 1 & Cue 2) and the predicted distribution for the double-cue condition using the standard model. All distributions are normalized Gaussians; hence, the heights of the distributions reflect their inverse variances (i.e., larger height = smaller variance = greater reliability)

Fig. 2 Participants' view of the landmarks from the starting location in the virtual environment of Chen et al. (2017). The waypoints appeared in the space between the starting location and the landmarks

conditions). Chen et al. found that in the majority of experimental conditions, observed performance in the double-cue conditions corresponded closely to predicted performance; in other words, navigators combined visual cues and body-based cues optimally according to the standard model.

Similar results have been obtained by other investigators (exceptions will be discussed in subsequent sections of this manuscript). These studies have greatly advanced the scientific understanding of human navigation, but they also raise important questions about the use of spatial cues in wayfinding. Is navigational performance affected by prior knowledge or beliefs about the task environment (e.g., that some locations are more likely to be goals than are others)? In real navigational scenarios, some errors have greater costs than do other errors. Do navigators account for these costs in a principled and quantifiable manner? The purpose of this manuscript is to explore the answers to these questions in the context of Bayesian decision theory (e.g., Berger, 1985; DeGroot, 1970; Robert, 2007).

The remainder of this manuscript is organized as follows: We begin with a summary of Bayesian models of cue combination in the context of human navigation and related behaviors. We then present a more formal treatment of cue combination using Bayesian decision theory. This section is followed by two applications of Bayesian decision theory to published findings that are inconsistent with the standard model of cue combination. We show that these findings can be explained using Bayesian decision-theoretic models. To be clear, our aim in this project is not to present a comprehensive Bayesian model of navigation, but rather to introduce researchers in spatial cognition and navigation, and allied fields, to the potential power of using Bayesian decision theory to investigate navigational behaviors. Our goal is to show that Bayesian decision theory provides a productive framework for future research on navigation, one that we believe will lead to a deeper understanding of navigational behaviors.

Bayesian models of cue combination

The scenarios to be discussed in this manuscript require navigators to estimate the location of a goal given various spatial cues (e.g., visual landmarks, body-based cues from self-motion, memory of the layout of objects). Let L represent the location to be estimated, and S and V represent two spatial cues. In Bayesian terms (Knill et al., 1996; Mamassian et al., 2002; Yuille & Bülthoff, 1996), the problem can be formalized as follows:

$$p(L|S,V) = \frac{p(S,V|L)\,p(L)}{p(S,V)} \quad (1)$$

The term on the left side of the equation, $p(L|S,V)$, is the probability of target locations given information provided by the spatial cues S and V. In Bayesian theory, this term is

referred to as the "posterior." The location corresponding to the mean of this distribution might be selected as the goal on a particular trial of a homing task, for example.

The prior, $p(L)$, is the probability distribution over locations in the absence of information from the spatial cues. The prior formalizes the extent to which the navigator believes or has information that some locations are more probable than others as goals before any information from S and V is available. If the navigator has reason to believe (e.g., previous experience) that some locations are more likely to be goals than are others, $p(L)$ will vary across locations, and hence, will be nonuniform; such priors are referred to as "informative." However, if the navigator believes than any location is equally likely to be a goal, $p(L)$ will be the same for all locations; such a prior is referred to as "uniform" and can be represented as $p(L) = 1$.[2] Priors are commonly used to represent biases of various kinds (e.g., Jacobs, 1999; Mamassian & Landy, 2001; Weiss et al., 2002). For example, Jacobs (1999) used a prior in a Bayesian model of cue integration in depth perception to represent bias to see an object as approximately as deep in 3D as it is wide in the image plane.

The term $p(S,V|L)$ is the likelihood function. The notation suggests that it is the probability of the spatial cues given a location in the environment. However, in a Bayesian analysis, the information in the sensory cues is assumed to be given (e.g., provided by various sensory-perceptual systems); $p(S,V|L)$ is therefore a function of L (for this reason, this term sometimes is written as, $\mathcal{L}(L|S,V)$, where $\mathcal{L}$ stands for likelihood). The likelihood function gives the likelihood of locations given the sensory cues in the absence of prior knowledge.

The term in the denominator of Eq. (1) is a scaling parameter (technically, the sum across L of the product of the likelihood function and the prior), as it ensures that the sum of the values of the posterior across locations is 1. For current purposes, it can be ignored, producing:

$$p(L|S,V) \propto p(S,V|L)\,p(L) \quad (2)$$

(the symbol, $\propto$, means "proportional to"). If $p(S,V|L)$ is factorable (e.g., S and V are conditionally independent given L), we obtain:

$$p(L|S,V) \propto p(S|L)\,p(V|L)\,p(L) \quad (3)$$

This equation can be used to predict observed performance if the prior and the likelihood functions for S and for V are specified mathematically or computationally. Xu et al. (2017)

---

[2] Unfortunately, in Bayesian terminology, a uniform prior is not the same as an uninformative prior; the latter term usually refers to priors that express vague or very general information about a variable and have desirable statistical properties.

and Wang et al. ([2018](#)) use approaches of this kind to model performance in reorientation tasks.

Another way to view a Bayesian model of navigation is as follows: A navigator may have beliefs (which may be incorrect) about the location of the goal before they have any sensory-perceptual information about its location. These beliefs are captured in the prior distribution. The navigator then obtains data about the goal location from their sensory-perceptual systems (e.g., by walking the outbound path in a homing task). These data are used to update the prior, producing the posterior. The posterior reflects the navigator's beliefs about the probability of locations being the goal after having gathered information from the spatial cues.

Applications of cue-combination models to spatial navigation have used a different formulation. Performance in an experimental condition in which participants estimate locations using two cues is predicted from performance in conditions in which participants estimate locations using single cues. Formally, this paradigm involves predicting the dual-cue posterior distribution from two single-cue posterior distributions (Jacobs, [1999](#); Landy et al., [1995](#)):

$$p(L|S, V) \propto p(L|S)p(L|V) \tag{4}$$

An important underlying assumption of this experimental paradigm is that the operations of the single-cue systems in the multiple-cue condition (e.g., navigation with both visual cues and body-based cues available) can be approximated by assessing performance in the single-cue conditions individually (e.g., navigation using only visual cues and navigation using only body-based cues).

Equation ([4](#)) can be derived from Eq. ([3](#)) by the application of Bayes's theorem. Expanding ([3](#)),

$$p(L|S, V) \propto \frac{p(L|S)p(S)}{p_S(L)} \frac{p(L|V)p(V)}{p_V(L)} p(L) \tag{5}$$

where $p_S(L)$ and $p_V(L)$ are scaling parameters for $p(L|S)\,p(S)$ and $p(L|V)\,p(V)$, respectively ($p_S(L)$, $p_V(L)$, and $p(L)$ are distinguished because they may differ in an experimental application, as discussed subsequently). If the scaling parameters are ignored, we obtain:

$$p(L|S, V) \propto p(L|S)\,p(L|V)\,p(S)\,p(V)\,p(L) \tag{6}$$

If one assumes that all priors are uniform (e.g., $p(S) = p(V) = p(L) = 1$), Eq. ([4](#)) is obtained.[3]

The formulation in Eq. ([4](#)) corresponds to a version of "weak-coupling" of sensory-perceptual modules in Yuille and Bülthoff's ([1996](#)) taxonomy and is the approach that has been used in nearly all investigations of cue combination in navigation (exceptions are discussed subsequently). Two

critical assumptions were made to produce Eq. ([4](#)): (a) the cues are conditionally independent and (b) the priors are uniform. A third assumption (c) is needed to yield the equations in Table [1](#)— namely, that the probability distributions are Gaussian (see, e.g., Bromiley, [2018](#)). This formulation is the standard model of cue combination. The standard model can be viewed as a special case of a Bayesian model, although as noted previously, the standard model can be derived without using Bayes's theorem. The assumption that priors are uniform is a limitation of previous applications of the standard model to navigation.

The discussion so far has been abstract. It is natural to ask how a Bayesian model might be realized in a homing task. As the participant walks the outbound path, she accumulates information from multiple sources about her position and the location of the goal (typically specified by instruction at the beginning of the outbound path). Let us assume that goal locations are not selected uniformly (e.g., they are more frequent in one quadrant of the environment) and that the navigator has learned this distribution from past experience in the task. The navigator combines her prior knowledge and information from the spatial cues to estimate the location of the goal. At the end of the outbound path, the navigator walks toward that location. In a Bayesian analysis (Eq. [3](#)), prior knowledge is represented as a distribution over locations in the environment (e.g., $p(L)$, with a peak in one quadrant) and each spatial cue (e.g., body-based cues from self-motion) is associated with a likelihood function (e.g., $p(S|L)$). The product of the prior distribution and the likelihood functions (appropriately scaled) forms the posterior distribution (e.g., $p(L|S, V)$). All estimates and inferences are based on the posterior distribution. For example, the mean of the posterior may constitute the navigator's estimate of the goal location. Because the sensory-perceptual information from the spatial cues and the goal location change from trial to trial, the likelihoods and the posterior also change from trial to trial (Ma, [2019](#)). The reader may be incredulous that Bayesian processes can be implemented at all in perceptual-cognitive systems, let alone on a trial-by-trial basis. There is evidence, however, that the computations needed to implement Bayesian processes may be natural consequences of certain characteristics of neural activity (Ma et al., [2006](#)). These results indicate that Bayesian processes can be implemented in the brain and on appropriate time scales.

Informative priors

In this section, we demonstrate that prior knowledge, as represented by an informative prior distribution, functions like an additional spatial cue in the typical cue combination paradigm. Suppose that an investigator is interested in the integration of external visual cues (V) and body-based self-motion cues (S) in a standard homing task. Following standard

---

[3] Consider that $p(X|L)p(L) \propto p(L|X)$. Assuming that the prior over locations is uniform ($p(L) = 1$), then substituting into Eq. ([3](#)) for S and V produces Eq. ([4](#)).

practice (e.g., Ernst & Banks, 2002; Nardini et al., 2008), performance is assessed under conditions in which participants have only one cue available (visual or self-motion), two consistent cues, or two cues in conflict. Assume that target locations are uniformly distributed across the navigational arena in the single-cue conditions but are more likely to occur in one region of the arena than in others in the double-cue conditions (referring back to Eq. 5, $p_S(L) = p_V(L) = 1$ but $p(L) \neq 1$); assume further that navigators learn this distribution over repeated trials. The mental representation of the distribution of target locations in the double-cue conditions corresponds to an informative prior. In this scenario, the correct formulation of behavior, with sufficient practice, in the double-cue conditions is:

$$p(L\,|S, V) \propto p(L \mid S)\, p(L \mid V)\, p(L). \qquad (7)$$

Observed performance in the double-cue conditions is determined by the posterior on the left side of Eq. (7). The distributions, $p(L \mid S)$ and $p(L \mid V)$, are estimated from observed performance in the single-cue conditions and the prior, $p(L)$, is under experimental control (e.g., the centroid and variability of this distribution are set by the experimenter). If the distributions in Eq. (7) are assumed to be Gaussian, then the formulae for the parameters of interest are (see Table 1 and Bromiley, 2018):

$$\mu_c = w_s \mu_s + w_v \mu_v + w_p \mu_p \qquad (8)$$

$$w_i = \frac{\sigma_c^2}{\sigma_i^2}, \qquad (9)$$

and

$$\sigma_c^2 = \left(\frac{1}{\sigma_s^2} + \frac{1}{\sigma_v^2} + \frac{1}{\sigma_p^2} \right)^{-1}, \qquad (10)$$

where μ_c and σ_c^2 are the mean and the variance of the posterior distribution (i.e., the optimal combination of the two cues and the prior); w_i are the optimal weights on the two cues and the prior; and σ_i^2 are the variances of the component distributions (the subscripts, s, v, and p, refer to the cues S and V, and the prior, respectively). Observed performance in the consistent cue and the conflicting cue conditions can be compared with predicted performance generated by Eqs. (8)–(10). These equations show that the prior functions like an additional spatial cue in this scenario (e.g., compare Eq. 8 to the mean of the optimal combination of two cues in Table 1).

Suppose that participants are insensitive to the manipulation of the probability of target locations in the double-cue conditions. If they nevertheless integrate visual cues and self-motion cues optimally, the correct model of performance will correspond to Eq. (4). In this scenario, the observed

variabilities in the double-cue conditions will be larger than those predicted by Eq. (10):

$$\sigma_C^2 = \left(\frac{1}{\sigma_s^2} + \frac{1}{\sigma_v^2} + \frac{1}{\sigma_p^2} \right)^{-1} < \left(\frac{1}{\sigma_s^2} + \frac{1}{\sigma_v^2} \right)^{-1} = \sigma_{C_{Obs}}^2, \qquad (11)$$

where σ_C^2 is as above, and $\sigma_{C_{Obs}}^2$ is the observed variance in the double-cue conditions if participants optimally combine information from the two spatial cues but ignore the prior. Put another way, observed performance in the double-cue conditions will appear to be suboptimal relative to performance predicted from the single-cue conditions and the prior, although it will be optimal relative to performance predicted from just the single-cue conditions.

Only a handful of projects have investigated effects of prior knowledge or experience on spatial memory and navigation. Huttenlocher et al. (2004) varied the distributions of target locations in a simple spatial memory task (the "dot-and-circle task" of Huttenlocher et al., 1991). They found that participants' memories of location were not affected by these manipulations. Sampaio et al. (2020), however, implemented a version of the dot-and-circle task in virtual reality (VR) and found that participants' memories of the locations of common objects on a table top were biased toward the customary locations of those objects (e.g., pizza on a plate toward the observer's front).

In the domain of navigation-related behaviors, Petzschner and Glasauer have developed elegant Bayesian models of the role of prior experience and of cue combination in visual path integration. In these experiments, participants executed and then reproduced a straight displacement or a single turn in desktop VR; hence, path integration depended solely on visual information. Petzschner and Glasauer (2011) showed that reproductions of distances and turns were sensitive to the statistics of the displacement and turn-angle samples (e.g., bias toward the mean), and that these biases were predicted by a Bayesian model with iteratively updated priors (this experiment will be explored in more detail subsequently). Petzschner et al. (2012) showed further that participants integrated a symbolic cue and prior experience into reproductions of linear displacements and that their behavior was well-predicted by Bayesian models (see also, Petzschner et al., 2015). To our knowledge, these methods and models have not been applied to more complex navigation tasks, such as homing.

Incorporating a loss function

A limitation of prior research on spatial cue combination in human navigation is that studies have not incorporated a complete Bayesian decision model (for examples of Bayesian decision models, see Körding & Wolpert, 2006; Ma, 2012;
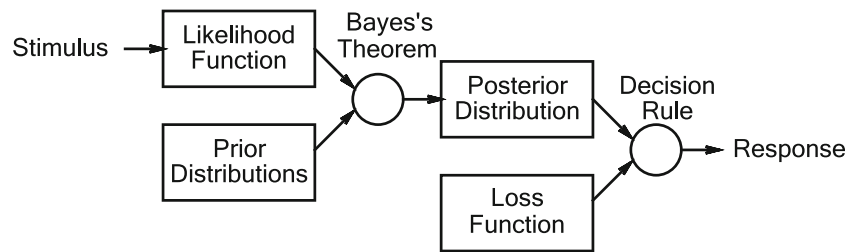
Fig. 3 A complete Bayesian decision model. Adapted from Mamassian et al. (2002, Fig.1.6)

Mamassian et al., 2002; Trommershäuser et al., 2008). One component of such a model (see Fig. 3) estimates the posterior distribution of the relevant variable using Bayesian principles. This component was described previously. In our experiments, this exercise has involved estimating the posterior distribution over locations given visual and self-motion cues (Chen et al., 2017; Sjolund et al., 2018). The other component of the model corresponds to the application of a decision rule that transforms the posterior distribution into an action. Decision rules are determined by the loss function (or equivalently, the gain function) that specifies the consequences of various actions. According to Bayesian decision theory, the action that minimizes expected losses given the posterior distribution should be chosen (Berger, 1985; Robert, 2007). In the context of a navigational problem, selecting an estimate of a location could be an action.

As an example, suppose that one is driving an automobile on the road shown in Fig. 4. This is a navigational problem in which the driver must (among other things) estimate and follow a safe path around the corner. To make the corner, the driver should generally try to stay in the middle of the road. However, a navigational error to the right, into the grass and shrubbery, is likely to be less costly than a navigational error to the left, which could result in jumping the guardrail and tumbling down a cliff. Hence, one might expect the driver to steer to the right of the midpoint of the road, reducing the likelihood of the more costly error, while still keeping the car on the road.

Loss functions can take any form depending on the costs of errors of estimation. The form of the loss function determines the appropriate estimate of the parameter of interest given the

Fig. 4 A driving scenario with an asymmetrical distribution of losses

posterior. For consistency with subsequent theory development, let θ be the parameter to be estimated (e.g., the correct target location) and a be the action (e.g., the actual location selected). Common loss functions are absolute-error loss,

$$L(\theta, a) = |\theta - a|, \tag{12}$$

and squared-error loss,

$$L(\theta, a) = (\theta - a)^2. \tag{13}$$

The use of absolute-error loss entails that the median of the posterior is the appropriate estimate of θ, and the use of squared-error loss entails that the mean of the posterior is the appropriate estimate of θ (e.g., Berger, 1985). Another commonly used estimate of the parameter of the posterior, especially in the vision community, is the *maximum a posteriori* (MAP) estimate, which is the mode of the posterior. The use of a MAP estimate entails that the action must be exactly correct and that the costs of all errors are equal. The MAP estimate is implied by the use of Dirac-delta loss (Yuille & Bülthoff, 1996),

$$L(\theta, a) = -\delta(\theta - a). \tag{14}$$

This loss function, which is based on the Dirac delta function, produces constant loss except if the action is correct ($\theta - a = 0$), where loss is infinitely small.

Returning to the example in Fig. 4, if we assume for simplicity that target locations at any point in time are represented on a single spatial dimension perpendicular to the direction of travel, the use of an asymmetrical absolute-error loss function ("multilinear loss") that assigned twice as much loss to errors to the left as to errors to the right would entail that the 67th percentile (2/3rd fractile) of the posterior is the optimal estimate of the target θ (e.g., Robert, 2007). As the driver winds her way along the highway, the high-risk zone may sometimes be on the left and may sometimes be on the right. The loss function must adapt to these environmental changes, such that the optimal target will "slide" back and forth across the posterior distributions. The guardrail may be missing on the cliff side of a given section of highway, in which case relative losses may increase dramatically (e.g., from 2:1 to 4:1, producing an optimal target of the 4/5th fractile). One might also expect to see substantial individual differences in relative

losses: A beginning driver might be especially prone to hug the safe side of the highway, whereas a highly skilled driver might veer little from the midline.

Previous investigations of spatial cue combination in navigation have employed squared-error loss, although there is no theoretical justification for using this loss function (Petzschner & Glasauer, 2011, examined alternative implicit loss functions in visual path integration). In fact, statistical decision theory argues against such a loss function, as it probably penalizes large errors too much and it increases without bound as error increases (Berger, 1985; Smith, 1988). The advantages of squared-error loss are that it produces tractable mathematical derivations, is familiar to psychological researchers (as least-squares), and yields an estimate of the parameter (viz., the mean of the posterior distribution) that is readily interpretable.

Bayesian decision theory

This section of the manuscript provides a more formal treatment of Bayesian decision theory. Bayesian decision theory is not a model of human performance or of any other natural phenomena; it is a mathematical framework for making decisions. Bayesian decision theory can be used, however, as the theoretical scaffolding for building psychological models. We present two such applications in subsequent sections of this manuscript. Using Ma's (2019) taxonomy, our analysis focuses on the inference, action, and response stages of Bayesian modeling. Our notation follows that of Berger (1985) and Robert (2007).

Formal development

Bayesian analysis is performed by combining information about an unknown parameter before data have been collected—the prior—with the actual or hypothesized distribution of the data given the parameter—the likelihood function—to produce the distribution of the parameter conditional upon the sample data—the posterior distribution (see Fig. 3). All inferences and decisions are based on the posterior distribution.[4]

Formally, the posterior is the conditional distribution of the parameter (θ) given the data (x), and is denoted, $\pi(\theta|x)$. By definition (e.g., Hogg & Craig, 1970), the conditional distribution is the joint distribution of the two variables divided by the marginal distribution of the conditioning variable:

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} \tag{15}$$

[4] The strengths of such an approach to statistical inference, in particular, have been discussed in many venues and will not be repeated here (see, e.g., Gallistel, 2009; Rouder et al., 2009; Vandekerckhove et al., 2018, and accompanying articles).

where $f(x|\theta)$ is the distribution of the data given the parameter (the likelihood function; but as noted previously, in a Bayesian analysis, the data are viewed as fixed, and the parameter varies), $\pi(\theta)$ is the distribution of the parameter before data have been collected (the prior), and $m(x) > 0$ is the marginal distribution of x:

$$m(x) = \int f(x|\theta)\pi(\theta)d\theta. \tag{16}$$

The marginal distribution of x is the predicted observed distribution of the data if the likelihood function and the prior distribution are true.

Equation (15) is equivalent to Eq. (1), but is expressed in the formalisms of Bayesian analysis and decision theory. Paraphrasing a summary of Bayesian analysis presented earlier: Prior to collecting data, we have beliefs or information about the nature of the parameter θ. These beliefs are captured in the prior distribution. We then collect data, which are assumed to be informative about θ. These data are used to update the prior, producing the posterior. The posterior reflects our eventual understanding of the nature of θ after having gathered information about it from the world.

The fundamental principle of Bayesian decision theory is to choose an action that minimizes expected loss given the distribution of θ at the time of decision making. This distribution is the posterior. The *posterior expected loss* of an action a, for the posterior $\pi(\theta|x)$, is defined as,

$$\rho(\pi(\theta|x), a) = \int L(\theta, a)\pi(\theta|x)d\theta. \tag{17}$$

This equation is the expected value of the loss function, $L(\theta, a)$, with respect to the posterior distribution, $\pi(\theta|x)$, and can be interpreted as the average value of the loss function across all values of θ weighted by their likelihood. A *Bayes action*, denoted $\delta^\pi(x)$, is the action that minimizes Eq. (17). In an estimation problem, where one is estimating the value of θ, the Bayes action is to choose the estimate that minimizes posterior expected loss given the data (x). In a navigational scenario, the navigator may be estimating the target location that minimizes expected loss given their estimate of the posterior distribution of locations using the sensory-perceptual information available. Table 2 contains several common loss functions and the associated Bayes actions.

For squared-error loss, the posterior expected loss is,

$$\rho(\pi(\theta|x), a) = \int (\theta - a)^2 \pi(\theta|x)d\theta. \tag{18}$$

The value of a which minimizes Eq. (18) is $E^\pi[\theta]$, the mean of the posterior distribution (e.g., Berger, 1985). Formally, this can be expressed as follows:

Table 2 Common Loss Functions and Their Associated Bayes Actions

Loss Function	Formula	Bayes Action ($\delta^\pi(x)$)
Dirac-delta	$L(\theta, a) = -\delta(\theta - a)$, δ is the Dirac delta function	mode of $\pi(\theta\|x)$
Absolute-error	$L(\theta, a) = \|\theta - a\|$	median of $\pi(\theta\|x)$
Multilinear	$L_{k_1,k_2}(\theta, a) = \begin{cases} k_2(\theta-a) & \text{if } \theta > a \\ k_1(a-\theta) & \text{if } \theta \leq a \end{cases}$	$k_2/(k_2 + k_1)$ fractile of $\pi(\theta\|x)$
Squared-error	$L(\theta, a) = (\theta - a)^2$	mean of $\pi(\theta\|x)$
Weighted squared-error	$L(\theta, a) = w(\theta)(\theta - a)^2$, $w(\theta)$ is nonnegative	General case: $\delta^\pi(x) = \frac{\int \theta \pi(\theta\|x) w(\theta) d\theta}{\int \pi(\theta\|x) w(\theta) d\theta}$ Special case: $w(\theta) = a \exp\left[\frac{-(\theta-b)^2}{2c^2}\right]$ and $\pi(\theta\|x) = N(\mu, \sigma^2)$, then $\delta^\pi(x) = \frac{\sigma^2 b + c^2 \mu}{\sigma^2 + c^2}$

Notes. θ is the parameter of interest. a is the action (e.g., parameter estimate). $\pi(\theta|x)$ is the posterior distribution. $N(\mu, \sigma^2)$ is the normal distribution. Multilinear loss is equivalent to absolute-error loss when $k_1 = k_2$. The Bayes action is the action that minimizes expected loss given the posterior distribution; it does not depend on the form of the posterior (as long as the relevant statistic exists), except for the special case of weighted squared-error loss

$$\delta^\pi(x) = E^\pi[\theta] = \int \theta \pi(\theta|x) d\theta = \int \theta \left[\frac{f(x|\theta)\pi(\theta)}{m(x)}\right] d\theta$$

$$= \frac{\int \theta f(x|\theta)\pi(\theta) d\theta}{\int f(x|\theta)\pi(\theta) d\theta}. \tag{19}$$

It is important to recognize that the Bayes action for squared-error loss is the mean of the posterior distribution regardless of the form of the posterior (assuming the mean exists).

A generalization of squared-error loss is weighted squared-error loss:

$$L(\theta, a) = w(\theta)(\theta - a)^2 \tag{20}$$

where $w(\theta)$ is a nonnegative function of θ. Weighted squared-error loss is a standard loss function in Bayesian analysis (Berger, 1985; Robert, 2007), although we are not aware of any applications of it in the psychological sciences. Weighted squared-error loss captures the notion that the cost of an error of estimation may depend on the value of θ in addition to the difference between the true value of θ and its estimate. Another attractive feature of weighted squared-error loss is that appropriate choices of $w(\theta)$ produce bounded loss functions (i.e., loss does not increase without bound as the magnitude of error increases). The Bayes action for weighted squared-error loss, with weight function $w(\theta)$ and posterior $\pi(\theta|x)$, is:

$$\delta^\pi(x) = \frac{\int \theta \pi(\theta|x) w(\theta) d\theta}{\int \pi(\theta|x) w(\theta) d\theta} \tag{21}$$

(see Appendix A for the derivation). Weighted squared-error loss is used in the next section to interpret some anomalous findings in the cue combination literature.

The reader may see that the weight function in Eq. (21) is analogous to the prior distribution in Eq. (19). Indeed, estimating θ using Eq. (21) and prior $\pi(\theta)$ is equivalent to estimating θ using Eq. (19) and prior $\pi_w(\theta) \propto \pi(\theta)w(\theta)$ (Robert, 2007). In the
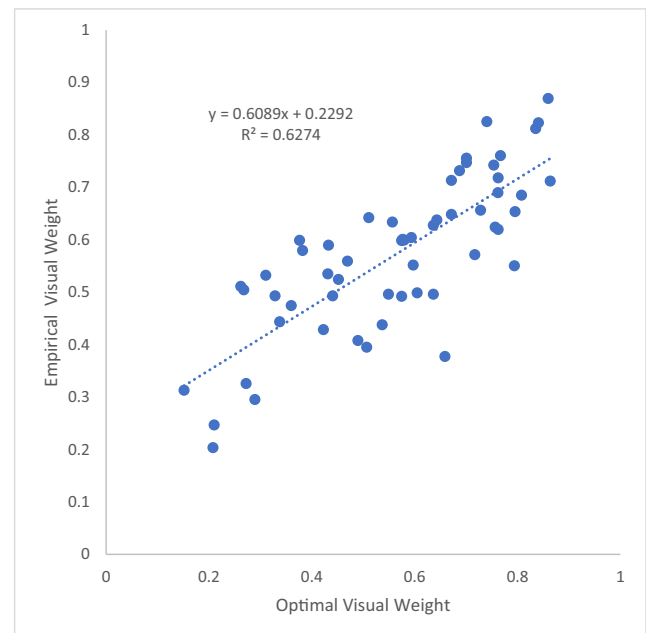

Fig. 5 Relationship between optimal visual weight and empirical visual weight in the conflict condition. Data from Chen et al. (2017) Experiments 1a–1b, 4. Each point corresponds to a single participant ($N = 57$)
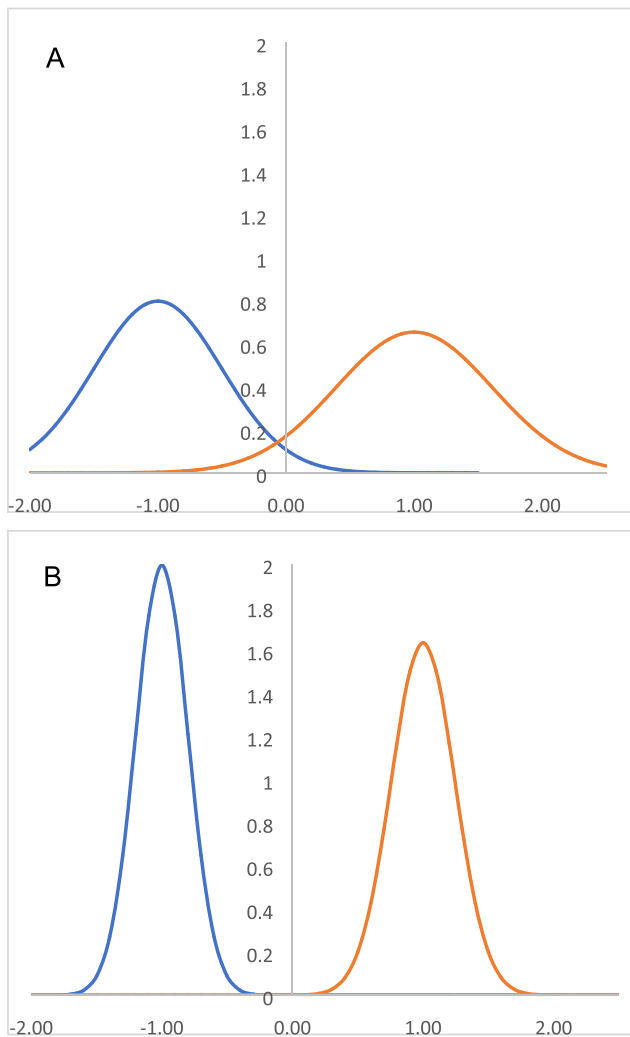
Fig. 6 Hypothetical distributions of single-cue performance. **a** Both cues have high variability and provide poor information about location. **b** Both cues have low variability and provide good information about location. Optimal cue weights are .6 and .4 for the more and the less reliable cue, respectively, in both panels

Fig. 7 Illustration of Bayesian decision theoretic model. Predicted performance is shifted toward the mean of the weight function, which is located halfway between the two single-cue distributions

An example of the application of a loss function

In this section of the manuscript, we show that a cue-combination model founded on Bayesian decision theory can predict findings in the literature that are inconsistent with the standard model. This Bayesian decision-theoretic model is not intended to be a general model of cue combination in navigation. Our goal is to demonstrate the potential power of using the tools of Bayesian decision theory to investigate human navigation.

Figure 5 summarizes key findings from three of Chen et al.'s (2017) experiments. Recall that in those experiments, participants walked an outbound path of three segments, and at the end, were required to walk back to the first waypoint (which was not visible). Four experimental conditions were implemented at the end of the outbound path, three of which are relevant here: In the vision condition, participants only had visual cues available when executing the return path; in the body-based condition, they only had body-based cues to execute the return path; and in the conflict condition, both visual cues and body-based cues were available, but the correct location defined by landmarks was different from the one defined by body-based cues. Figure 5 plots the relationship between optimal weights on visual cues and empirical weights on visual cues (see Table 1) in the conflict condition. Although the magnitude of the relationship is reasonably strong (given the substantial individual differences), the relationship is not consistent with the standard model of cue combination: The intercept and the slope of the regression line differ from 0 and 1, respectively. [5] This relationship is not

standard cue-combination paradigm, if all distributions are Gaussian and a uniform prior is assumed (which is typical), the distribution formed by the product of the weight function and the posterior is identical to that formed by the product of the weight function and the single-cue distributions composing the posterior (see Bromiley, 2018, for relevant derivations); in such an application, the weight function is effectively a prior. In a cognitive process model, however, the properties of a prior and a weight function may differ in important ways. A prior based on preexperimental knowledge is likely to be stable throughout an experiment, whereas a weight function may vary from experimental condition to experimental condition depending on the costs of errors of estimation.
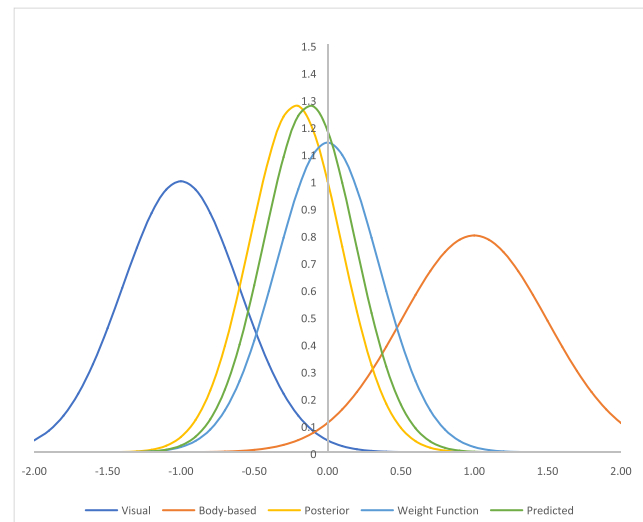
---

[5] The pattern in Fig. 5 is not an artifact of using simple linear regression, which only models error in Y. The same pattern is evident in orthogonal regression, which models error in Y and in X simultaneously: intercept = .167 ± .0443, slope = .719 ± .0707 (standard errors are Jackknife estimates).

Table 3 Model Comparisons and Parameter Estimates

Experimental condition	Bayes factor	r
Exp. 1a: One landmark	1.32×10^{-4}	.627
Exp. 1a: Three landmarks	2.68×10^{-3}	.377
Exp. 1b: Rotation	1.33	.215
Exp. 1b: No rotation	8.73×10^{-2}	.410
Exp. 4: Day 1	1.17×10^{-3}	.501
Exp. 4: Day 2	1.31×10^{-1}	.301

Note. Bayes factor = likelihood(standard model)/likelihood(Bayesian decision theory model). Bayes factors $\leq 3.33 \times 10^{-1}$ favor the Bayesian decision theory model. See Appendix B for details

caused by regression to the mean in estimated distances, as the optimal and empirical cue weights are independent of the correct and reproduced distance. By contrast, the relationship between predicted variability and observed variability in the conflict condition was consistent with the standard model: The slope did not differ from 1 and the intercept did not differ from 0 (see Chen et al., 2017).

The results in Fig. 5 indicate that participants were walking to a target location that was closer to the midpoint between the single-cue conditions than their own single-cue performance would predict. Participants tended to underweight the cue that produced relatively good performance and overweight the cue that produced relatively poor performance. Why might this be? Chen et al. (2017) conjectured that participants were hedging against navigational errors that could result from fully weighting a cue with high relative reliability. A given cue will have high relative reliability if other cues have even worse reliability (see Table 1). Navigators might have been unwilling to trust fully a cue with high relative reliability when other cues were available because their response could still be inaccurate. Figure 6 illustrates the problem. The weight assigned to each cue is the same in Fig. 6a–b, even though in Fig. 6a both cues provide relatively poor information about location, whereas in Fig. 6b, both cues provide relatively good information about location. The assumption is that participants were operating primarily in the regime illustrated in Fig. 6a and underweighted cues with high relative reliability to reduce the risk of making large absolute errors. Chen et al. (2017) conjectured that this trade-off could be captured in a complete Bayesian decision model using an appropriate loss function but did not pursue this notion further.

The reader may question whether loss can even be defined in this task, as navigators did not receive feedback and there were no consequences for errors. Loss in this context is subjective and therefore differs from typical applications of decision rules where loss is determined by the experimenter (see, e.g., Maloney & Zhang, 2010). Although it is customary in the psychological sciences to view loss as objective (e.g., an incorrect response yields a penalty imposed by the experimenter

or the environment), the concept of loss in Bayesian decision theory is fundamentally subjective.[6] In the context of spatial navigation, the subjective nature of loss implies that navigators may make decisions based on their own assessments of the costs of errors. Returning again to the driving scenario in Fig. 4, a risk-averse driver might impose a highly asymmetrical loss function (e.g., errors to the left are 10 times more costly than are errors to the right), whereas a risk-tolerant driver might impose a more symmetrical loss function (e.g., errors to the left are 1.5 times more costly than are errors to the right). One of the aims of the scientist is to determine which subjective loss functions navigators are using.

To explain the pattern of results in Fig. 5, we developed a model of the homing task that incorporated a complete Bayesian decision model of the cognitive processes on each trial. Because the means and the variances of predicted performance in the conflict condition are independent in the model, the modeling of each will be discussed separately.

A standard loss function for implementing bias in decision processes is weighted squared-error loss (e.g., Berger, 1985; Robert, 2007). The apparent bias in return paths toward the midpoint between the locations indicated by landmarks and by self-motion cues can be captured with weighted squared-error loss using a Gaussian weight function:

$$w(\theta) = a \exp\left[\frac{-(\theta-b)^2}{2c^2}\right], \qquad (22)$$

where a is the maximum height of the function, and b and c^2 are its mean and variance, respectively (we use the generic Gaussian notation in the weight function to make clear the sources of parameters in subsequent formulae). In Appendix

---

[6] Skeptical readers are encouraged to see Berger (1985) and Robert (2007) for numerous examples of subjective loss functions. DeGroot (1970) aptly characterized the subjective nature of statistical decision theory in the opening paragraph to his chapter on utility (the complement of loss): "A statistician's subjective probabilities … are numerical representations of his beliefs and information. His *utilities* … are numerical representations of his tastes and preferences" (p. 86, emphasis original).

A, we show that if the posterior distribution $\pi(\theta|x)$ is normal $N(\mu, \sigma^2)$, then the Bayes action [Eq. (21)] is:

$$\delta^\pi(x) = \frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}. \tag{23}$$

Rearranging terms,

$$\delta^\pi(x) = \mu - \left(\frac{\sigma^2}{\sigma^2 + c^2}\right)(\mu - b) \tag{24}$$

In the conflict condition of Chen et al. (2017), participants had to estimate a goal location (i.e., the location of the first waypoint) using information from discrepant spatial cues. According to the standard model of cue combination, the optimal estimate of the goal location is μ, the mean of the posterior distribution constructed from the combination of the two spatial cues. According to the decision-theoretic model embodied in Eq. 24, the optimal estimate of the goal location is the mean of the posterior distribution (μ) adjusted by the difference between the mean of the posterior and the mean of the Gaussian weight function (b). The effect of this adjustment is to "pull" the optimal estimate toward the value b. The magnitude of the adjustment depends on the relative variances of the posterior and the weight function. For fixed σ^2, as c^2 increases in magnitude (i.e., as the weight function becomes flatter), the adjustment approaches zero and the optimal estimate approaches μ; whereas as c^2 decreases in magnitude (i.e., the weight function becomes more peaked), the adjustment approaches the difference between μ and b, and the optimal estimate approaches b. The model is illustrated in Fig. 7. The weight function operates like an informative prior in the decision process (Berger, 1985, p. 161).

According to the model, observed performance in the conflict condition reflected participants' use of the optimal estimate of the target location using weighted squared-error loss rather than just squared-error loss. The mean of the resulting distribution of performance should therefore be:

$$\mu_{Conflict} = \mu - \left(\frac{\sigma^2}{\sigma^2 + c^2}\right)(\mu - b), \tag{25}$$

where μ and σ^2 are the mean and the variance, respectively, of the optimal combination of the two cues using the equations in Table 1; and b and c^2 are the corresponding parameters of the weight function.

Chen et al.'s (2017) experiments were not designed to test models for individual participants, or even to test models with multiple free parameters, so several simplifying assumptions were made to ensure that we had sufficient power for model comparisons. We assumed that the ratio,

$$r = \frac{\sigma^2}{\sigma^2 + c^2} \tag{26}$$

was constant and the same for all participants within an experimental condition. This ratio is the relative reliability of the weight function. The value of b varied across participants and was set equal to the midpoint between mean performance in the vision condition and mean performance in the body-based condition for each participant. With these assumptions, only one parameter had to be estimated, r:

$$\mu_{Conflict} = \mu - r(\mu - b). \tag{27}$$

This alternative model was compared with the null model:

$$\mu_{Conflict} = \mu. \tag{28}$$

The null model is the standard model of cue combination, which is defined by the equations in Table 1. [7]

The decision-theoretic model in Eq. (27) was fit to data from Experiments 1a, 1b, and Experiment 4 of Chen et al. (2017). (See Appendix B for details.) These experiments were selected because they used similar methods. Each of these experiments comprised two within-participant conditions, and these were modeled separately. All six of these experimental conditions used the homing task described previously: Participants walked from a fixed starting location to three waypoints in sequence and then had to return to the first waypoint using their memories of its location. These experiments manipulated the number of landmarks (1 vs. 3), the integrity of body-based cues (at the end of the outbound path, participants rotated 270° and back or did not rotate), and experience in the task (Day 1 vs. Day 2).

The results are summarized in Table 3. The Bayes factors are the likelihoods of the standard model (the null hypothesis) relative to the likelihoods of the decision-theoretic model (the alternative hypothesis) given the data. Bayes factors are interpreted directly without reference to statistical tables; for example, a Bayes factor of 0.10 means that the alternative is 10 times more likely than the null given the data (Kass & Raftery, 1995). We consider Bayes factors greater than or equal to 3 as evidence in favor of the standard model and Bayes factors less than or equal to 1/3 as evidence in favor of the decision-theoretic model; Bayes factors between these bounds favor neither model. In five of six data sets, the decision-theoretic model better predicted mean performance than did the standard cue combination model. Importantly, the

---

[7] We also estimated a common value of c^2 across participants, allowing the ratio r to vary (σ^2 is estimated from the observed variances in the single-cue conditions). These fits were usually slightly worse than those produced by estimating r, although the differences were not so large as to justify a choice between the two. To our minds, a fixed ratio across participants (allowing σ^2 and c^2 to vary) seems more plausible than a fixed variance of the weight function.

computations of the model likelihoods penalized the decision-theoretic model for having one free parameter (see Appendix B). In Experiments 1a and 4, the estimated values of r indicate that the influence of the weight function is smaller (r decreases) with greater perceptual certainty (1 vs. 3 landmarks) and more practice in the task (Day 1 vs. Day 2); that is, participants placed greater trust in the information from their sensory-perceptual systems as the quality of that information improved. A parallel effect does not seem to hold in

Experiment 1b, but the model is not supported in the condition in which body-based cues were disrupted by body rotation.

According to the decision-theoretic model, the empirical weights computed using the formula in Table 1 are not pure measures of cue weighting because of the influence of the weight function. It is possible, however, to recover the weights through the model. In Appendix C, we show that if the decision-theoretic model is an accurate characterization of performance, then the empirical weight on the visual cue should be:

$$w'_V = \widehat{w}_V - \frac{r}{d_{V-B}}(\mu-b), \qquad (29)$$

where $\widehat{w}_V$ is the empirical weight on the visual cue using the formula in Table 1; d_{V-B} is the distance between the means of the vision condition and body-based condition; and r, μ, and b are as defined previously. Figure 8 plots the relationships between empirical visual weights and optimal visual weights for one of the datasets modeled previously (Experiment 4, Day 1; corresponding figures for the remaining datasets can be found in Appendix C). The parameter r in Eq. (29) was set equal to the value estimated in the model fitting (.501 for this data set).

The two panels in Fig. 8 are representative of the pattern evinced in each data set: The regressions of empirical weights on optimal weights (e.g., Fig. 8a) yielded intercepts greater than zero and slopes less than one, whereas the regressions of model empirical weights (Eq. 29) on optimal weights (e.g., Fig. 8b) yielded intercepts that did not differ significantly from zero and slopes that did not differ significantly from one. The latter pattern is predicted by Bayesian models of cue combination (under assumptions identified previously). It appears, then, that a fully specified Bayesian decision model can explain the problematic relationship between optimal weights and empirical weights obtained by Chen et al. (2017).

We also assessed the extent to which this decision-theoretic model could reproduce the observed variability in the conflict condition. We computed predictions for a process model of the task (see Appendix D for details). According to the model, on each trial of the homing task, a posterior distribution of target locations is constructed from sensory-perceptual information provided by visual and body-based systems. The mean of the posterior is used as the preliminary estimate of the target's location for that trial. Information from the visual and the body-based systems is also used to estimate the value of the weight function. This feature of the model provides a mechanism for the posterior and the weight function to be correlated (Oruç et al., 2003). Consistent with the modeling described in Appendix B, the average of two single-cue estimates is used as the mean of the weight function (b), reflecting a bias to walk toward the midpoint between the locations indicated by the two cues. The value of the weight function is obtained by sampling from the appropriate Gaussian
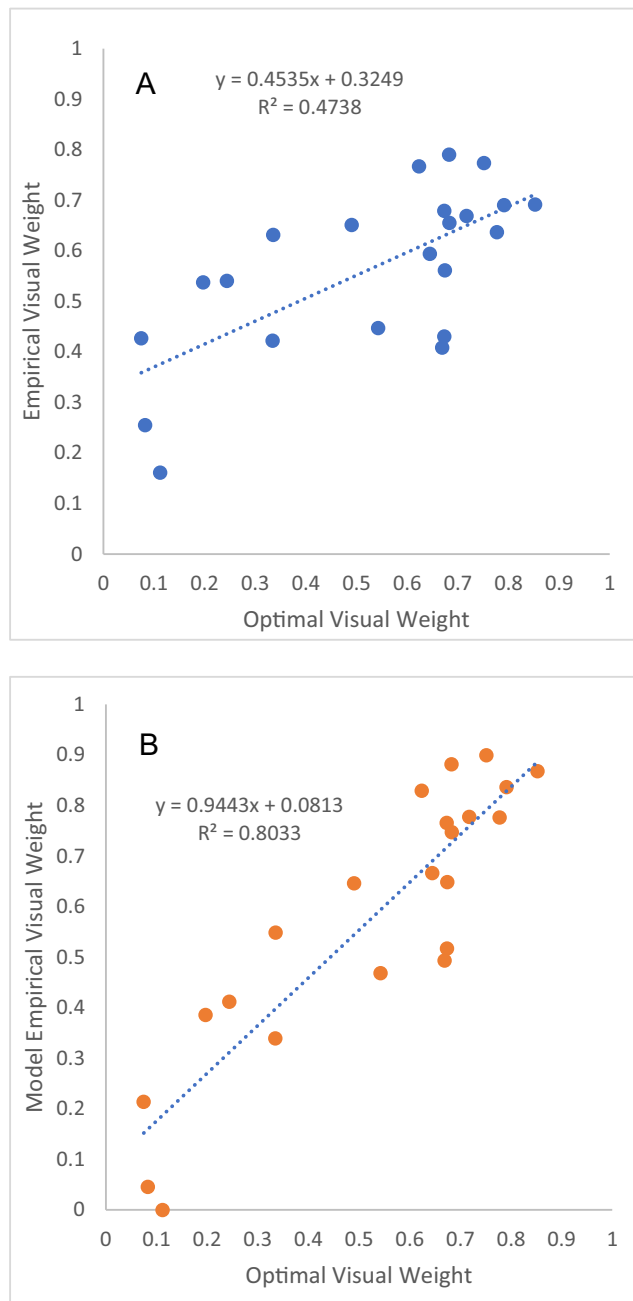
Fig. 8 Relationships between empirical visual weights and optimal visual weights for Chen et al. (2017) Experiment 4, Day 1. **a** Empirical visual weights as in Table 1. **b** Model empirical weights as discussed in text
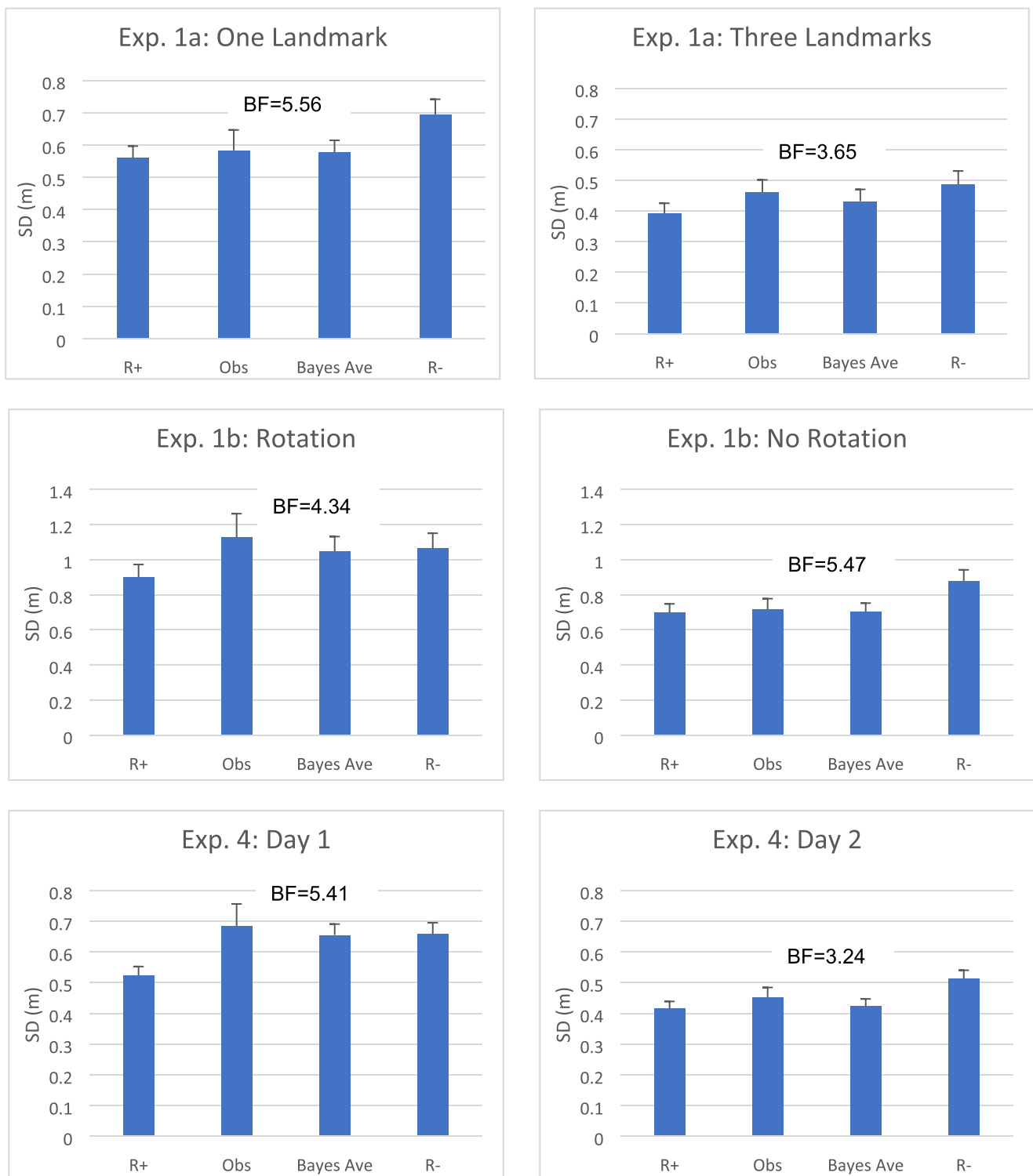
Fig. 9 Predicted and observed mean standard deviations (in meters) for six experiments. Resampling model (R^+), nonresampling model (R^-), observed performance in conflict condition (Obs), Bayes average of R^+ and R^- (Bayes Ave). Obs is bracketed by R^+ and R^- (± 1 SE), consistent with a probability mixture of two processes. Bayes factors (BF) favor the null hypothesis of no difference between Obs and Bayes Ave. Error bars are 1 SE of mean

distribution. The mean of the posterior and the value of the weight function are then combined using Eq. (25) to obtain the Bayes action, which is the location to which the navigator

walks. We assumed that the relative reliability of the weight function, which determines the magnitude of bias, is stable for an experimental condition and reflects properties of the task

Table 4 Comparisons of Observed vs. Predicted Data (Data vs. BDT) and Models (Standard Model vs. BDT)

Experimental condition	Bayes factor[1] (Data vs. BDT)	Bayes factor[2] (Standard model vs. BDT)
Exp. 1a: One landmark	5.56	1.92
Exp. 1a: Three landmarks	3.65	9.91×10^{-2}
Exp. 1b: Rotation	4.34	5.72×10^{-1}
Exp. 1b: No rotation	5.47	3.03×10^{-1}
Exp. 4: Day 1	5.41	7.31×10^{-2}
Exp. 4: Day 2	3.24	8.43×10^{-1}

Note. [1] Bayes factors (JZS, $r = 1$; Rouder et al., 2009) test the null hypothesis of no difference between the observed variability and the predicted Bayes average in the conflict condition. The null hypothesis is supported in all experiments (all Bayes factors ≥ 3). [2] Bayes factor = likelihood(standard model)/ likelihood(BDT). BDT = Bayesian decision theory model. The standard model is not favored in any experiment (no Bayes factor ≥ 3); the BDT is favored in three experiments (Bayes factors $\leq 3.33 \times 10^{-1}$).

environment (e.g., the number of landmarks, experience with the homing task). These values were the same as those estimated previously (values of r reported in Table 3).

It may seem as though the model allows the navigator to change the loss function after perceiving the environment; it is as if the player of a card game were allowed to change the rules after seeing their cards. Such a process is not permitted in Bayesian decision theory. In fact, the loss function remains the same from trial to trial but one of its parameters changes as a result of perception; this parameter is the mean of the weight function. This situation is analogous to the driving scenario discussed previously: Just as the high-risk zone and the relative costs of errors could change over time as a result of perception, so too the navigator in our experiments is, according to the model, learning on each trial which areas of space (values of θ) are those in which errors are especially important. The decision rule can be characterized as "to be safe, err toward the middle of the two cues when they differ," but the location of the "middle" changes from trial to trial. The influence of the weight function, which is determined by its relative reliability (the parameter r), is not affected by trial-to-trial sensory-perceptual information. Perceptual uncertainty in estimating the parameter b is incorporated into the model, as is appropriate (see Appendix D).

We envisioned two versions of this model: In one version, the posterior and the weight function were computed from the same sample of the single-cue systems (R^-, for no resampling), introducing a positive correlation between the two. The effect of a correlation between decision components is to increase the variability of their combination (Oruç et al., 2003). In the other version, the posterior and the weight function were computed from independent samples of the single-cue systems and hence were uncorrelated (R^+, for resampling). This version of the model embodies the conjecture that failures of attention or working memory may necessitate resampling of information from single-cue systems. Predicted variability of the Bayes action for both versions of the model can be computed analytically for individual participants.

Because both versions of the model could be true, even for a given participant on different trials, we computed the Bayes average of their predictions (Claeskens & Hjort, 2008; Raftery, 1995; Wasserman, 2000) to compare to observed performance. Figure 9 contains predictions for the two model versions, their Bayes average, and observed performance in the conflict condition for each of the six experimental conditions. We want to emphasize that no additional parameters were estimated to obtain the results in Fig. 9. The fits to variability are a form of cross-validation of the model.

In each experimental condition, the two versions of the model bracket observed performance: Performance in the conflict condition is never less than the prediction of the resampling model (R^+) and never exceeds the prediction of the non-resampling model (R^-) by more than one standard error. This pattern is consistent with our conjecture that mean performance is a probability mixture of two processes. Table 4 contains statistical comparisons using Bayes factors between the Bayes average and observed performance in the conflict condition and between the Bayes average and the standard model. These results show that the decision-theoretic model predictions do not differ from the observed results (in Column 2, all Bayes factors ≥ 3) and are never convincingly worse than the standard model predictions (in Column 3, all Bayes factors < 3; all but one < 1), and in three cases, are convincingly better (Bayes factors $\leq 1/3$). Across all six experiments ($N = 115$), the standard model is convincingly worse than the decision-theoretic model (Bayes factor = 2.5×10^{-3}). The squared correlation between observed and predicted variability across the six experimental conditions is .995. Combined with the results summarized in Table 3, the decision-theoretic model predicts performance in the homing task better than does the standard model.

In summary, Chen et al.'s (2017) findings were mostly consistent with predictions of the standard model except for the linear relation between optimal cue weights (computed from the variances of the component distributions) and empirical cue weights (computed from the means of those

distributions; see Table 1). This relation indicated that participants underweighted reliable cues and overweighted unreliable cues when executing the return paths. According to the decision-theoretic model, the visual cues and the body-based cues were combined optimally, but participants did not fully trust the resulting distribution of estimates of the home location because they were aware of the intrinsic variabilities, and hence, inaccuracies, of the component spatial cue systems. Relying heavily on a cue with high relative reliability could still result in large homing errors. This concern led participants to walk toward a subjective "safe zone" between the two disparate spatial cues. This bias in return paths was implemented via the loss function that defined the decision rule.

An alternative non-Bayesian explanation of this pattern of results is that navigators alternated across trials between the location indicated by the posterior (i.e., the standard model prediction) and the location corresponding to the average of the two cues. Mixture models of this type have been investigated in several perceptual tasks (e.g., Laquitaine & Gardner, 2018; Wozny et al., 2010). A key prediction of such models is that response distributions are bimodal. There was no evidence of bimodality in Chen et al.'s (2017) data but the disparity between cues (15°) was probably not sufficiently large to reveal bimodality.

For completeness, we tested this non-Bayesian mixture model by first fitting it to mean performance, allowing the mixture probability to be a free parameter, and then to the variability of performance using the same mixture probabilities. The mixture model and decision-theoretic model make identical predictions for mean performance (because of the way that we implemented the decision-theoretic model). Letting p and $1-p$ be the mixture probabilities for the posterior mean and the mean of the two cues, respectively, then $1-p$ is equal to the relative reliability of the weight function (r in Table 3). This equality is purely arithmetic and has no conceptual interpretation. The two models make different predictions, however, for the variability of performance. Across all six experiments, the mixture model is convincingly worse than the decision-theoretic model (Bayes factor = 5.8 × 10^{-5}). The squared correlation between observed and predicted variability across the six experimental conditions is .902 (cf. .995 for the decision-theoretic model).

The modeling of variances exemplifies an important category of situations in which the response distribution is affected by trial-to-trial variation in sensory-perceptual systems and therefore must be compared to an appropriate model-based distribution of performance (Ma, 2019). This model is Bayesian but it is not optimal (e.g., Norton et al., 2019), as the trial-to-trial variation in sensory-perceptual systems does not affect the relative reliability of the weight function (and hence the magnitude of bias) in each experimental condition. As noted previously, we assumed that the relative reliability of the weight function is stable for an experimental condition and

is determined by the nature of the task environment. This assumption could be tested in appropriately designed experiments. One can construct an optimal version of the model by sampling the value of the weight function from a distribution with a constant mean across trials (in effect, ignoring the sensory-perceptual information on a given trial). The weakness of such a model is that it does not reflect the cognitive processes that we hypothesize are operating on each trial; in addition, such a model predicts lower variability in the conflict condition than was observed.[8]

Examples of the application of an informative prior

In the previous section of the manuscript, we showed that experimental findings that were problematical for the standard model of cue combination could be explained by specifying an explicit loss function in a complete Bayesian decision model. In this section, we turn our attention to the other essential component of a Bayesian model, the prior distribution. The first example uses a prior to represent knowledge of properties of the stimuli, and the second uses a prior to represent a cognitive bias.

Regression and range effects in path integration Petzschner and Glasauer (2011) developed Bayesian models to explain biases in estimates of linear and angular displacements during visual path integration (see also, Lakshminarasimhan et al., 2018). Only the distance estimation data will be summarized here. Participants used a joystick to traverse a straight path in a virtual desert environment. When the sample distance for the trial had been reached, movement was terminated automatically. Participants then had to reproduce the just experienced distance using the same locomotion mode. Velocity was constant during production and reproduction but jittered between the two to eliminate temporal cues to distance. Three overlapping ranges of distances were experienced in separate sessions of 200 trials each (1–10 m, 5–14 m, 10–19 m). Participants were blind to the distance traveled and estimated on each trial, and did not receive feedback. Petzschner and Glasauer's findings are reproduced in Fig. 10a. The key results are that estimates within each range were biased toward the mean of the range (e.g., shorter distances overestimated, longer distances underestimated) and the magnitude of bias increased with the average magnitude of displacement (i.e., smaller for the 1–10-m range than for the 10–19-m range).

The Bayesian model is illustrated in Fig. 10b. The model assumes that participants learn the average range of the distances experienced within each session and represent this knowledge as

---

[8] An optimal model can also be constructed by allowing the relative reliability of the weight function (r) to depend on the variability of the visual and body-based systems. The problem with this model is that the magnitude of bias is inversely related to the variability of the sensory-perceptual systems, which is inconsistent with the proposed role of the weight function in the decision model.
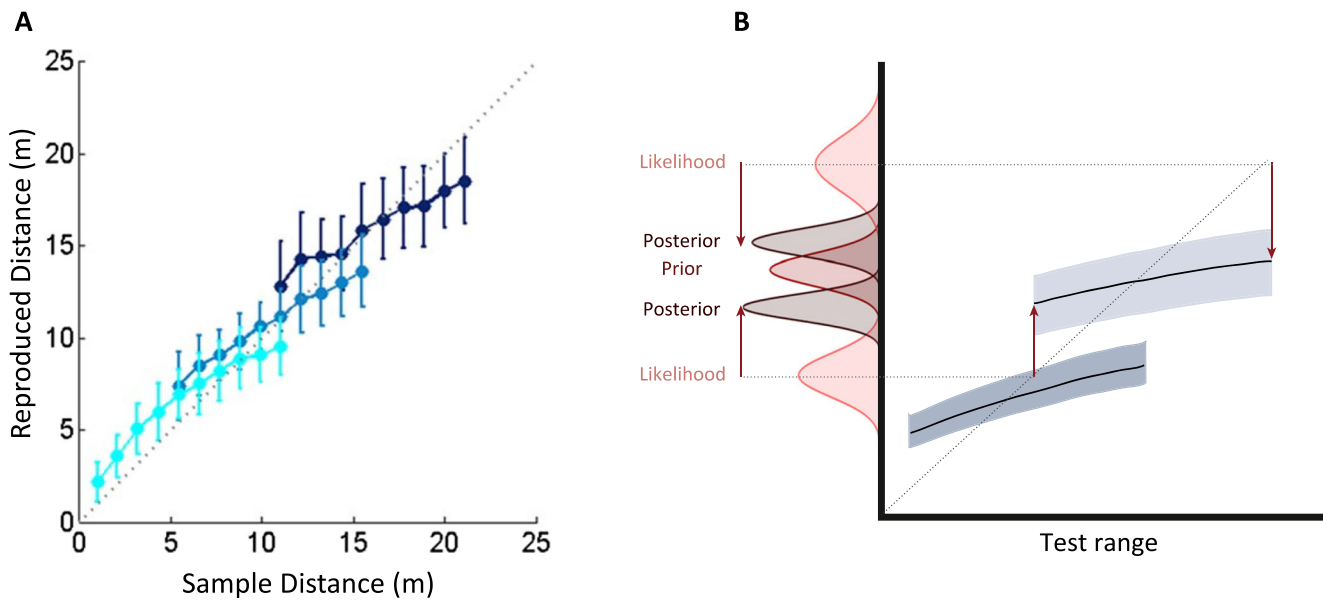
Fig. 10 Distance estimation results and Bayesian model from Petzschner and Glasauer (2011). **a** Distance estimates from visual path integration show characteristic patterns of regression to the mean and range effects (greater regression to the mean with increasing magnitude of the distance range). Dotted line indicates equality between reproduced and sample distances. Adapted from Petzschner and Glasauer (2011), Fig. 3. **b** Bayesian model posits an informative prior centered on the middle of each distance range (only the upper range distributions are illustrated). Likelihood functions are unbiased. Bayesian combination of the prior with each likelihood function yields posterior distributions from which distance estimates are obtained. Posterior distributions are closer to the prior than are likelihoods, and the magnitude of bias (vertical arrows) is larger for the larger distance (a result of representing internal estimates of distances on a logarithmic scale, such that variances increase with magnitude). Adapted from Petzschner et al. (2015), Fig. 1

an informative prior. In the example, the prior is centered on the mean of the higher range of distances. Sensory-perceptual likelihood functions are illustrated for the minimum and the maximum distances within that range; they are assumed to be unbiased (i.e., means = actual distance). The scaled products of the likelihood functions and the prior yield the posterior distributions (e.g., Eq. 15). The estimated distances, which are obtained from the posteriors, are closer to the mean of the range, producing overestimation of the shorter distance and underestimation of the longer distance. The amount of bias is larger for the longer distance in the range because its likelihood function has greater variance than does the likelihood function for the shorter distance. This model predicted the results of the experiment quite well ($R^2 = .97$).[9] The model included a parameter sensitive to the loss functions potentially used by participants (Dirac-delta, absolute-error, squared-error; see Table 2). Values of the fitted parameter indicated that cohorts of about the same size used each loss function.

The biases produced by the prior in this model are formally analogous to those produced by the loss function in our decision-theoretic model of homing. In Petzschner and Glasauer's (2011) model, the prior biases the estimate away from the likelihood function toward the mean of the prior; in our decision-theoretic

model, the weight function (which operates like a prior in the decision stage) biases the estimate away from the posterior distribution toward the mean of the weight function. Mathematically, the computations are identical; conceptually and theoretically, however, the components of the models correspond to different entities computed at different stages of cognitive processing.

Dissociation between homing variability and homing direction in path integration Zhao and Warren (2015b) employed a homing task very similar to that used by Chen et al. (2017) but varied the disparity between visual cues and body-based cues in the conflict conditions over a much wider range (15°, 30°, 45°, 90°, 115°, and 135°). They obtained two major findings:

1. The observed variability of homing direction was consistent with predictions of the standard model in the combined cue condition and in most of the conflict conditions, even for large cue disparities.
2. The observed mean of homing direction was not consistent with the standard model of cue combination and indicated that participants relied on single cues to establish a walking direction. One group of participants ($N = 11$) walked in the direction indicated by the visual landmarks, until the discrepancy between visual cues and body-based cues exceeded 90°, at which point some participants continued to follow the visual cues and others switched to body-based cues. A

---

[9] Petzschner and Glasauer also developed a model in which the prior was iteratively updated across trials using the distance just experienced. The iterative model did not predict the results any better than the model described in the text, but had some qualitative strengths, and hence, was preferred by the authors.

second group of participants ($N = 7$) walked in the direction indicated by body-based cues in all conflict conditions, apparently ignoring the visual cues entirely.

There was no evidence in Zhao and Warren's (2015b) results of trial-to-trial cue-switching (e.g., Laquitaine & Gardner, 2018) or of model-based decision strategies (e.g., Wozny et al., 2010). In particular, bimodality diagnostic of those processes was not evident (see Fig. 3 of Zhao & Warren, 2015b). Heading direction was determined by one cue, whereas heading variability was consistent with the optimal combination of visual and body-based cues (as defined in Table 1) at nearly all cue disparities.

Although Zhao and Warren's results are inconsistent with the standard model of cue combination, they can be explained in a Bayesian model with informative priors. The conflict conditions in their experiment were blocked by increasing cue disparity and always occurred after the single-cue and combined-cue conditions. Participants apparently learned early in the 15° disparity block that the two cues were incongruent, as even in this condition, participants' homing direction was determined by only one cue.[10] It is likely that once participants learned that the cues were incongruent they attended to this feature of the environment as the disparity increased across blocks of conflict trials.

Our conjecture is that participants in Zhao and Warren's (2015b) experiment treated the conflict condition as a single-cue condition, but with an informative prior centered on one of the cues. This prior represented the bias to use only one cue in the presence of conflicting information. Integration of a single-cue distribution and an informative prior will produce a combined distribution with smaller variability than the two component distributions, just as integration of two single-cue distributions produces a combined distribution with reduced variability (see Table 1). The prior need only have variability similar to that of the ignored cue to produce a reduction in variability in the conflict condition equal to that predicted by the optimal combination of the single-cue distributions. The mean, however, will be equal to the shared mean of the single-cue distribution and the prior distribution. This effect is illustrated in Fig. 11. Additional processes are needed to predict the shift from the visual cue to the body-based cue when the disparity between the two was sufficiently large (none of the participants switched from body-based cues to visual cues), but Bayesian models of similar phenomena have been developed for perceptual localization tasks (Körding et al., 2007; Roach et al., 2006) and could be applied to homing tasks.
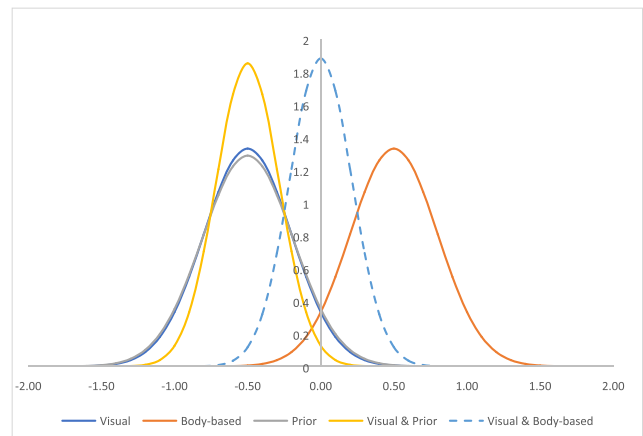

Fig. 11 Idealized distributions of performance for two single-cue conditions, a prior distribution centered on the visual cue, and the predicted distribution for the optimal combination of the visual cue and the prior. The predicted distribution for the optimal combination of the visual cue and the body-based cue is included for comparison. The variance of the prior was set to be slightly greater than the variance of the visual cue so that both distributions would be visible in the figure

It may seem strange or even magical that the reliability of the prior would match the reliability of the ignored cue. We hypothesize that the bias to rely on one spatial cue and to ignore the other spatial cue when they conflict is a top-down, cognitive bias. This bias must be of sufficient magnitude to attenuate the influence of the ignored cue (through facilitation of the favored cue, inhibition of the ignored cue, or a combination). Strength of bias in a Bayesian model is represented by reliability (or its inverse, variability); hence, the reliability of the prior—which represents the bias to use the favored cue—should approximately match the reliability of the ignored cue. The top-down cognitive bias that we are proposing follows from Chen et al.'s (2017, Figure 14) model of cue weighting.

Consider the Stroop task as an analogy. To name the color of the words efficiently, one must focus attention on color naming and inhibit the currently inappropriate response of reading the words. The magnitude of inhibition is not related to color naming but to the automaticity of reading (low for beginning readers, high for adult readers). Models of conflict tasks, such as the Simon task and the Eriksen flanker task, have incorporated processes of the kind we are proposing (Jo et al., 2021; Ridderinkhof, 2002; although these are not Bayesian models).

Zhao and Warren's (2015b) results naturally raise questions about the differences between participants who attended to visual cues (until the disparity between visual and body-based cues became sufficiently large) and participants who only attended to body-based cues. Chen et al. (2017, Experiment 4) found that participants who relied relatively more on body-based cues and performed relatively better with body-based cues also had higher scores on a standardized test of mental rotation ($\bar{r} = .50$). Given that mental rotation correlates positively with several skills related to navigation

---

[10] Chen et al. (2017) and Nardini et al. (2008) found that navigators integrated cues optimally in the 15° disparity condition. A potentially important difference between their experiments and Zhao and Warren's is that all cue conditions were intermixed within blocks over the course of the experiment in Chen et al.'s and Nardini et al.'s projects.

(Hegarty et al., 2006; Weisberg et al., 2014), one would predict that participants in Zhao and Warren's experiment who attended to body-based cues ("distal-path-integration group") would be the better navigators (*ceteris paribus*). Participants in Zhao and Warren's experiment might have sorted themselves based on their perceived navigational skill (Hegarty et al., 2002) into followers of landmarks or followers of their own body-based sense of direction.

Examples of nonoptimal cue combination

The results of several investigations of cue combination in navigation have shown that participants do not combine cues optimally, or even at all in some circumstances. Petrini et al. (2016) examined cue combination in a path reproduction task. Participants experienced an outbound path with vision only, self-motion only, or vision and self-motion, and then reproduced the path by walking in darkness. They found that adults did not integrate visual and self-motion cues in this task, whereas children did so optimally. Tcheang et al. (2011) also did not find evidence of cue integration in a homing task, and like Petrini et al.'s study, one or more cues were available during the outbound path but only one cue (self-motion information) was available during testing. Newman and McNamara (2021) have shown that in a homing task, adults do not integrate visual and self-motion cues until they execute the return path; if the same principle holds in the paradigms used by Petrini et al. and Tcheang et al., the nature of the experimental conditions could have precluded combining cues because only one cue was available at the time of testing.

Other findings may be irreconcilable with Bayesian models. For example, Chrastil et al. (2019) found little evidence that cues were combined at all, let alone combined optimally, in a novel homing task (see also, L. Zhang et al., 2019). For reasons described subsequently, we view such findings as expected and informative.

General discussion

The aim of this project was to apply Bayesian decision theory to a complex cognitive activity, spatial navigation. Our particular interest was to understand how navigators combine spatial cues to estimate the location of a target location, or goal. According to a Bayesian analysis of this problem, navigators combine any prior knowledge they may have about the distribution of goals in the environment with spatial information provided by sensory-perceptual systems to construct a distribution of possible goal locations conditioned on the sensory-perceptual information. This distribution, which is known formally as the posterior, is the basis of all inferences and decisions about the goal. A decision rule is then used to transform the posterior distribution into the choice of a goal location.

Decision rules are determined by the loss function that specifies the consequences of various actions. According to Bayesian decision theory, the location that minimizes expected losses given the posterior distribution should be chosen as the goal.

Several projects have investigated spatial cue combination in navigation (cited previously), but all of them have assumed that prior knowledge, if it existed, did not privilege some locations over others (i.e., prior knowledge was uniform across locations) and none used an explicit decision rule. We showed that Bayesian decision theoretic models were able to predict experimental results that were inconsistent with predictions of the standard model. The use of an explicit decision rule explained discrepancies between optimal and empirical cue weights in Chen et al.'s (2017) experiments and the use of informative priors explained the incongruity between heading variability and heading direction in Zhao and Warren's (2015b) experiment.

Incorporating informative priors, specifying a complete decision model, or allowing nonindependence between stimuli or decision processes gives Bayesian decision theory considerable explanatory power. To date, investigations of human spatial navigation have not capitalized on this explanatory power. We are not aware of any investigations of human navigation that have manipulated the probability of locations being goals, prior experiences with spatial cues, or the costs of navigational decisions. One study that comes very close is that of H. Zhang et al. (2010). They examined optimality of route planning in an economic navigation task where participants traced paths on a computer screen with their fingers. Versions of that experiment that involved physical locomotion could be implemented in immersive VR using readily available technology. These are promising and completely unexplored areas of experimental and theoretical investigation.

As stated previously, our aim in this manuscript is to demonstrate the potential utility of using Bayesian decision theory to investigate navigational behaviors. The baseline model to which the Bayesian model has been compared is the standard model of cue combination. We believe that this model is the appropriate reference point in the current context because it has been used to design and interpret nearly all previous investigations of cue combination in navigation (for an important exception, see Petzschner et al., 2012). We have no doubt, however, that other models, both Bayesian and non-Bayesian, could be developed to explain Chen et al.'s (2017) and Zhao and Warren's (2015b) findings, and we hope that this project serves to stimulate others to create and to test such models.

One limitation of previous studies of cue combination in navigation is that they have quantified performance in a single dimension: either distance between the stopping point and the target (e.g., Chen et al., 2017; Nardini et al., 2008) or homing direction (e.g., Zhao & Warren, 2015b). An important step for future research is to extend the models to the environments in

which participants actually navigate. Friedman et al. (2013) published an excellent tutorial on applications of the standard model to two-dimensions, and in Appendix A we provide the derivation of Gaussian weighted squared-error loss in two dimensions (this derivation generalizes to the incorporation of a prior with individual spatial cues in two dimensions). A major benefit of extending the models to two dimensions is that formal models of single-cue likelihood functions (e.g., models that generate bivariate distributions of locations on the ground plane) can be developed and used to derive the posterior. Performance in multiple cue conditions can then be predicted in terms of parameters of the likelihood functions instead of performance statistics collected in single-cue experimental conditions (e.g., Xu et al., 2017). This type of modeling would represent a significant step forward in understanding the computations involved in integrating spatial cues to estimate one's position and the locations of navigational goals in the environment (e.g., Fetsch et al., 2013; Fetsch et al., 2012; Seilheimer et al., 2014).

Incorporating loss functions into navigational planning and decisions has practical application. The use of mapping apps is ubiquitous. For driving, these apps minimize total travel time, which means that the loss function assigns higher costs to routes that take longer, accounting for both distance and the speed limit (as well as some real-time data, such as traffic density and wrecks). Anyone who has used such apps in areas not well-served by interstate highways will at some point find themselves driving on small roads in remote places. It is often very difficult to determine in advance the status of a section of road that a mapping app selects for the route (e.g., the first author once found himself on a muddy unpaved road in rural Colorado on a drive from Nashville to Jackson, Wyoming, because that short segment produced the optimal solution for total travel time). These apps typically have settings that allow the user to avoid highways, tolls, and ferries, thereby customizing the loss function to some extent. However, a driver who is pulling a travel trailer may have a maximum speed well below the speed limit on many sections of highway, and therefore may want to minimize distance; other drivers may wish to minimize route complexity or maximize the use of interstate highways; and a parent may want their child to use a route that maximizes safety. All of these options for alternative loss functions could easily be made available with current technology, but as far as we know, none is currently implemented.

Criticisms of Bayesian models

Bayesian models have a rich history of application in many domains of the psychological sciences (e.g., Griffiths et al., 2008; Yuille & Bülthoff, 1996; Zaki, 2013). Although these models have been successful in predicting a wide array of behaviors, they also have been scathingly criticized (Bowers & Davis, 2012). One line of criticism focuses on the (prior) plausibility of Bayesian models in light of the many demonstrations of nonoptimality in human behavior. This is a valid point, but not one that applies to our agenda, as we view Bayesian decision theory as a conceptual framework for designing experiments, developing models, and interpreting data (as discussed subsequently), and expect experimental findings to adjudicate between Bayesian and non-Bayesian cognitive models.

A second common criticism is that Bayesian models have so many degrees of freedom that they are unfalsifiable. We believe that this criticism is unwarranted. Bayesian models typically have few parameters to estimate; most of the structure is imposed by the model itself (Ma, 2019). In many investigations of cue combination in navigation, there were no free parameters (Chen et al., 2017; Sjolund et al., 2018; Zhao & Warren, 2015b). Bayesian models are more complex than many cognitive models, as they include three independent components: The prior distribution, the likelihood function, and the decision rule. Criticizing a Bayesian model because it takes advantage of the structure available to it is not reasonable to our minds. However, we appreciate that Bayesian models may seem ad hoc given the numerous choices available to the researcher in constructing and testing a model. These implicit degrees of freedom are not captured in typical measures of model complexity. This concern is real and important, and one that applies to many domains of modeling, and indeed, to the conduct of psychological science in general (Simmons et al., 2011). The solution to this problem is transparency in model development and evaluation. The selection of model components (e.g., the prior) must be justified in advance of data collection, preferably on the basis of independent theoretical or empirical considerations, and researchers must identify revisions to the model (e.g., additional parameters) that were motivated by the findings. Other researchers can then evaluate plausibility for themselves.

A third criticism is that non-Bayesian models are not given sufficient consideration in discussions of Bayesian models. This weakness might have existed in the past but it is much less of a problem now. As just two examples: Norton et al. (2019) tested five Bayesian models and six non-Bayesian models and Laquitaine and Gardner (2018) tested 13 Bayesian models and one non-Bayesian model. Importantly, the Bayesian models tested by Laquitaine and Gardner included variants that differed in prior distributions and loss functions. The results of both projects favored non-Bayesian alternatives, which should allay concerns that Bayesian models, even complex ones, are not falsifiable.

Limitations of the current project

This project is limited in at least two ways. First, the models were applied to existing data sets, obtained from experiments that were not designed to test models of the kind proposed here. Future studies need to test directly components of Bayesian-decision-theoretic models: Are navigators sensitive to Bayesian priors (e.g., nonuniform distributions of goal locations), and if so, do they use them in a manner consistent with Bayesian decision theory? Do navigators account for the costs of navigational errors and paths in principled and quantifiable ways, and if so, which loss functions are used? Are there systematic individual differences in the selection of decision rules?

The second limitation follows from the first: The models were not tested at the level of individual participants and the model comparisons were therefore limited to mean performance. Chen et al. (2017), in particular, did not collect enough data per participant per condition to permit reliable model fitting for individual participants. Future experiments need to be designed so that model parameters can be estimated for individual participants, preferably using methods that specify structure to individual variation in parameter values (Lee & Wagenmakers, 2013; Shiffrin et al., 2008).

Coda

Bayesian decision theory is not a model of human performance. Bayesian decision theory, as applied to psychological phenomena, is an ideal observer (Geisler, 2011): It specifies a mathematically optimal way to use information from multiple sources to make decisions. Human spatial navigation is precisely the type of problem that Bayesian decision theory was developed to explore. Bayesian decision theory should not be expected to predict all of the details of human performance in a task; put another way, one should not expect human performance always to be consistent with Bayesian principles. Investigations of human performance generally, and navigation performance specifically, are at least as informative when behavior deviates from the predictions of Bayesian decision theory as when it is consistent with those predictions, as such inconsistencies can lead to productive questions about cognitive processes: Where do the deviations occur? Is it possible to bring theory and data into correspondence by relaxing one or more assumptions of the theory, and if so, what do these modifications imply about the underlying cognitive processes? What cognitive or perceptual processes may lead to non-Bayesian performance? One does not need Bayesian decision theory to conduct experiments that investigate the effects of prior experience on navigation performance or the influence of the costs of errors on the selection of navigational paths. Bayesian decision theory, however, provides a coherent conceptual framework for designing such experiments, developing process models of the tasks, and interpreting the findings.

Midday on October 22, 1707, the day of the Isles of Scilly disaster, Admiral Shovell gathered the sailing masters (the principal navigators) of the fleet on board his ship, the *Association*, and consulted them about the fleet's position (Cooke, 1883). All but one thought that the fleet was a safe distance from the Isles of Scilly, the lone dissenter of the opinion that the fleet was only 3 hours sail from the islands. Shovell adopted the majority opinion, and set sail on a course that he thought was taking the fleet into the open English Channel, but in fact took the ships directly into the deadly rocks in foul weather. It does not appear that the Admiral considered the enormous variability of the sailing masters' estimates of the fleet's position (May, 1960) or the relative costs of navigational errors under the various estimates of the fleet's position. Had he done either, the outcome might have been very different.

Appendix A

This appendix contains the derivation of the Bayes action (sometimes referred to as the Bayes rule) for weighted squared-error loss. Our aim is to present these derivations in a manner that is accessible to graduate students in psychology who have completed one year of graduate statistics and have sufficient familiarity with calculus to understand differentiation and integration.

Information form of univariate normal

The most common way to parameterize the univariate Gaussian probability density function (pdf; i.e., normal distribution) in psychology is as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right). \tag{A1}$$

This parameterization is often referred to as the *moment* form (π in this formula is the mathematical constant, not a probability distribution). An alternative parameterization is the *canonical* or *information* form, which is useful for many derivations, especially in the multivariate case (Bromiley, 2018; Shön & Lindsten, 2011). We begin by deriving the information form for the univariate Gaussian pdf:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x^2 - 2x\mu + \mu^2)}{2\sigma^2}\right), \tag{A2}$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{x\mu}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2}\right], \tag{A3}$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\mu^2}{2\sigma^2}\right] \exp\left[\frac{x\mu}{\sigma^2} - \frac{x^2}{2\sigma^2}\right]. \tag{A4}$$

By taking the natural logarithm of the first two terms, they can be included within the exponential as a single term that does not depend on x:

$$Z = \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\mu^2}{2\sigma^2}\right]\right)$$
$$= -\frac{1}{2}\left(ln(2\pi) + ln(\sigma^2) + \frac{\mu^2}{\sigma^2}\right), \tag{A5}$$

$$f(x) = \exp\left[Z + \frac{x\mu}{\sigma^2} - \frac{x^2}{2\sigma^2}\right]. \tag{A6}$$

The natural parameters of this function are the scaled mean $\left(\frac{\mu}{\sigma^2}\right)$ and the inverse variance, or reliability $\left(\frac{1}{\sigma^2}\right)$.

General derivation of Bayes action for weighted squared-error loss

The Bayes action for weighted squared-error loss is stated in Berger (1985) and in Robert (2007), but both sources leave the proof to the reader. We are not aware of a published, complete derivation, and provide one here.

The loss function for weighted squared-error loss is,

$$L(\theta, a) = w(\theta)(\theta - a)^2, \tag{A7}$$

where θ is the parameter to be estimated (e.g., the correct target location), a is the action (e.g., the actual location selected), and $w(\theta)$ is a non-negative function of θ. The aim is to minimize the *posterior expected loss* of the action a, for the loss function $L(\theta, a)$ and posterior distribution $\pi(\theta|x)$:

$$\rho(\pi(\theta|x), a) = \int w(\theta)(\theta - a)^2 \pi(\theta|x)d\theta. \tag{A8}$$

The variable x refers to the data that are informative about θ.

We first expand the quadratic expression and distribute the integral:

$$\rho(\pi(\theta|x), a) = \int w(\theta)\left(\theta^2 - 2a\theta + a^2\right)\pi(\theta|x)d\theta, \tag{A9}$$

$$= \int \theta^2 w(\theta)\pi(\theta|x)d\theta - 2a\int \theta \, w(\theta)\pi(\theta|x)d\theta$$
$$+ a^2 \int w(\theta)\pi(\theta|x)d\theta. \tag{A10}$$

The minimum is obtained by taking the derivative of (A10) with respect to a, setting the result to 0, and solving for a:

$$0 = \frac{d}{da}\left[\int \theta^2 w(\theta)\pi(\theta|x)d\theta - 2a\int \theta \, w(\theta)\pi(\theta|x)d\theta + a^2\int w(\theta)\pi\left(\theta|x\right)d\theta\right], \tag{A11}$$

$$0 = 0 - 2\int \theta \, w(\theta)\pi(\theta|x)d\theta + 2a\int w(\theta)\pi(\theta|x)d\theta, \tag{A12}$$

$$2\int \theta \, w(\theta)\pi(\theta|x)d\theta = 2a\int w(\theta)\pi(\theta|x)d\theta, \tag{A13}$$

$$\frac{\int \theta w(\theta)\pi(\theta|x)d\theta}{\int w(\theta)\pi(\theta|x)d\theta} = a = \delta^\pi(x), \tag{A14}$$

which is the result we set out to prove. (This derivation assumes that all integrals converge and that $\int w(\theta)\pi(\theta|x)d\theta > 0$. The superscript on δ designates the distribution with respect to which loss is minimized.)

Derivation of Bayes action for Gaussian weighted squared-error loss in 1D

The goal is to evaluate (A14) for a Gaussian weight function and normal posterior. Although one can derive this function using existing sources (e.g., Bromiley, 2018; Oruç et al., 2003), doing so requires establishing equivalences between entities that are not obviously related or expressing the probability density functions in terms of reliabilities. For simplicity and transparency, we provide a complete derivation in the current notation here.

Let $w(\theta) = a\exp\left[\frac{-(\theta-b)^2}{2c^2}\right]$ and suppose that $\pi(\theta|x) = N(\mu, \sigma^2)$. ($a$, b, and c are the generic parameters of the Gaussian function.) The goal is to derive the Bayes action $\delta^\pi(x)$ of Equation (A14). We begin by expressing both functions in information form to make the subsequent derivations more tractable.

$$Z_w = \ln(a) - \frac{1}{2}\left(\frac{b^2}{c^2}\right), \tag{A15}$$

$$w(\theta) = \exp\left[Z_w + \frac{\theta b}{c^2} - \frac{\theta^2}{2c^2}\right]. \tag{A16}$$

These two equations are obtained in the same manner as (A5) and (A6), respectively.

$$Z_\pi = -\frac{1}{2}\left(ln(2\pi) + ln(\sigma^2) + \frac{\mu^2}{\sigma^2}\right), \tag{A17}$$

$$\pi(\theta|x) = \exp\left[Z_\pi + \frac{\theta\mu}{\sigma^2} - \frac{\theta^2}{2\sigma^2}\right]. \tag{A18}$$

These two equations are the same as Eq. (A5) and Eq. (A6) with θ replacing x.

The product of $w(\theta)$ and $\pi(\theta|x)$ can be computed by summing the terms within the exponentials:

$$w(\theta)\pi(\theta|x)$$
$$= \exp\left[(Z_w + Z_\pi) + \theta\left(\frac{b}{c^2} + \frac{\mu}{\sigma^2}\right) - \frac{1}{2}\theta^2\left(\frac{1}{c^2} + \frac{1}{\sigma^2}\right)\right], \tag{A19}$$

$$= \exp\left[(Z_w + Z_\pi) - Z_{w\pi} + Z_{w\pi} + \theta\left(\frac{b}{c^2} + \frac{\mu}{\sigma^2}\right) - \frac{1}{2}\theta^2\left(\frac{1}{c^2} + \frac{1}{\sigma^2}\right)\right], \tag{A20}$$

$$= \exp[(Z_w + Z_\pi) - Z_{w\pi}] \exp\left[Z_{w\pi} + \theta\left(\frac{b}{c^2} + \frac{\mu}{\sigma^2}\right) - \frac{1}{2}\theta^2\left(\frac{1}{c^2} + \frac{1}{\sigma^2}\right)\right], \quad (A21)$$

$Z_{w\pi}$ is *constructed* from elements that do not depend on θ so that the second exponential has the Gaussian form in Eq. (A6):

$$Z_{w\pi} = -\frac{1}{2}\left(\ln(2\pi) + ln\left[\left(\frac{1}{c^2} + \frac{1}{\sigma^2}\right)^{-1}\right] + \frac{\left(\frac{b}{c^2} + \frac{\mu}{\sigma^2}\right)^2}{\left(\frac{1}{c^2} + \frac{1}{\sigma^2}\right)}\right). \quad (A22)$$

To see the correspondence, note that:

$$\frac{b}{c^2} + \frac{\mu}{\sigma^2} = \frac{b\sigma^2}{c^2\sigma^2} + \frac{\mu c^2}{\sigma^2 c^2} = \frac{b\sigma^2 + \mu c^2}{\sigma^2 c^2}, \quad (A23)$$

$$= \frac{b\sigma^2 + \mu c^2}{\sigma^2 c^2} \times \frac{\frac{1}{\sigma^2 + c^2}}{\frac{1}{\sigma^2 + c^2}}, \quad (A24)$$

$$= \frac{\left(\frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}\right)}{\left(\frac{\sigma^2 c^2}{\sigma^2 + c^2}\right)}, \quad (A25)$$

And that:

$$\frac{1}{c^2} + \frac{1}{\sigma^2} = \frac{\sigma^2 + c^2}{\sigma^2 c^2}, \quad (A26)$$

$$\left(\frac{1}{c^2} + \frac{1}{\sigma^2}\right)^{-1} = \frac{\sigma^2 c^2}{\sigma^2 + c^2}. \quad (A27)$$

The last term in Eq. (A22) is the square of Eq. (A25) divided by Eq. (A26),

$$\frac{\left(\frac{b}{c^2} + \frac{\mu}{\sigma^2}\right)^2}{\left(\frac{1}{c^2} + \frac{1}{\sigma^2}\right)} = \frac{\left[\frac{\left(\frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}\right)}{\left(\frac{\sigma^2 c^2}{\sigma^2 + c^2}\right)}\right]^2}{\left(\frac{\sigma^2 + c^2}{\sigma^2 c^2}\right)}, \quad (A28)$$

which simplifies to,

$$= \frac{\left(\frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}\right)^2}{\left(\frac{\sigma^2 c^2}{\sigma^2 + c^2}\right)}. \quad (A29)$$

Rewriting the second exponential in Eqs. (A21) and (A22), we have, respectively:

$$\exp\left[Z_{w\pi} + \theta\left(\frac{\left(\frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}\right)}{\left(\frac{\sigma^2 c^2}{\sigma^2 + c^2}\right)}\right) - \frac{1}{2}\theta^2\left(\frac{\sigma^2 + c^2}{\sigma^2 c^2}\right)\right], \quad (A30)$$

$$Z_{w\pi} = -\frac{1}{2}\left(\ln(2\pi) + ln\left[\left(\frac{\sigma^2 + c^2}{\sigma^2 c^2}\right)^{-1}\right] + \frac{\left(\frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}\right)^2}{\left(\frac{\sigma^2 c^2}{\sigma^2 + c^2}\right)}\right). \quad (A31)$$

By defining $\mu_{w\pi} = \frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}$ and $\sigma_{w\pi}^2 = \frac{\sigma^2 c^2}{\sigma^2 + c^2}$, we obtain:

$$f(\theta) = \exp\left[Z_{w\pi} + \theta\left(\frac{\mu_{w\pi}}{\sigma_{w\pi}^2}\right) - \frac{1}{2}\theta^2\left(\frac{1}{\sigma_{w\pi}^2}\right)\right], \quad (A32)$$

$$Z_{w\pi} = -\frac{1}{2}\left(\ln(2\pi) + ln\left(\sigma_{w\pi}^2\right) + \frac{\mu_{w\pi}^2}{\sigma_{w\pi}^2}\right), \quad (A33)$$

which are of the same form as Eq. (A6) and Eq. (A5), respectively. This means that the second exponential in Eq. (A21) is the equation for a Gaussian distribution with mean and variance, respectively:

$$\mu_{w\pi} = \frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}, \quad (A34)$$

$$\sigma_{w\pi}^2 = \frac{\sigma^2 c^2}{\sigma^2 + c^2} = \left(\frac{1}{c^2} + \frac{1}{\sigma^2}\right)^{-1}. \quad (A35)$$

We can now evaluate the integral in the denominator:

$$\int w(\theta)\pi(\theta|x)d\theta = \int \exp[(Z_w + Z_\pi) - Z_{w\pi}]\exp\left[Z_{w\pi} + \theta\left(\frac{\mu_{w\pi}}{\sigma_{w\pi}^2}\right) - \frac{1}{2}\theta^2\left(\frac{1}{\sigma_{w\pi}^2}\right)\right]d\theta, \quad (A36)$$

$$= \exp[(Z_w + Z_\pi) - Z_{w\pi}]\int \exp\left[Z_{w\pi} + \theta\left(\frac{\mu_{w\pi}}{\sigma_{w\pi}^2}\right) - \frac{1}{2}\theta^2\left(\frac{1}{\sigma_{w\pi}^2}\right)\right]d\theta, \quad (A37)$$

because the first exponential does not depend on θ. The integral of a probability density function is 1; hence,

$$\int w(\theta)\pi(\theta|x)d\theta = \exp[(Z_w + Z_\pi) - Z_{w\pi}]. \quad (A38)$$

If $w(\theta)$ is normal (i.e., $a = (2\pi c^2)^{-\frac{1}{2}}$), then Eq. (A38) is normal with mean μ (assuming that b is viewed as the variable) and variance $c^2 + \sigma^2$.

Now, for the integral in the numerator:

$$\int \theta\, w(\theta)\,\pi(\theta|x)d\theta = \int \theta\, \exp[(Z_w + Z_\pi) - Z_{w\pi}]\, \exp\left[Z_{w\pi} + \theta\left(\frac{\mu_{w\pi}}{\sigma_{w\pi}^2}\right) - \frac{1}{2}\theta^2\left(\frac{1}{\sigma_{w\pi}^2}\right)\right]d\theta,$$

$$(A39)$$

$$= \exp[(Z_w + Z_\pi) - Z_{w\pi}]\int \theta\, \exp\left[Z_{w\pi} + \theta\left(\frac{\mu_{w\pi}}{\sigma_{w\pi}^2}\right) - \frac{1}{2}\theta^2\left(\frac{1}{\sigma_{w\pi}^2}\right)\right]d\theta. \qquad (A40)$$

The integral in this expression is the expected value of θ distributed as a Gaussian distribution with mean,

$$\mu_{w\pi} = \frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}. \qquad (A41)$$

Hence,

$$\delta^\pi(x) = \frac{\exp[(Z_w + Z_\pi) - Z_{w\pi}]\,\mu_{w\pi}}{\exp[(Z_w + Z_\pi) - Z_{w\pi}]}, \qquad (A42)$$

$$\delta^\pi(x) = \mu_{w\pi} = \frac{b\sigma^2 + \mu c^2}{\sigma^2 + c^2}. \qquad (A43)$$

The mean, $\mu_{w\pi}$, can also be expressed as,

$$\mu_{w\pi} = \sigma_{w\pi}^2\left(\frac{b}{c^2} + \frac{\mu}{\sigma^2}\right), \qquad (A44)$$

where,

$$\sigma_{w\pi}^2 = \frac{\sigma^2 c^2}{(\sigma^2 + c^2)}. \qquad (A45)$$

Rearranging terms in Eq. (A43),

$$\delta^\pi(x) = \mu - \left(\frac{\sigma^2}{\sigma^2 + c^2}\right)(\mu - b). \qquad (A46)$$

Extension to 2D

See Friedman et al. (2013) for a lucid exposition of the standard cue-combination model in two dimensions. Here, we derive the Bayes action in two dimensions for Gaussian weighted squared-error loss. The same derivation can be used for the application of a Gaussian prior.

Following Bromiley (2018), the information form of a single multivariate Gaussian is:

$$f(\theta) = \exp\left[Z + \eta^T\theta - \frac{1}{2}\theta^T\Lambda\theta\right], \qquad (A47)$$

where,

$$Z = -\frac{1}{2}\left(d\ln 2\pi - \ln|\Lambda| + \eta^T\Lambda^{-1}\eta\right); \qquad (A48)$$

$\Lambda = \Sigma^{-1}$ is the inverse of the covariance matrix; $\eta = \Lambda\mu$; $|\Lambda|$ is the determinant of Λ; and d is the number of variates. θ and μ in this context are d-dimensional vectors. Bromiley shows

that the product of N multivariate Gaussians is a scaled Gaussian, analogous to the univariate case. Our interest is in the product of a 2D Gaussian weight function and 2D Gaussian posterior. We assume here for simplicity of notation that the weight function is multivariate normal. The two-dimensions, or variates, are v and y (v rather than x because x has been used for another purpose).

$$w(\theta)\pi(\theta|x)$$

$$= \exp[\Sigma Z_i - Z_{w\pi}]\exp\left[Z_{w\pi} + \eta_{w\pi}^T\theta - \frac{1}{2}\theta^T\Lambda_{w\pi}\theta\right], \qquad (A49)$$

where,

$$\Sigma Z_i = -\frac{1}{2}\left[4\ln 2\pi - (\ln|\Lambda_w| + \ln|\Lambda_\pi|) + (\eta_w^T\Lambda_w^{-1}\eta_w + \eta_\pi^T\Lambda_\pi^{-1}\eta_\pi)\right], \qquad (A50)$$

$$Z_{w\pi} = -\frac{1}{2}\left[2\ln 2\pi - \ln|\Lambda_{w\pi}| + \eta_{w\pi}^T\Lambda_{w\pi}^{-1}\eta_{w\pi}\right]. \qquad (A51)$$

The components of these formulae are as follows:

$$\Sigma_w = \begin{bmatrix} c_v^2 & c_{vy} \\ c_{vy} & c_y^2 \end{bmatrix}, \Sigma_\pi = \begin{bmatrix} \sigma_v^2 & \sigma_{vy} \\ \sigma_{vy} & \sigma_y^2 \end{bmatrix}, \qquad (A52)$$

$$\Lambda_w = \Sigma_w^{-1} = |\Sigma_w|^{-1}adjugate(\Sigma_w)$$

$$= \begin{bmatrix} \dfrac{c_y^2}{|\Sigma_w|} & -\dfrac{c_{vy}}{|\Sigma_w|} \\ -\dfrac{c_{vy}}{|\Sigma_w|} & \dfrac{c_v^2}{|\Sigma_w|} \end{bmatrix}, \qquad (A53)$$

$$\Lambda_\pi = \begin{bmatrix} \dfrac{\sigma_y^2}{|\Sigma_\pi|} & -\dfrac{\sigma_{vy}}{|\Sigma_\pi|} \\ -\dfrac{\sigma_{vy}}{|\Sigma_\pi|} & \dfrac{\sigma_v^2}{|\Sigma_\pi|} \end{bmatrix}, \qquad (A54)$$

$$\eta_w = \Lambda_w\begin{bmatrix} b_v \\ b_y \end{bmatrix} = \begin{bmatrix} \dfrac{c_y^2 b_v - c_{vy}b_y}{|\Sigma_w|} \\ \dfrac{c_v^2 b_y - c_{vy}b_v}{|\Sigma_w|} \end{bmatrix}, \qquad (A55)$$

$$\eta_\pi = \begin{bmatrix} \dfrac{\sigma_y^2\mu_v - \sigma_{vy}\mu_y}{|\Sigma_\pi|} \\ \dfrac{\sigma_v^2\mu_y - \sigma_{vy}\mu_v}{|\Sigma_\pi|} \end{bmatrix}, \qquad (A56)$$

$$\Lambda_{w\pi} = (\Lambda_w + \Lambda_\pi)$$

$$= \begin{bmatrix} \dfrac{c_y^2}{|\Sigma_w|} + \dfrac{\sigma_y^2}{|\Sigma_\pi|} & -\left(\dfrac{c_{vy}}{|\Sigma_w|} + \dfrac{\sigma_{vy}}{|\Sigma_\pi|}\right) \\ -\left(\dfrac{c_{vy}}{|\Sigma_w|} + \dfrac{\sigma_{vy}}{|\Sigma_\pi|}\right) & \dfrac{c_v^2}{|\Sigma_w|} + \dfrac{\sigma_v^2}{|\Sigma_\pi|} \end{bmatrix}, \qquad (A57)$$

and

$$\eta_{w\pi} = (\eta_w + \eta_\pi) = \begin{bmatrix} \dfrac{c_y^2 b_v - c_{vy} b_y}{|\Sigma_w|} + \dfrac{\sigma_y^2 \mu_v - \sigma_{vy} \mu_y}{|\Sigma_\pi|} \\[2mm] \dfrac{c_v^2 b_y - c_{vy} b_v}{|\Sigma_w|} + \dfrac{\sigma_v^2 \mu_y - \sigma_{vy} \mu_v}{|\Sigma_\pi|} \end{bmatrix}. \quad (A58)$$

Equation (A14) generalizes to the multidimensional case. Recall that the aim is to minimize expected loss given the weight function and the posterior. The expected loss,

$$\rho(\pi(\theta|x), a) = \int w(\theta)(\theta - a)^T (\theta - a) \pi(\theta|x) d\theta, \quad (A59)$$

where θ, a, and x are d-dimensional vectors and $w(\theta)$ is a nonnegative real-valued function of θ. Differentiating with respect to a we have:

$$\frac{\partial \rho(\pi(\theta|a), a)}{\partial a} = \frac{\partial}{\partial a} \left[\int w(\theta)(\theta - a)^T (\theta - a) \pi(\theta|x) d\theta \right], \quad (A60)$$

$$= \int w(\theta) \frac{\partial}{\partial a} \left[(\theta - a)^T (\theta - a) \right] \pi(\theta|x) d\theta, \quad (A61)$$

$$= -2 \int w(\theta)(\theta - a) \pi(\theta|x) d\theta, \quad (A62)$$

$$= -2 \int \theta w(\theta) \pi(\theta|x) d\theta + 2a \int w(\theta) \pi(\theta|x) d\theta, \quad (A63)$$

assuming that the conditions of differentiation under the integral are met (Casella & Berger, 2002; Rudin, 1976).

Setting Eq. (A63) equal to zero:

$$0 = -2 \int \theta w(\theta) \pi(\theta|x) d\theta + 2a \int w(\theta) \pi(\theta|x) d\theta \Rightarrow, \quad (A64)$$

$$\int \theta w(\theta) \pi(\theta|x) d\theta = a \int w(\theta) \pi(\theta|x) d\theta, \quad (A65)$$

$$a = \frac{\int \theta w(\theta) \pi(\theta|x) d\theta}{\int w(\theta) \pi(\theta|x) d\theta} = \delta^\pi(x). \quad (A66)$$

For the specific 2D case of interest, we have:

$$\delta^\pi(x) = \frac{\int \theta \exp[\sum Z_i - Z_{w\pi}] \exp\left[Z_{w\pi} + \eta_{w\pi}^T \theta - \frac{1}{2} \theta^T \Lambda_{w\pi} \theta \right] d\theta}{\int \exp[\sum Z_i - Z_{w\pi}] \exp\left[Z_{w\pi} + \eta_{w\pi}^T \theta - \frac{1}{2} \theta^T \Lambda_{w\pi} \theta \right] d\theta}, \quad (A67)$$

$$\delta^\pi(x) = \frac{\exp[\sum Z_i - Z_{w\pi}] \int \theta \exp\left[Z_{w\pi} + \eta_{w\pi}^T \theta - \frac{1}{2} \theta^T \Lambda_{w\pi} \theta \right] d\theta}{\exp[\sum Z_i - Z_{w\pi}] \int \exp\left[Z_{w\pi} + \eta_{w\pi}^T \theta - \frac{1}{2} \theta^T \Lambda_{w\pi} \theta \right] d\theta}, \quad (A68)$$

$$\delta^\pi(x) = \int \theta \exp\left[Z_{w\pi} + \eta_{w\pi}^T \theta - \frac{1}{2} \theta^T \Lambda_{w\pi} \theta \right] d\theta, \quad (A69)$$

$$\delta^\pi(x) = E^{w\pi}[\theta] = \mu_{w\pi}. \quad (A70)$$

To obtain the unscaled mean, $\mu_{w\pi}$, we need to evaluate $\Lambda_{w\pi}^{-1} \eta_{w\pi}$:

$$\Lambda_{w\pi}^{-1} \eta_{w\pi} = (\Lambda_w + \Lambda_\pi)^{-1} (\eta_w + \eta_\pi), \quad (A71)$$

$$= \left(\Sigma_w^{-1} + \Sigma_\pi^{-1} \right)^{-1} \left(\Sigma_w^{-1} b + \Sigma_\pi^{-1} \mu \right). \quad (A72)$$

Note that this matric equation is of the same form as Eq. (A44) above (refer also to Eq. A35), with Σ_i^{-1} playing the role of $\frac{1}{\sigma_i^2}$.

In the special case where the covariances, $c_{vy} = \sigma_{vy} = 0$, then,

$$\Sigma_w^{-1} = \begin{bmatrix} \dfrac{1}{c_v^2} & 0 \\[2mm] 0 & \dfrac{1}{c_y^2} \end{bmatrix}, \Sigma_\pi^{-1} = \begin{bmatrix} \dfrac{1}{\sigma_v^2} & 0 \\[2mm] 0 & \dfrac{1}{\sigma_y^2} \end{bmatrix}. \quad (A73)$$

Simplifying (A72):

$$\delta^\pi(x) = \begin{bmatrix} \dfrac{1}{c_v^2} + \dfrac{1}{\sigma_v^2} & 0 \\[2mm] 0 & \dfrac{1}{c_y^2} + \dfrac{1}{\sigma_y^2} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{b_v}{c_v^2} + \dfrac{\mu_v}{\sigma_v^2} \\[2mm] \dfrac{b_y}{c_y^2} + \dfrac{\mu_y}{\sigma_y^2} \end{bmatrix}, \quad (A74)$$

$$\delta^\pi(x) = \begin{bmatrix} \dfrac{\sigma_v^2 c_v^2}{c_v^2 + \sigma_v^2} & 0 \\[2mm] 0 & \dfrac{\sigma_y^2 c_y^2}{c_y^2 + \sigma_y^2} \end{bmatrix} \begin{bmatrix} \dfrac{\sigma_v^2 b_v + c_v^2 \mu_v}{\sigma_v^2 c_v^2} \\[2mm] \dfrac{\sigma_y^2 b_y + c_y^2 \mu_y}{\sigma_y^2 c_y^2} \end{bmatrix}, \quad (A75)$$

$$\delta^\pi(x) = \begin{bmatrix} \dfrac{\sigma_v^2 b_v + c_v^2 \mu_v}{c_v^2 + \sigma_v^2} \\[2mm] \dfrac{\sigma_y^2 b_y + c_y^2 \mu_y}{c_y^2 + \sigma_y^2} \end{bmatrix}. \quad (A76)$$

In words, the Bayes action, assuming $c_{vy} = \sigma_{vy} = 0$, is the vector corresponding to the weighted combination of the mean vectors of the weight and posterior distributions. Each dimension is combined independently, with each distribution weighted according to its relative reliability on that dimension.

Appendix B

Chen et al. (2017) assessed the variability and accuracy of performance using the 2D distance between locations (e.g., the participant's stopping point on the return path and the correct target location; see pp. 113–114 of that article for a discussion of the relative merits of various measures of performance). To ensure that our results would be consistent with theirs, the modeling was conducted on distance data. This means that the models were fit to one-dimensional data (for possible extensions to 2D see the main text and Appendix A). The relevant data and computations can be found in the supplementary Excel file "BDT_Data_and_Models."

Response variabilities in the vision and the body-based conditions were obtained from Chen et al. (2017). These

values were used as estimates of σ_1^2 and σ_2^2 in Table 1. Optimal weights for visual and body-based cues were calculated using the formulae for w_1 and w_2 in Table 1. Chen et al. used four target locations, symmetrically distributed about the midline defined by the participant's initial facing direction on each trial. For the current project, distances between the centroid of observed performance in the conflict condition and the centroids of observed performance in the vision condition and the body-based condition were computed on each side of the midline and then averaged across the two sides (Chen et al. computed the weights for each side and averaged those across sides; the two approaches yield virtually identical estimates). These computations were carried out for each participant. These distances defined the distributions used for modeling and were used to compute the empirical weights (see Table 1). One participant in Experiment 1a: 3 Landmarks was excluded as an outlier because of a very large discrepancy between the optimal weight and the empirical weight on visual cues (.97 vs. .13, respectively).

The distributions for the vision condition, conflict condition, and body-based condition were located on a single dimension by arbitrarily multiplying the mean distance between the vision condition and the conflict condition by −1. This transformation entails that the conditions were ordered, vision, conflict, body-based, with the observed conflict mean equal to 0, for all participants. The predicted mean of the posterior distribution, μ in Eq. (30), was computed for each participant using the formula in Table 1 for the mean of the optimal combination. The mean of the weight function, b in Eq. (30), was set equal to the midpoint between the means of the vision and the body-based conditions.

The ratio r was estimated using a least squares criterion in Solver (Microsoft Excel 16 for Mac) using GRG nonlinear estimation. Sums of squared errors (SSE) between predicted and observed performance were transformed to adjusted log likelihoods using the normal linear model and these were exponentiated to estimate the marginal likelihoods composing the Bayes factor (Claeskens & Hjort, 2008; Raftery, 1995; Wasserman, 2000):

$$\ln m_i = -\frac{n}{2}\left(\ln 2\pi + \ln\left(\frac{SSE_i}{n}\right) + 1\right) - \frac{p_i}{2}\ln(n), \tag{B1}$$

$$B_{ij} = \frac{\exp(\ln m_i)}{\exp(\ln m_j)}, \tag{B2}$$

where n is the number of participants in the experimental condition and p_i is the number of parameters estimated in the model (0 for the null model; 1 for the alternative model). Note that (B1) is ½ the Bayesian information criterion (BIC). Straightforward algebra shows that (B2) is equal to:

$$B_{ij} = \left(\frac{SSE_i}{SSE_j}\right)^{-\frac{n}{2}} n^{\frac{(p_j - p_i)}{2}}. \tag{B3}$$

Appendix C

In this appendix, we derive the predicted empirical weights on single cues according to the decision-theoretical model. We refer to these as the "model empirical weights." Relevant computations can be found in the supplementary Excel file, "BDT_Data_and_Models."

The weight function in the weighted squared-error loss function is assumed to be Gaussian,

$$w(\theta) = a\exp\left[\frac{-(\theta - b)^2}{2c^2}\right], \tag{C1}$$

and the posterior is normal, $\pi(\theta|x) = N(\mu, \sigma^2)$. If the model accurately characterizes performance, then the observed mean in the conflict condition should be (see the main text and Appendix A),

$$\mu_C = \mu - \left(\frac{\sigma^2}{\sigma^2 + c^2}\right)(\mu - b). \tag{C2}$$

As discussed in the main text, we estimated a common value of the relative reliability of the weight function across participants:

$$\mu_C = \mu - r(\mu - b). \tag{C3}$$

The empirical weights are computed from the observed distances between the mean of the conflict condition and the means of the vision and the body-based conditions:

$$d_V = \mu_C - \mu_V, \tag{C4}$$

$$d_B = \mu_B - \mu_C \tag{C5}$$

(assuming that the means are ordered, vision, conflict, body-based, as discussed in Appendix B).

Our goal is to estimate the weights based on the true posterior, which is not observed behaviorally:

$$d'_V = \mu - \mu_V. \tag{C6}$$

Rearranging terms in Eq. (C3) and substituting into Eq. (C6) we obtain:

$$d'_V = [\mu_C + r(\mu - b)] - \mu_V, \tag{C7}$$

$$d'_V = (\mu_C - \mu_V) + r(\mu - b), \tag{C8}$$

$$d'_V = d_V + r(\mu - b). \tag{C9}$$

Similarly,

$$d'_B = d_B - r(\mu - b) \tag{C10}$$

(the second term is subtracted in Eq. C10 because the signs are reversed in Eq. C5 relative to Eq. C4).

The model empirical weight on the visual cue is:

$$w'_V = \frac{d'_B}{(d'_V + d'_B)} \tag{C11}$$

$$w'_V = \frac{d_B - r(\mu - b)}{(d'_V + d'_B)}. \tag{C12}$$

Because $d'_V + d'_B = d_V + d_B$, we obtain from Eq. (C12):

$$w'_V = \widehat{w}_V - \frac{r}{(d_V + d_B)}(\mu - b). \tag{C13}$$

Note also that $d_V + d_B = \mu_B - \mu_V$ in this model implementation.
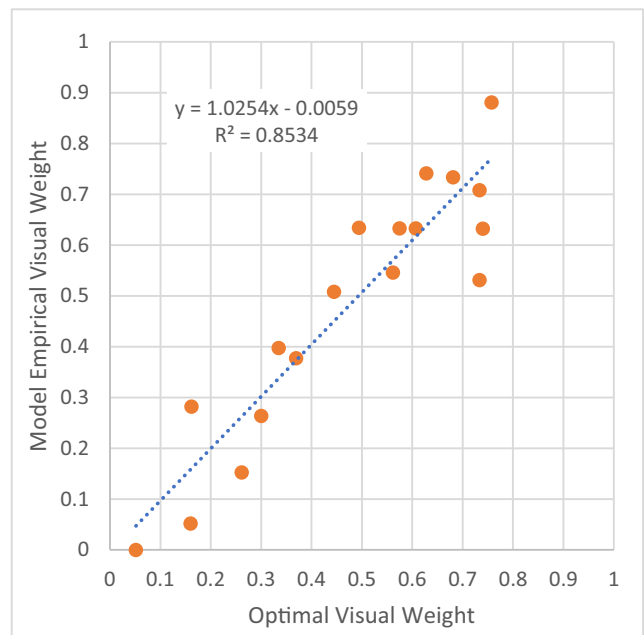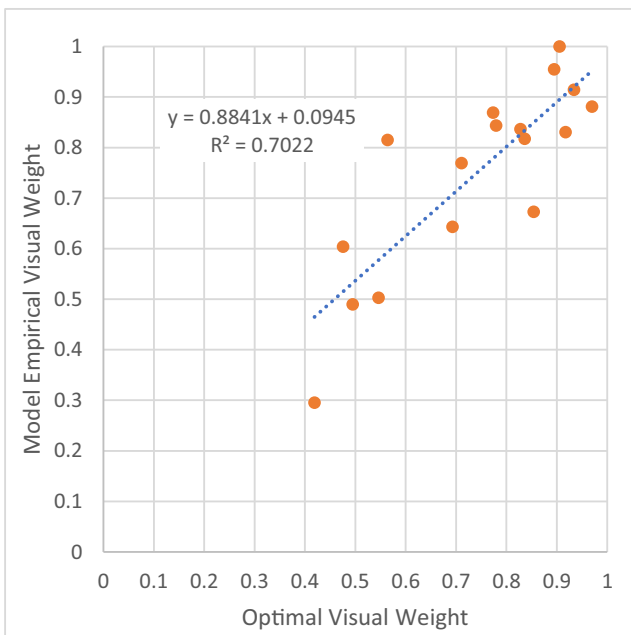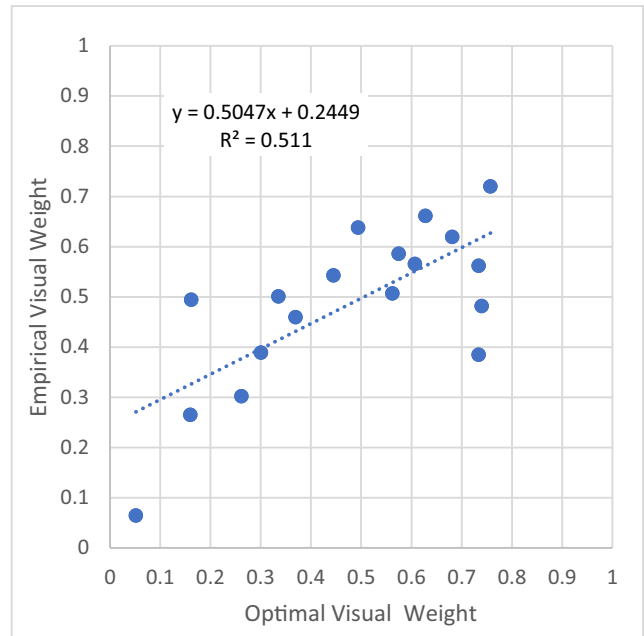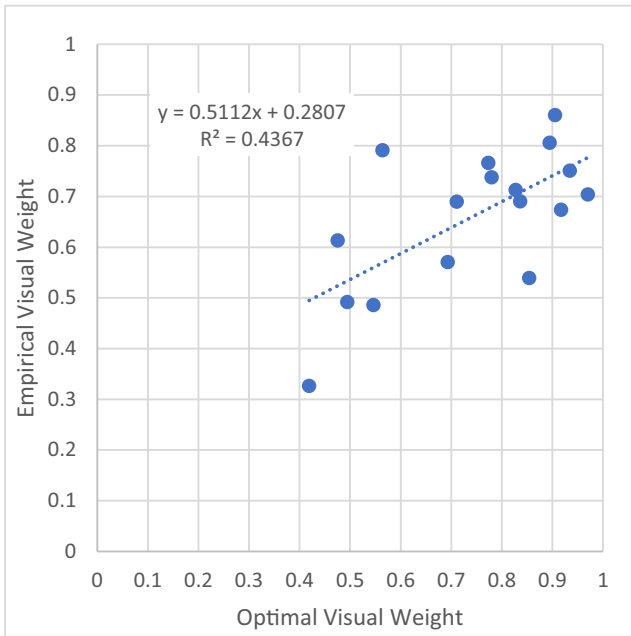
Fig. 12 Regressions of empirical weights on optimal weights for Experiment 1a: 3 Landmarks

Fig 13 Regressions of empirical weights on optimal weights for Experiment 1a: 1 Landmark
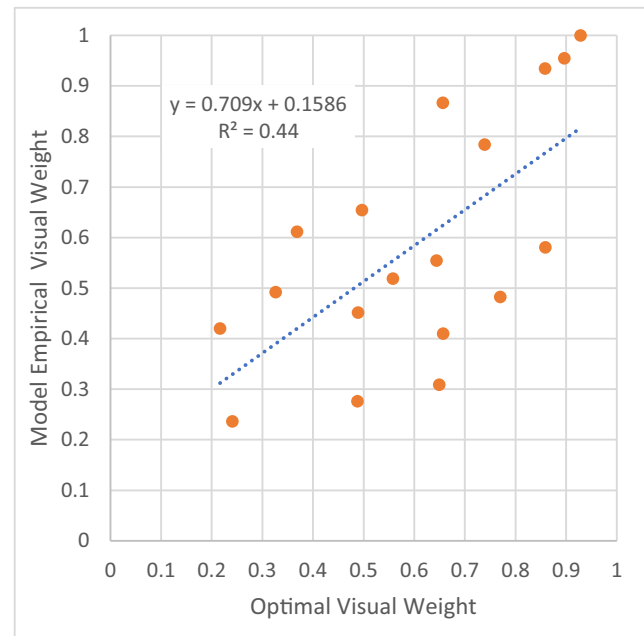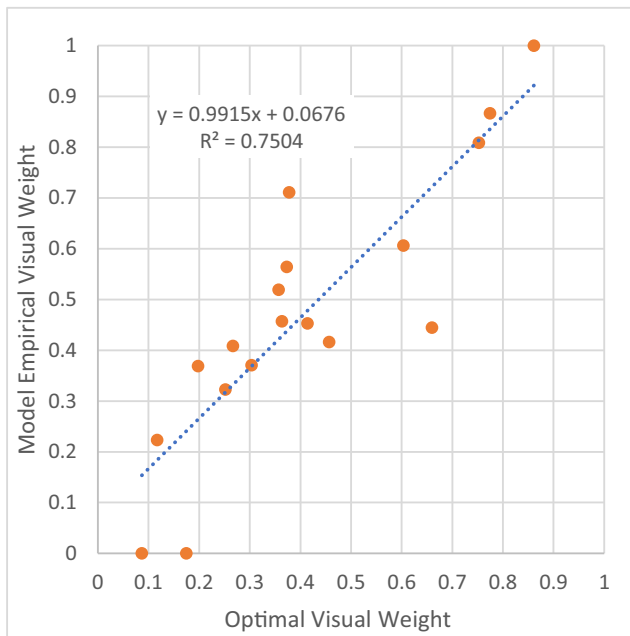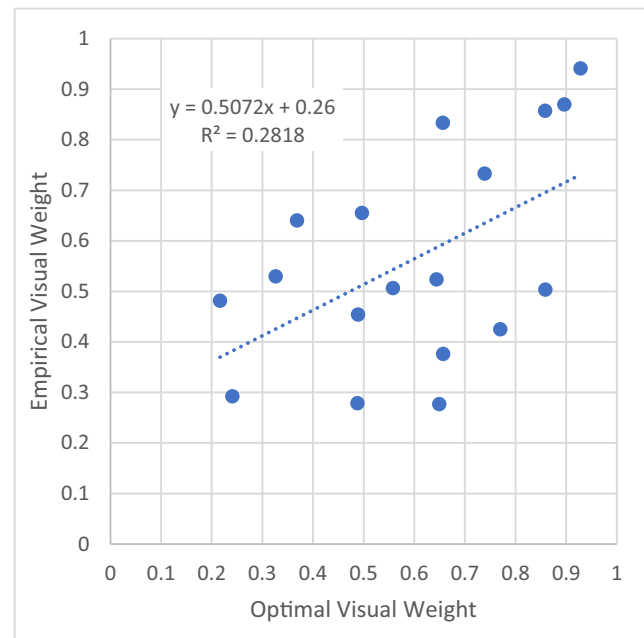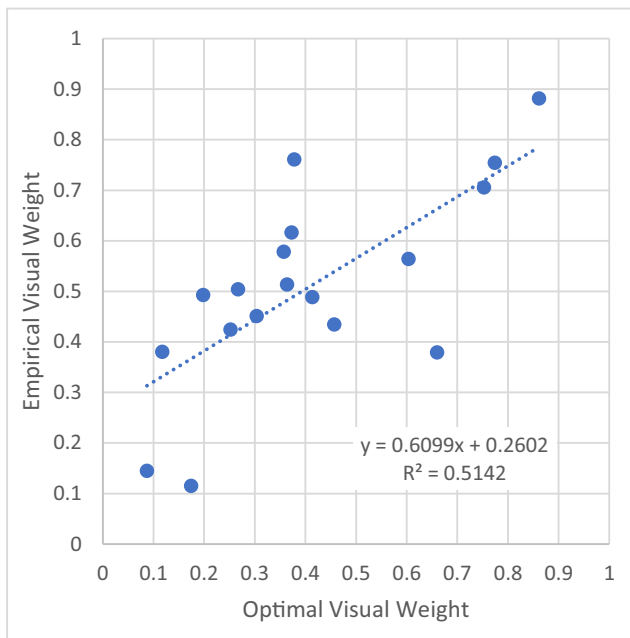
Fig. 14 Regressions of empirical weights on optimal weights for Experiment 1b: No rotation

The model empirical weights computed from Eq. (C13) were regressed on optimal weights using the observed variances in the vision and body-based conditions. Values from Eq. (C13) less than 0 or greater than 1 were set to 0 or 1, respectively. Figures 12, 13, 14, 15 and 16 include these plots and for comparison plots of the (unadjusted) empirical weights regressed on the same optimal weights. Graphs for Chen et al.'s (2017) Experiments 1a, 1b, and Experiment 4: Day 2 are included in this appendix; graphs for Experiment 4: Day 1 are included in the main text.

Fig. 15 Regressions of empirical weights on optimal weights for Experiment 1b: Rotation

Appendix D

This appendix describes the modeling of variances for the decision-theoretic model. These computations can be found in the supplementary Excel file, "BDT_Variances_V2."

Assume that single-cue systems, V and B, are Gaussian with means and variances, μ_V, σ_V^2, μ_B, and σ_B^2, respectively; the posterior is Gaussian with mean, $\mu_P = w_V \mu_V + w_B \mu_B$, $w_V = \frac{\sigma_B^2}{\sigma_V^2 + \sigma_B^2}$ and $w_B = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_B^2}$, and variance, $\sigma_P^2 = \frac{\sigma_V^2 \sigma_B^2}{\sigma_V^2 + \sigma_B^2}$; and the weight function is Gaussian with mean, b, and variance, c^2.
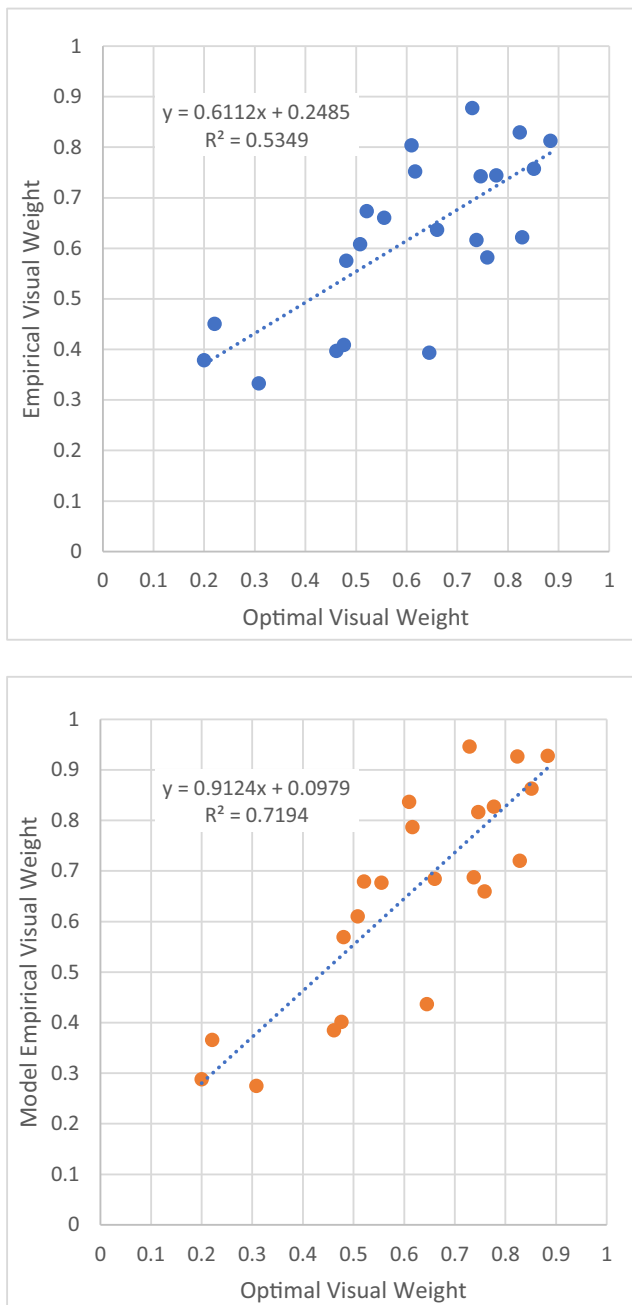
Fig. 16 Regressions of empirical weights on optimal weights for Experiment 4: Day 2

Let x_V and x_B be samples from the single cue distributions on a given trial. The value of the posterior for that trial is,

$$post = w_V x_V + w_B x_B. \tag{D1}$$

One can view the values x_V and x_B as location estimates produced by each of the sensory-perceptual systems (these estimates may themselves be the means of posterior distributions computed by upstream processes). For the R⁺ model, a new sample, x'_V and x'_B, is taken from the single-cue distributions to compute the value of the weight function. These values are averaged, and the weight estimate is sampled from

a Gaussian with mean, $(x'_V + x'_B)/2$, and variance, c^2. The Bayes action for that trial is,

$$BA = post - r(post - weight), \tag{D2}$$

where r is the ratio of variances in Table 3 for the relevant experimental condition.

Rearranging terms,

$$BA = (1-r)post + r\,weight. \tag{D3}$$

Using the standard formula for the variance of the sum of random variables,

$$Var(BA) = (1-r)^2\,Var(post) + r^2\,Var(weight)$$
$$+ 2(1-r)r\,Cov(post, weight) \tag{D4}$$

$$Var(BA) = (1-r)^2\sigma_P^2$$
$$+ r^2\left(c^2 + \left(\frac{1}{2}\right)^2\sigma_V^2 + \left(\frac{1}{2}\right)^2\sigma_B^2\right)$$
$$+ 2(1-r)r\,Cov(post, weight). \tag{D5}$$

Trial-to-trial variability in the single-cue systems affects the variability of the sample from the weight function, and this variability propagates to the Bayes action.

For the R⁺ model, (D5) was evaluated for each participant with $Cov(post, weight) = 0$ (as the samples x_V, x_B and x'_V, x'_B are independent) using estimates of σ_P^2, σ_V^2, and σ_B^2 from their task performance and an estimate of c^2 using Eq. (26) from the main text:

$$c^2 = \frac{\sigma_P^2}{r} - \sigma_P^2. \tag{D6}$$

For the R⁻ model, the single-cue estimates x_V and x_B were used to compute both the posterior and the mean of the weight function; otherwise, the models were identical. This feature of the model causes the covariance between the posterior and the weight function to be positive. The covariance in (D5) can be computed for each participant using the property of bilinearity:

$$Cov(post, weight)$$
$$= Cov\left(w_V x_V + w_B x_B, \frac{1}{2}x_V + \frac{1}{2}x_B\right), \tag{D7}$$

$$Cov(post, weight) = w_V \frac{1}{2}Cov(x_V, x_V) + w_V \frac{1}{2}Cov(x_V, x_B) + w_B \frac{1}{2}Cov(x_B, x_V) +$$
$$w_B \frac{1}{2}Cov(x_B, x_B), \tag{D8}$$

$$Cov(post, weight) = \frac{1}{2}\left(w_V \sigma_V^2 + w_B \sigma_B^2\right), \tag{D9}$$

because $Cov(x_i, x_i) = \sigma_i^2$ and $Cov(x_i, x_j) = 0$, $j \neq i$. In this special case, the covariance between the linear combination defining the posterior and the linear combination defining the weight function is the variance of the posterior itself. [11]

The predicted standard deviation was computed for each participant and each of the two models. Sums of squared errors between predicted and observed performance were transformed to log likelihoods and these were exponentiated to obtain the marginal likelihoods (see Appendix B; no parameters were estimated for either model; hence, $p_i = 0$). The Bayes average was defined as the weighted sum of the predicted variances of the two models (consistent with the mixing of two distributions), with each model weighted by its relative posterior probability (Claeskens & Hjort, 2008; Raftery, 1995; Wasserman, 2000):

$$\Pr(M_j) = \frac{\exp(\ln m_j)}{\sum_k \exp(\ln m_k)}. \tag{D10}$$

The Bayes average was computed for each participant using common relative posterior probabilities, converted to the standard deviation, and then averaged across participants to obtain the means and standard errors in Fig. 9.

References

Allen, G. A. (1999). Spatial abilities, cognitive maps, and wayfinding: Bases for individual differences in spatial cognition and behavior. In R. G. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 46–80). The Johns Hopkins University Press.

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

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd ed.). : Springer-Verlag.

Boorstin, D. (1983). *The discoverers*. Random House.

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin, 138*(3), 389–414.

Bromiley, P. A. (2018). *Products and convolutions of Gaussian probability density functions*. TINA Memo 2003-003. Division of Informatics, Imaging and Data Sciences, School of Health Sciences, University of Manchester.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Cengage.

Chen, X., He, Q., Kelly, J. W., Fiete, I. R., & McNamara, T. P. (2015). Bias in human path integration is predicted by properties of grid cells. *Current Biology, 25*, 1771–1776.

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

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

Chrastil, E. R., Nicora, G. L., & Huang, A. (2019). Vision and proprioception make equal contributions to path integration in a novel homing task. *Cognition, 192*, Article 103998. https://doi.org/10.1016/j.cognition.2019.06.010

Claeskens, G., & Hjort, N. L. (2008). *Model selection and model averaging*. Cambridge University Press.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society B, 4*(Suppl. 1), 102–118.

Cooke, J. H. (1883). *The shipwreck of Sir Cloudesley Shovell, on the Scilly Islands in 1707*. Paper presented at the Meeting of the Society of Antiquaries. http://www.hmssurprise.org/shipwreck-sir-cloudesley-shovell

DeGroot, M. H. (1970). *Optimal statistical decisions*. McGraw-Hill.

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

Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience, 14*, 429–442.

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

Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 195–215.

Friedman, A., Ludvig, E. A., Legge, E. L. G., & Vuong, Q. C. (2013). Bayesian combination of two-dimensional location estimates. *Behavioral Research, 45*, 98–107.

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

Gagnon, K. T., Thomas, B. J., Munion, A., Creem-Regehr, S. H., Cashdan, E. A., & Stefanucci, J. K. (2018). Not all those who wander are lost: Spatial exploration patterns and their relationship to gender and spatial memory. *Cognition, 180*, 108–117.

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

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116*(2), 439–453.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51*(7), 771–781.

Golledge, R. G. (1999). Human wayfinding and cognitive maps. In R. G. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 5–45). Johns Hopkins University Press.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge handbook of computational cognitive modeling* (pp. 59–100). Cambridge University Press.

Halberda, J. (2016). Epistemic limitations and precise estimates in analog magnitude representation. In D. Barner & A. S. Baron (Eds.), *Core knowledge and conceptual change* (pp. 171–190). Oxford University Press.

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual

---

[11] The value of the weight function can be formalized as, $\frac{1}{2}x_V + \frac{1}{2}x_B + e$, with $e \sim N(0, c^2)$. Because e is uncorrelated with x_V and x_B and does not exist in the posterior (i.e., it has coefficient 0), it cannot contribute to the covariance and can be ignored.

differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*(2), 151–176.

Hegarty, M., Richardson, A. E., Montello, D. R., Lovelace, K., & Subbiah, I. (2002). Development of a self-report measure of environmental spatial ability. *Intelligence, 30*, 425–447.

Hogg, R. V., & Craig, A. T. (1970). *Introduction to mathematical statistics*. Macmillan.

Huttenlocher, J., Hedges, L. V., Corrigan, B., & Crawford, L. E. (2004). Spatial categories and the estimation of location. *Cognition, 93*, 75–97.

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

Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology, 52*(2), 93–129.

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

Jo, S., Kim, J. Y., & Han, S. W. (2021). Top-down control of attention under varying task loads. *Acta Psychologica, 216,* Article 103310. https://doi.org/10.1016/j.actpsy.2021.103310

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

Knill, D. C., Kersten, D., & Yuille, A. L. (1996). Introduction: A Bayesian formulation of visual perception. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 1–21). Cambridge University Press.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLOS ONE, 2*(9), Article e943. https://doi.org/10.1371/journal.pone.0000943

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences, 10*(7), 319–326.

Lakshminarasimhan, K. J., Petsalis, M., Park, H., DeAngelis, G. C., Pitkow, X., & Angelaki, D. E. (2018). A dynamic Bayesian observer model reveals origins of bias in visual path integration. *Neuron, 99*, 194–206.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research, 35*(3), 389–412.

Laquitaine, S., & Gardner, J. L. (2018). A switching observer for human perceptual estimation. *Neuron, 97*, 462–474.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling*. Cambridge University Press.

Lindberg, E., & Gärling, T. (1981). Acquisition of locational information about reference points during blindfolded and sighted locomotion: Effects of a concurrent task and locomotion paths. *Scandinavian Journal of Psychology, 22*, 101–108.

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. In R. G. Golledge (Ed.), *Wayfinding behavior: Cognitive mapping and other spatial processes* (pp. 125–151). Johns Hopkins University Press.

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences, 16*(10), 511–518.

Ma, W. J. (2019). Bayesian decison models: A primer. *Neuron, 104*, 164–175.

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

Maloney, L. T., & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research, 50*, 2362–2374.

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research, 41*, 2653–2668.

Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 13–36). MIT Press.

May, W. E. (1960). The last voyage of Sir Clowdisley Shovel. *Journal of Navigation, 13*, 324–332.

Meilinger, T. (2008). The network of reference frames theory: A synthesis of graphs and cognitive maps. In C. Freksa, N. S. Newcombe, P. Gärdenfors, & S. Wölfl (Eds.), *Spatial cognition VI* (pp. 344–360). Springer.

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

Montello, D. R., Lovelace, K. L., Golledge, R. D., & Self, C. M. (1999). Sex-related differences and similarities in geographic and environmental spatial abilities. *Annals of the Association of American Geographers, 89*(3), 515–534.

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

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

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

Norton, E. H., Acerbit, L., Ma, W. J., & Landy, M. S. (2019). Human online adaptation to changes in prior probability. *PLoS Computational Biology, 15*(7), Article e1006681.

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

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

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.

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences, 19*(5), 285–293.

Petzschner, F. H., Maier, P., & Glasauer, S. (2012). Combining symbolic cues with sensory input and prior experience in an iterative Bayesian framework. *Frontiers in Integrative Neuroscience, 6*(Article 58), 1–18.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111–163.

Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research, 66*, 312–323.

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflicts: A strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society B, 273*, 2159–2168.

Robert, C. P. (2007). *The Bayesian choice* (2nd ed.). Springer.

Rohde, M., van Dam, L. C. J., & Ernst, M. O. (2016). Statistically optimal multisensory cue integration: A practical tutorial. *Multisensory Research, 29*(4/5), 279–492.

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

Rudin, W. (1976). *Principles of mathematical analysis* (3rd ed.). McGraw-HIll.

Sampaio, C., Jones, M., Engelbertson, A., & Williams, M. (2020). Bayesian priors in estimates of object location in virtual reality.

Psychonomic Bulletin & Review, 27, 1309–1316. https://doi.org/10.3758/s13423-020-01782-z

Seilheimer, R. L., Rosenberg, A., & Angelaki, D. E. (2014). Models and processes of multisensory cue combination. *Current Opinion in Neurobiology, 25,* 38–46.

Shelton, A. L., Marchette, S. M., & Furman, A. J. (2013). A mechanistic approach to individual differences in spatial learning, memory, and navigation. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 59, pp. 223–259). Elsevier.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32,* 1248–1284.

Shön, T. B., & Lindsten, F. (2011). Manipulating the multivariate Gaussian density (Technical report). Division of Automatic Control, Linköping University.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366.

Sjolund, L. A., Kelly, J. W., & McNamara, T. P. (2018). Optimal combination of environmental cues and path integration during navigation. *Memory & Cognition, 46,* 89–99.

Smith, J. Q. (1988). *Decision analysis: A Bayesian approach.* Chapman and Hall.

Souman, J. L., Frissen, I., Sreenivasa, M. N., & Ernst, M. O. (2009). Walking straight into circles. *Current Biology, 19,* 1538–1542.

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 of the United States of America, 108*(3), 1152–1157.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning, and statistical decision theory. *Trends in Cognitive Sciences, 12*(8), 291–297.

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

Vandekerckhove, J., Rouder, J. N., & Kruschke, J. K. (2018). Editorial: Bayesian methods for advancing psychological science. *Psychonomic Bulletin & Review, 25*(1), 1–4.

Wang, L., Mou, W., & Dixon, P. (2018). Cue interaction between buildings and street configurations during reorientation in familiar and unfamiliar outdoor environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 631–644.

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

Wang, R. F., & Brockmole, J. R. (2003). Human navigation in nested environments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 398–404.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44,* 92–107.

Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2014). Variations in cognitive maps: Understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(3), 669–682.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598–604.

Wozny, D. R., Beierholm, U. R., & Shams, L. (2010). Probability matching as a computational strategy used in perception. *PLOS Computational Biology, 6*(8), Article e1000871.

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

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge University Press.

Zaki, J. (2013). Cue integration: A common framework for social cognition and physical perception. *Perspectives on Psychological Science, 8*(3), 296–312.

Zhang, H., Maddula, S. V., & Maloney, L. T. (2010). Planning routes across economic terrains: Maximizing utility, following heuristics. *Frontiers in Psychology, 1,* Article 214.

Zhang, L., Mou, W., Lei, X., & Du, Y. (2019). Cue combination used to update the navigator's self-localization, not the home location. Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication. https://doi.org/10.1037/xlm0000794

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

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