



# Finding Home: Landmark Ambiguity in Human Navigation

Simon Jetzschke<sup>1,2\*</sup>, Marc O. Ernst<sup>2,3</sup>, Julia Froehlich<sup>1</sup> and Norbert Boeddeker<sup>1,2</sup>

<sup>1</sup> Department of Biology, Cognitive Neuroscience, Bielefeld University, Bielefeld, Germany, <sup>2</sup> Cognitive Interaction Technology–Cluster of Excellence, Bielefeld University, Bielefeld, Germany, <sup>3</sup> Appl. Cognitive Psychology, Faculty for Computer Science, Engineering, and Psychology, Ulm University, Ulm, Germany

Memories of places often include landmark cues, i.e., information provided by the spatial arrangement of distinct objects with respect to the target location. To study how humans combine landmark information for navigation, we conducted two experiments: To this end, participants were either provided with auditory landmarks while walking in a large sports hall or with visual landmarks while walking on a virtual-reality treadmill setup. We found that participants cannot reliably locate their home position due to ambiguities in the spatial arrangement when only one or two uniform landmarks provide cues with respect to the target. With three visual landmarks that look alike, the task is solved without ambiguity, while audio landmarks need to play three unique sounds for a similar performance. This reduction in ambiguity through integration of landmark information from 1, 2, and 3 landmarks is well modeled using a probabilistic approach based on maximum likelihood estimation. Unlike any deterministic model of human navigation (based e.g., on distance or angle information), this probabilistic model predicted both the precision and accuracy of the human homing performance. To further examine how landmark cues are integrated we introduced systematic conflicts in the visual landmark configuration between training of the home position and tests of the homing performance. The participants integrated the spatial information from each landmark near-optimally to reduce spatial variability. When the conflict becomes big, this integration breaks down and precision is sacrificed for accuracy. That is, participants return again closer to the home position, because they start ignoring the deviant third landmark. Relying on two instead of three landmarks, however, goes along with responses that are scattered over a larger area, thus leading to higher variability. To model the breakdown of integration with increasing conflict, the probabilistic model based on a simple Gaussian distribution used for Experiment 1 needed a slide extension in from of a mixture of Gaussians. All parameters for the Mixture Model were fixed based on the homing performance in the baseline condition which contained a single landmark. from the 1-Landmark Condition. This way we found that the Mixture Model could predict the integration performance and its breakdown with no additional free parameters. Overall these data suggest that humans use similar optimal probabilistic strategies in visual and auditory navigation, integrating landmark information to improve homing precision and balance homing precision with homing accuracy.

**Keywords:** human landmark navigation, multisensory cue integration, landmark integration breakdown, virtual reality, landmark reliabilities, probabilistic navigation model

## OPEN ACCESS

### Edited by:

Martin Giurfa,

UMR5169 Centre de Recherches sur la Cognition Animale (CRCA), France

### Reviewed by:

Ken Cheng,

Macquarie University, Australia

Tobias Meilinger,

Max Planck Institute for Biological Cybernetics, Germany

### \*Correspondence:

Simon Jetzschke

sjetzschke@uni-bielefeld.de

**Received:** 04 January 2017

**Accepted:** 03 July 2017

**Published:** 18 July 2017

### Citation:

Jetzschke S, Ernst MO, Froehlich J and Boeddeker N (2017) Finding Home: Landmark Ambiguity in Human Navigation.

Front. Behav. Neurosci. 11:132.

doi: 10.3389/fnbeh.2017.00132

## INTRODUCTION

Humans often combine different types of information to find back to a goal location. The proprioceptive, kinaesthetic, and vestibular senses generate information about body posture and movement (Marlinsky, 1999; Mittelstaedt and Mittelstaedt, 2001; Kearns et al., 2002; Angelaki and Cullen, 2008; Souman et al., 2009; Fetsch et al., 2010). A strategy that is used by many animals to find back to a starting position involves integrating these cues for different parts of a journey into a vector pointing home (Etienne and Jeffery, 2004). This mechanism is referred to as *path integration*. Path integration, however, is only effective for backtracking to a previously visited place and depends on potentially noisy (imprecise) and biased (inaccurate) cues (Cheung and Vickerstaff, 2010; Jetzschke et al., 2016). Therefore, many animals, including humans, memorize external reference points for navigation, such as landmarks (Cartwright and Collett, 1987; Gillner and Mallot, 1998; Wehner, 2003; Nardini et al., 2008; Dittmar et al., 2014; Zhao and Warren, 2015). It has been proposed that honeybees and ants use visual snapshots when solving homing tasks. Using this mechanism, the current input is matched to a previously acquired template (Cartwright and Collett, 1987), which can contain information about the spatial configuration of all available landmarks (Waller et al., 2000), about patterns in the skyline (Graham and Cheng, 2009; Philippides et al., 2011), or about the fractional position of mass of a visual pattern (Lent et al., 2013), to name just a few possible sources of information. That is, natural environments can be highly cluttered with landmark cues. It is still largely unclear whether navigating animals explicitly identify and extract specific landmark objects or if they memorize the whole surrounding scenery from the home position, which they then use for visual matching (Gillner et al., 2008; Basten and Mallot, 2010; Philippides et al., 2011; Baddeley et al., 2012; Zeil, 2012; Stürzl et al., 2015). In this context it is also unclear how the different sources of information are combined or whether they are all equally important. Certainly distinct visual shapes which provide precise and accurate spatial information are no doubt important (see Collett et al., 2013 for a review on visual guidance and memory), but is there a weighing of information based on how informative the sources of information are. Framed in terms of spatial constraints the different sources of information only inform about the existence of and the relationship to a certain point in space and all sources are equally important. Additionally, any deterministic model would make no prediction on the precision of navigation performance. That is, they do not probabilistically inform about the precision and thus the importance of the information that a landmark or another spatial cue might provide (cf. Weiss, Simoncelli, Adelson, Nat. Neuro. 2002 for a similar discussion in the visual motion domain). Would that really be a reasonable integration strategy?

When landmark cues are unreliable (imprecise) and ambiguous (inaccurate), such matching might be better performed probabilistically with a likelihood function assigned to each individual cue (source of landmark information). What does this imply? The Bayesian framework provides a principled approach how to optimally combine different sources of sensory

information that might contain uncertainty and how to integrate this sensory information with any prior knowledge the observer may have about the task at hand. Thereby, the uncertain and ambiguous sensory information can be expressed in form of a likelihood function, that contains all available information about e.g., distance and orientation of the landmark to the home position. The likelihood function defined as the probability that a certain pattern of sensory signals ( $S$ ) is obtained given the state of the world ( $W$ ):  $P(S|W)$ . In other words, each source of information does not only provide one single value with absolute certainty, such as a given position in space, but an entire probability distribution indicating the likelihood of each value. Combining different sources of information in this framework is simply achieved by multiplying the different likelihood functions (which if all distributions are Gaussians leads to the well-known weighting scheme (e.g., Ernst and Banks, 2002). Most often it is then assumed that the perceptual response is given by the maximum of the resulting combined likelihood function (the maximum likelihood estimate or MLE). Using the Bayesian framework, we could in principle also include prior knowledge. Prior knowledge  $P(W)$  is combine with the likelihood function according the Bayes Theorem, which-in essence -again corresponds to a product of these probability distributions. In this case, it is often assumed that the resulting perceptual response is based on the maximum a posteriori estimate (or MAP estimate). For the purposes of this manuscript, however, we do not consider prior knowledge, which is why we relax to a maximum likelihood estimation (MLE) scheme (for reviews on the application of the Bayesian Framework in cognitive science see e.g., Ernst and Bühlhoff, 2004; Cheng et al., 2007; Körding, 2007).

This study addresses the question how different sources of landmark information that contain uncertainty are combined. To this end, we presented participants with a limited set of artificial landmarks in an indoor experimental environment, which allowed us to fully control all configurational cues that are available for homing. The task in these experiments was to return to a previously visited “home” location, which is surrounded by a varying number of either visual or auditory landmarks at different locations. After an initial learning period, the participants were relocated to a previously unknown place and we asked them to return to their starting position. During the outbound travel, we made sure to eliminate any body-based cues that could help them to find back home using path integration. Each single landmark alone did not provide any directional information on their own, because they looked or sounded alike from all directions and the surroundings were uniform in all directions. Between conditions, we changed the ambiguity of the landmark array by manipulating the number and thus the configuration of the available landmarks. In the second experiment, we further tested participants performance in a landmark relocation task. What strategies do humans use to find home when the cues are uncertain and ambiguous? We compared different landmark layouts analyzing participants’ homing performance and we evaluated whether humans might use a common integration principle, regardless of the modality (visual or auditory) in which the

information was presented. That is, we asked how the navigation performance depended on the number of the landmarks and the modality of presentation. To interpret the behavioral data we applied a probabilistic model based on the maximum likelihood approach (e.g., Ernst and Banks, 2002). From the performance to a single landmark cue we know participants basic homing performance. We therefore asked further whether the MLE model can predict the integration performance based on the single landmark responses without fitting any additional free parameters. In other words, we asked whether humans integrate the available landmark information in a statistically optimal way.

## MATERIALS AND METHODS

Overall 25 participants (17 female) aged between 19 and 30 (mean 23) took part: 10 in the auditory homing study; 15 in the visual homing study of which 5 took part in both visual homing experiments. All participants gave their written informed consent before participating and they were paid 6 € per h. The experiments were approved by the ethics committee of Bielefeld University.

### Conditions and Procedure in Auditory Homing

Participants were blindfolded during all auditory homing conditions (Experiment 1-audition) and received no feedback about their performance. These auditory experiments took place in the sports facility (size  $27 \times 45$  m) of Bielefeld University. Throughout the auditory homing task, we used small battery powered speakers (Technaxx MusicMan Mini Soundstation,  $50 \times 50 \times 50$  mm) as landmarks. They were placed on the ground pointing upwards such that the sound spread omni-directionally and thus provided the same information independent of the direction of approach. All landmarks were calibrated for equal output volume prior to the experiment. Landmarks were not moved within one experimental block. The landmarks played a white noise sounds with a uniform random amplitude modulation. The gain factor for the amplitude modulation varied randomly with 4 Hz between  $g = 0.25$  and 1 (Figure 1A). With this modulation, individual landmarks were distinguishable but not recognizable as such. To aid auditory sound source segmentation, we used an individual landmark sound, which was modulated with a fixed frequency of  $f = 4, 8,$  and 13 Hz in an additional condition (see Supplement Material). This alternative sound enables participants to identify and learn individual landmarks by the frequency of amplitude changes (Figure 1A). The sounds were played continuously throughout the entire condition.

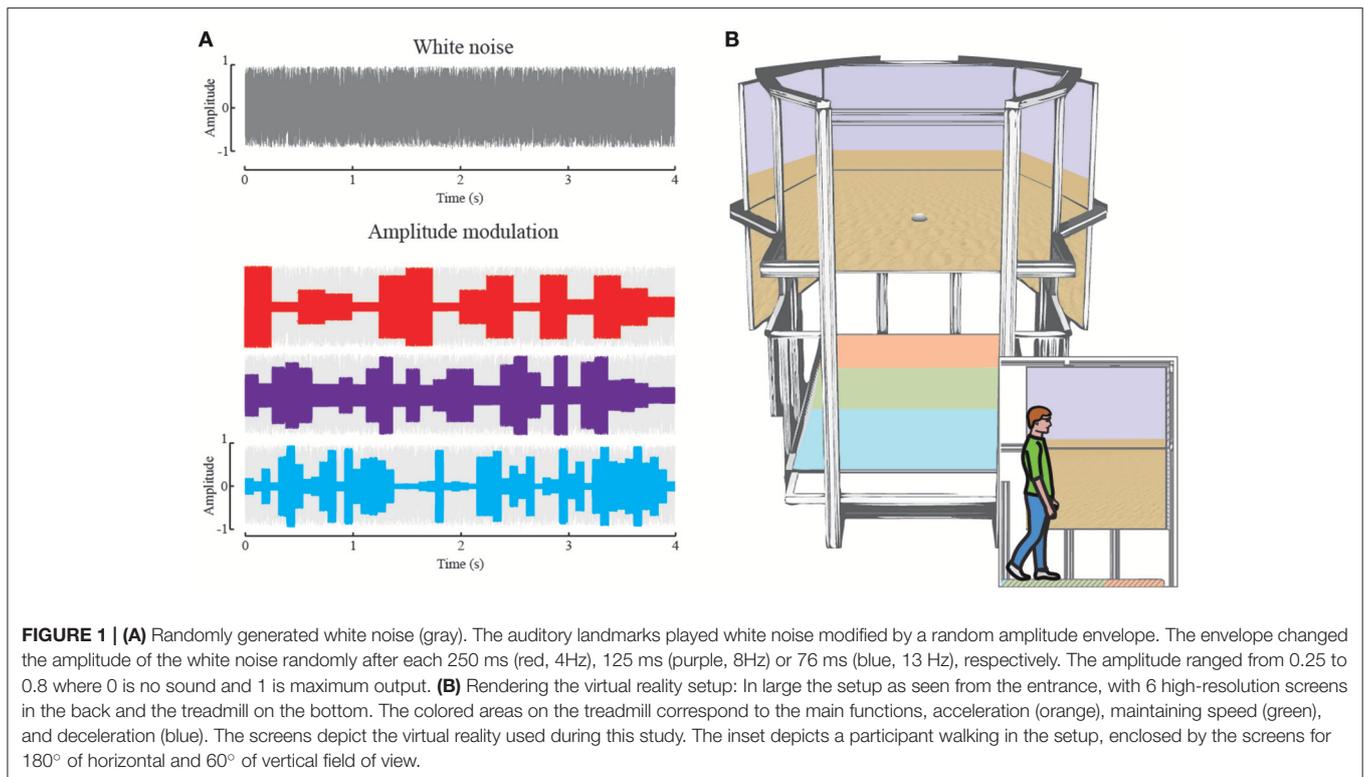
Prior to the experiment we tested participant's ability to discriminate spatially distributed audio sources. Participants were able to identify only up to three auditory objects that played sounds simultaneously. Consequently, increasing the numbers of landmarks did not improve perception as the landmarks became indistinguishable (see Supplement Material). Based on this result, we decided to use a maximum of 3 landmarks.

There were three conditions with either 1, 2, or 3 landmarks placed each at a distance of 2 m around the home position (see Figure 2A). Participants learned this home position in an initial training phase which lasted 60 s. That is, participants were placed at the home position when we switched on the individual landmarks. During this training phase participants were not allowed to walk away, but they were able to freely rotate around their body axis. After the training phase, we immediately started the experimental phase. To get participants from this trained home position to the remote start position for the homing experiment (release point), we displaced them passively on a swivel chair. We took randomly chosen, curvy detours to prevent them from getting useful information about the spatial layout of the environment. From the release point they autonomously had to find back to the trained home position. The swivel chair also allowed for quick turns during the detour to further perturb any reliable input from the vestibular organ. In total, there were four different release points. Upon arriving at a release point, participants left the chair and their task was to walk back to the trained position where they thought home was. When they believed to have reached this position, they stopped walking and we recorded their final position with a goniometer with a top-mounted rangefinder, which was set up at the base of the experimental area. This allowed us to later compute the participants end position of each trial with an error of about 5 cm (roughly 0.5% of the entire path walked). We tested 40 trials for each of the three landmark conditions in the audio homing experiments with a break of 10 min between each condition. The sequence of conditions was randomized for each participant.

Subsequently, we run a fourth condition as a control, which was identical to the previous 3-landmark condition except that now we used three auditory landmarks that played individual sounds each. Here we recorded the participants' final positions with a video camera system which had recently been mounted under the Bielefeld University sporting hall ceiling (Gigabit-Ethernet Camera TMC-1327 GE by Jai with a fish-eye lens by Fujinon, 17). The positions were extracted offline and rectified using custom written software (Wilhelm, 2011) resulting in an error between 3 cm in the central visual field of the camera and up to 5 cm in the peripheral area (Figure 2A).

### Conditions and Procedure in Visual Homing

For the visual homing task (Experiment 1-vision), we used a virtual reality setup that consisted of a treadmill platform (custom  $2.5 \times 1.5$  m belt manufactured by Maschinenbau Kitz, Troisdorf, Germany) and six 65" Sony displays arranged in a 3 by 2 layout (Figure 1B) in front of the participant. The participants' feet and heads were tracked by 14 Optitrack Flex 13 cameras (5 cameras aiming on the feet, 9 cameras aiming on the head area). The position of each foot was tracked by three reflective markers. To control the speed of the treadmill for locomotion in virtual reality, the treadmill was divided into three control zones. While walking in Zone 1 (approximately the front 10% of the treadmill length), the treadmill was accelerating until the participant



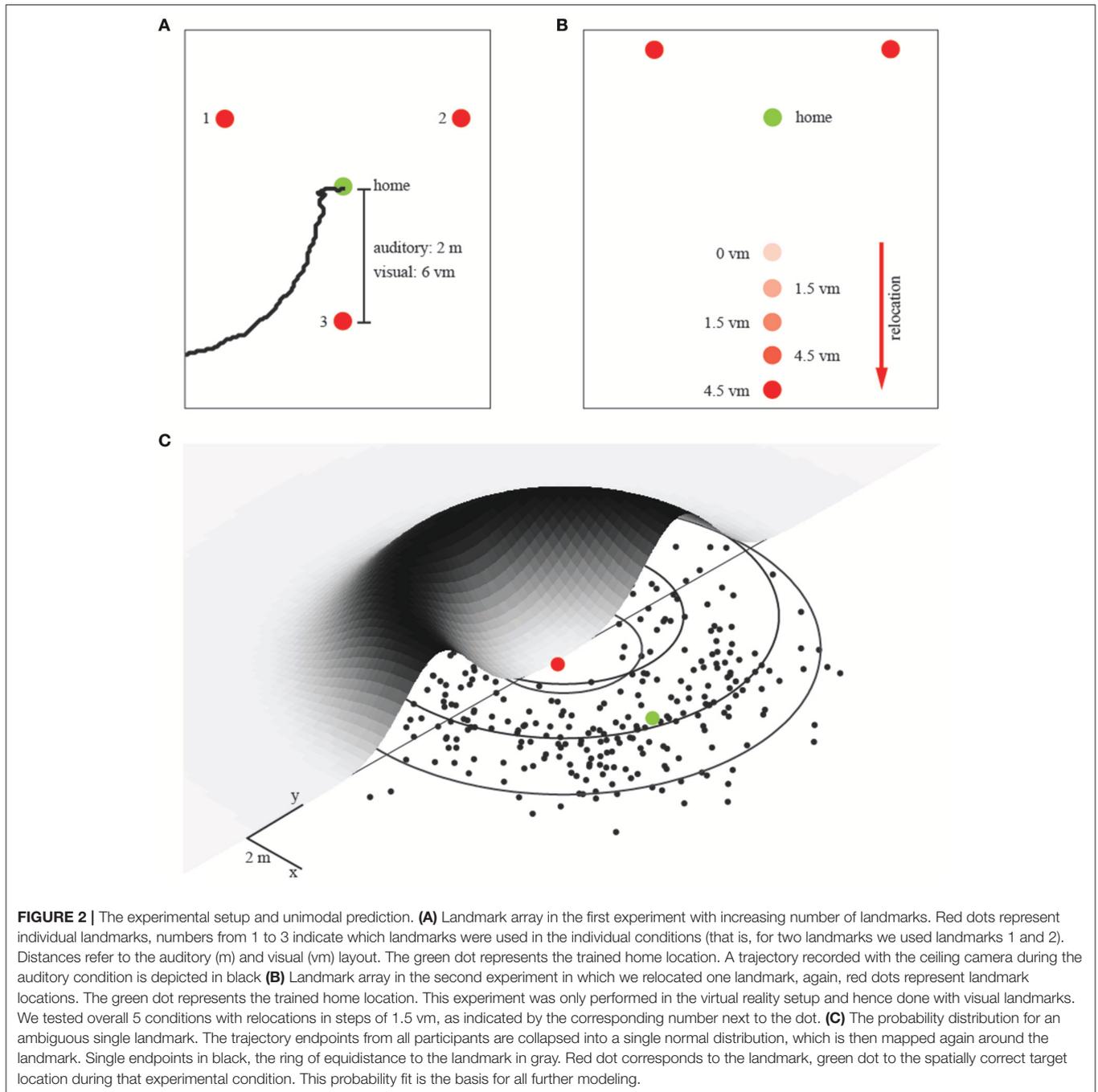
was transported back on the treadmill into Zone 2. While walking in Zone 2 the speed of the treadmill remained constant (approximately the middle 50% of the treadmill length). When the participant slowed down during walking, he/she would be transported further back on the treadmill until they entered Zone 3 of the treadmill in which case also the treadmill slowed down until finally it stopped (approximately last 40% of the treadmill). This way participants could freely adjust their speed of walking and the treadmill would follow their speed. The treadmill allowed only linear forward movements. To control the orientation of visual locomotion, we tracked the position of the participants' head and its orientation with six reflective markers, mounted on a frame worn as glasses. To enable participants to change direction of virtual locomotion, in the virtual environment we used the following control algorithm: When the head was aligned with the treadmill, participants moved straight forward. Deviation from straight-ahead in head orientation was registered by the tracking system and led to rotation of the visual environment in the given direction. To control the directional motion, there was a sensitive angular zone that started at a deviation angle of 22.5° from straight-ahead. That is, there was a central dead-zone where no visual rotation would occur. This was done to allow naturalistic gaze behavior in the frontal body space. The speed of rotation was linearly linked to the heads' deviation from straight-ahead, increasing in speed to a maximum of 45° per second at 90° head rotation. At this speed, it took participants 8 s to do a full turn within the virtual environment, while enabling them to do finer movements within smaller head turn angles. Due to the nature of the linear treadmill and the head-tracking for rotating

the virtual environment, participants could not turn actively on the treadmill. That is, participants were always walking forward within the setup, without the possibility to turn around or walk backwards.

The virtual visual scene was an endless desert, which participants could freely walk through. The sky was colored monochrome blue, offering no directional cues. The whole scenery was rendered in real-time using InstantReality software package (Fraunhofer Institute for Computer Graphics Research). We used small white hemispheres on the ground as local visual landmarks. The overall experimental approach was therefore very similar to the auditory homing experiment.

We tested three different conditions in which we provided either 1, 2, or 3 landmarks placed at 6 (virtual) meters around the home position in the virtual environment (**Figure 1B**). We placed the visual landmarks further apart than the auditory landmarks due to restrictions of the visual field in the virtual setup. Participants learned these positions in a 60 s training phase in the same way as described above for audio homing. That is, they could freely turn around but not walk away. When homing visually, however, participants pressed a button after each trial to be teleported to the next starting position, randomly assigned from a fixed distance of 40 virtual meters to the target location. We tested 40 trials per condition in the visual homing experiments with a break of 10 min between each condition. The order of the conditions was randomly chosen for each participant.

In a second visual homing experiment (Experiment 2), we manipulated the distance between the individual landmarks.



While the overall procedure was exactly the same as in the first visual homing experiment, we changed the position of one single landmark when the participant was teleported to a new starting position (**Figure 2B**). That is, in 5 conditions we relocated one of the three equal looking landmarks between 0 and 6 meters, in steps of 1.5 meters, away in an orthogonal direction to the remaining two landmarks (**Figure 1C**). The order of these 5 displacement-conditions was randomly chosen for each of the 10 participants.

### Modeling

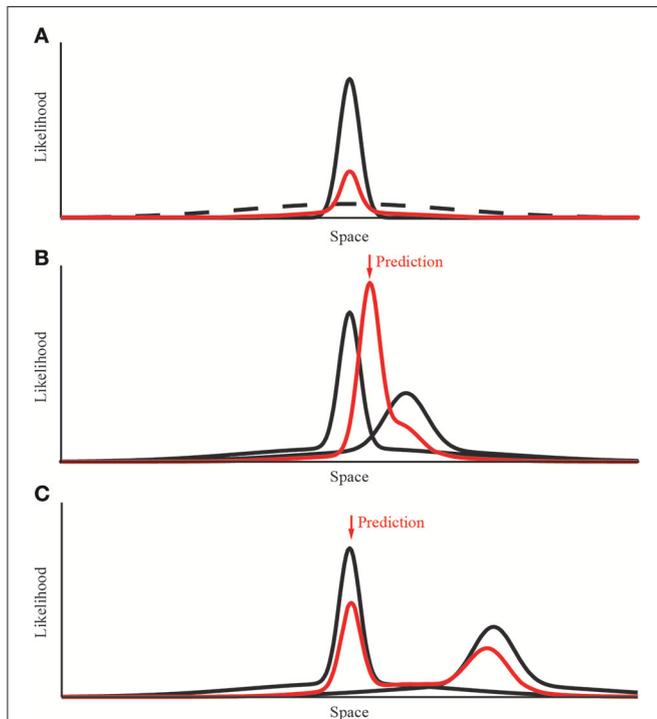
We used probabilistic models based on maximum likelihood estimation to predict human homing behavior. We started with a simple “Gaussian Model” with a probability distribution in form of a 2D donut that has a Gaussian shaped ridge (**Figure 2**). Later, we extended this simple model in order to model the breakdown of integration, replacing the Gaussian ridge by a mixture of two Gaussians (“Mixture Model”). This is because the product of two Gaussians used for modeling the integration will again always

result in a Gaussian, such that it is impossible to model the breakdown of integration using simple Gaussians. However, it has been shown previously that with a tiny modification to the shape of the probability distribution—i.e., changing the Gaussian to a Gaussian with heavy tails—this limitation can be overcome (e.g., Knill, 2007; Ernst and Di Luca, 2011). In practice the heavy tails are best modeled by a mixture of two Gaussians where both of them are aligned and where the second Gaussian that is forming the heavy tails has a very large variance (Figure 3; for more details see below).

The information provided by a single (uniform) landmark is the distance to the home position—it is ambiguous in direction. Thus, the likelihood function provided by a single landmark can be modeled as a ring with a radius corresponding to the distance to home. Because there is uncertainty in the distance cue, the likelihood function corresponds to a donut shaped probability distribution as shown in Figure 2B. To quantify participants' homing performance, we fitted such a 2D, donut-shaped probability distribution to the participants'

spatial distribution of end-positions. The distribution around one landmark is ring-shaped due to the ambiguity constraints (Figure 2B) and is described by two parameters, the distance to the landmark and the variance of the Gaussian distributed ridge. We obtained both parameters from fitting the model to the data in the single landmark condition: (1) we converted the endpoints into polar coordinates with the landmark in the center, and (2) collapsed the data from all directions onto a single dimension, such that we could fit a Gaussian distribution and determine its variance on the distance axis. We call this the "Gaussian" fit. All further probability distributions for different configurations of landmarks are linear combinations (i.e., multiplications) of this uni-landmark distribution. That is, to model the response probabilities for more than one landmark, we linearly combined the response probabilities of multiple single landmarks by multiplying their distributions. This way we can predict the performance with two and three landmarks without fitting any additional free parameters.

In the second visual experiment, we relocated the landmarks between training and test. Also in this case of a landmark relocation, we used the same parameters as in the first experiment. However, as mentioned above to be able to model the breakdown of integration, we need to extend the Gaussian Model slightly by relaxing the assumption that the donut-shaped probability distribution has a ridge that is Gaussian. Instead, we need to replace this Gaussian assumption by a mixture of two Gaussians, which corresponds to a "Mixture Model" (cf. Knill, 2007; Ernst and Di Luca, 2011; Figure 3). Both Gaussians in the Mixture Model are centered on the home-landmark distance from the landmark. However, they differ in variance: the main information about the distance to the home position provided by the landmark is given by the central Gaussian with a variance coding for the uncertainty with which the home position can be localized. The second Gaussian in the mixture has a much larger variance and essentially acts like a pedestal forming heavy tails (Figure 3). So this model has four parameters: (1) the distance home-target which is the same value as in the previous experiments, (2) the variance of the central Gaussian, which is obtained as before by collapsing the data in the uni-landmark condition, (3) the variance of the Gaussian forming the heavy tails, which is achieved by increasing the variance to cover the experimental space. This way, the heavy tails are essentially flat in the working range that is of interest to the experiment. Increasing the variance of this second Gaussian thus makes only marginal differences. Finally (4), the relative contribution of the central and heavy tail Gaussian which is determined by the relative amplitude of both Gaussians. We fixed this parameter at 12% which provides enough energy in the heavy tails to model the breakdown of integration. The exact number of this parameter is not very critical as long as there is enough energy in the heavy tails while at the same time the central Gaussian remains dominant. Given we wanted to determine all parameters from the One-Landmark Condition allowing us to make parameter-free predictions for all the other conditions, we do not have enough "outlier data" from the homing performance in the periphery of our workspace such that we could fit these two "heavy-tail parameters" reliably. This is why for practicality we simply chose



**FIGURE 3 |** Modeling the integration and breakdown of separate landmark cues. **(A)** Construction of the Gaussian mixture probability distribution. The data derived from the experimental condition is shown in black. We added an additional underlying distribution with 10 times the variance and a relative amplitude of 12%. The resulting distribution is normalized after the summation of both distributions. **(B)** Combining the response probabilities for two different cues. When both distributions overlap, the resulting distribution has a decreased variance (red line). The shift along the x-axis is determined by the quality of both distributions, that is the variance of both cues. **(C)** Modeling the breakdown of cue integration is possible due to the mixture of Gaussian probability distribution. When the conflict between both cues is too large their respective probability distributions do not overlap. In this case the prediction returns to the single cue with less variance.

them using a reasonable logic and fixed them henceforth for making predictions. As long as the second Gaussian forming the heavy tails is practically flat in the region of interest—here the workspace—it is sufficient for predicting the breakdown of integration. The modeling is therefore quite robust against the exact choice of variance and amplitude ratio of this second Gaussian as long as it completely covers the navigational space that participants could use. With this minor addition to the traditional MLE approach, the Gaussian Mixture Model is very effective for modeling the breakdown of integration. Such a breakdown of integration has been described before for cues, which are highly conflicting (e.g., Gepshtein et al., 2005).

When different likelihood functions, all made up of the same mixtures of Gaussians, are combined by multiplication, the point of the largest resulting likelihood determines the location that participants should aim for (maximum likelihood estimation). This mixture model can naturally predict the breakdown of integration because of the probability in the heavy tails of the likelihood functions. With small conflicts the center distributions dominate and the product between the likelihood functions corresponds to a unimodal weighted average (cf. **Figure 3B**). With large conflicts, however, due to the heavy tails the product of the likelihood functions becomes bimodal—i.e., the two distributions remain separate and not integrated (cf. **Figure 3C**, Knill, 2007; Ernst, 2012).

To emphasize again, all parameters for both the simple “Gaussian” model and the extended “Mixture of Gaussian” model, are derived from the one-landmark conditions. The two- and three-landmark conditions as well as the landmark relocations are parameter-free predictions of the model. That is, for all conditions, the likelihoods derived from the single landmark condition are simply multiplied and the maximum-likelihood estimate is taken as the estimate.

## RESULTS

### Landmark Ambiguity

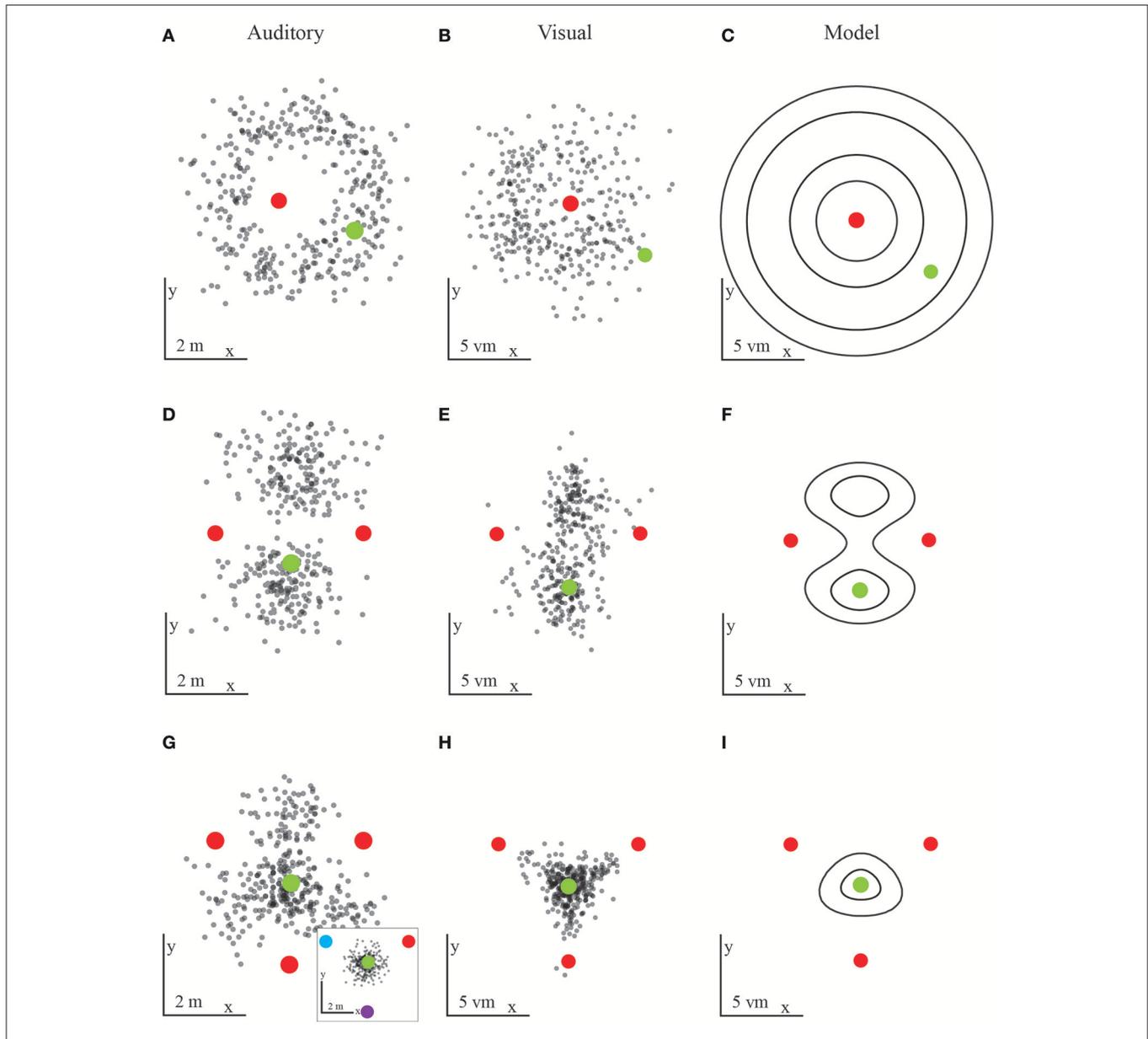
In both, the visual and the auditory experiments we find—maybe unsurprisingly—that participants cannot reliably locate their home position when cues provide ambiguous information. However, as just mentioned above the performance to this baseline condition is needed for fixing all the necessary parameters used for the modeling of the integration behavior of all later conditions. With one omnidirectional uniform landmark, a ring around that landmark represents an infinite number of potentially correct positions with a radius provided by the home-to-landmark distance. This high degree of ambiguity should thus result in responses that fall close to the aforementioned ring around the landmark. This ambiguity constraint is confirmed empirically, both in the auditory and in the visual conditions (**Figures 4A–C**).

When placing a second omnidirectional landmark into the landscape at another location next to the first landmark, we expect the overall ambiguity of the landmark layout to decrease. Because the second landmark is placed at the same distance to the home position as the first one, we expect two ambiguous locations to remain. That is, due to the spatial constraints we

predict that the likelihood functions of both landmarks should form a ring of ambiguity around them with a radius defined by the home-landmark distance. The rings formed by the likelihood functions intersect at exactly two positions (unless home is right on the line between the two landmarks, in which case the rings just touch). Thus, the product of the two likelihood rings reduces to a bimodal distribution with its two peaks corresponding to two equally likely positions. These two positions are the learned home position and a second mirror-symmetric position on the other side of the landmarks. Since both landmarks are similar and provide no direction information by themselves, the remaining ambiguity should lead to responses scattered around these two positions. Similarly, the variance of the responses is decreasing when trajectories lead to only two ambiguous locations. This is again true for both the auditory and the visual homing task (**Figures 4D–F**). With decreasing ambiguity, we thus see increasing accuracy and a decrease in variance. We define accuracy as the distance of all responses from the home location and precision as the spread in the data around the mean. For the Gaussian model this spread corresponds to the standard deviation (i.e., the square root of the variance) of the Gaussian. Therefore, unlike any deterministic model, the probabilistic models used here make clear predictions for both the accuracy and the precision of the homing performance (**Figures 5A–D**).

We chose a third scenario with three landmarks which by their configuration provide unambiguous information with respect to a home location. That is, we chose three omnidirectional landmarks forming a triangle with the home location in its center. In this unique case, all landmarks are evenly distributed and equidistant. That is, the rings formed by the likelihood functions of each landmark intersect at exactly one position and thus the product of the likelihood functions provides an unambiguous unimodal peak in the center of the three landmarks. For the visual homing task with the three landmarks we observed the predicted behavior: That is, with this layout all participants could accurately and precisely locomote to the learned home position (**Figures 4G–I**). By contrast, the accuracy for the auditory homing task was predicted less well. Instead, for the auditory condition with three identical landmarks we found a noticeable difference compared to the performance observed in the visual homing task. Therefore, it seems that the configuration of three identical auditory landmarks is not as unambiguous as it was in the visual task. We come to this conclusion because many responses also scatter in-between always two individual landmarks and they are not concentrated in the center of the three landmarks. Finding responses in the areas between two landmarks does not fit the unambiguous geometry of the learned home position given the three landmarks, but rather resembles the data from two landmark experiments.

Why is that? In audition it is harder to segment different sources producing comparable sounds into separate objects. As such, in audition the spatial layout of the scene is much more difficult to perceive. To help auditory segmentation, we next played sounds with a different frequency of amplitude modulation from each of the three landmarks. This way the individual sound sources became identifiable, which added directional information to the landmark array and further



**FIGURE 4 |** Results from the first experiment with auditory (**A,D,G**) and visual (**B,E,H**) landmarks. Each black dot is the final position of one participants’ trajectory, red dots mark landmark locations, and green dots the target location. The predictions (**C,F,I**) are based on the probabilities to reach a certain position. (**A,B**) Results for one landmark in the auditory (**A**) or visual (**B**) condition. The results fall within a ring around the landmark, as predicted by spatial constraints. The probabilistic model is based on the visual condition performance and implies an underlying Gaussian distribution (**C**). Human participants display large variability in the distance to the landmark. (**D,E**) Results for two landmarks in the auditory (**D**) or visual (**E**) condition. The prediction shows similar regions with similarities to the target locations (**F**). Here the probabilistic model combines two ambiguous landmarks by combining the rings obtained for one individual landmark. (**G,H**) Results for three equal landmarks in the auditory (**G**) or visual (**H**) condition. In the auditory conditions participants could not identify the target location in all trials. The inlayed figure shows the result for three unique auditory landmarks. When every landmark played a slight modified sound participants’ performance increased and is similar to the visual performance. The probabilistic model combines the responses from three individual landmarks in a way closely resembling the human performance (**I**).

reduced ambiguity. The predicted search position, which is based on the geometry of the setup remains the same, i.e. the center of the three landmarks. Given that the three auditory landmarks were now relatively easy to segment, we found that participants could locate the previously learned home location with high

reliability and similar accuracy and precision as in the visual homing task (**Figure 4G**- inlay).

The probabilistic model predicts the probabilities to reach a certain location which coincides well with the participants’ behavior. By measuring the performance distribution with

respect to only one landmark, we can reliably model the response to alternating numbers of ambiguous cues, both in the endpoints of trajectories (accuracy) and the variance of the data (precision). We find an increase in accuracy (i.e., decrease in the average distance to home) along with the number of landmarks (Figures 5A,B). Similarly, we find a decrease in variance (Figures 5C,D). The one condition which did not follow this pattern of results, i.e., which was different from the model predictions, was the abovementioned three audio landmark condition where all three auditory landmarks were identical. As discussed already earlier the reason for this deviance can be found in the difficulty to segment audio sound sources that are identical.

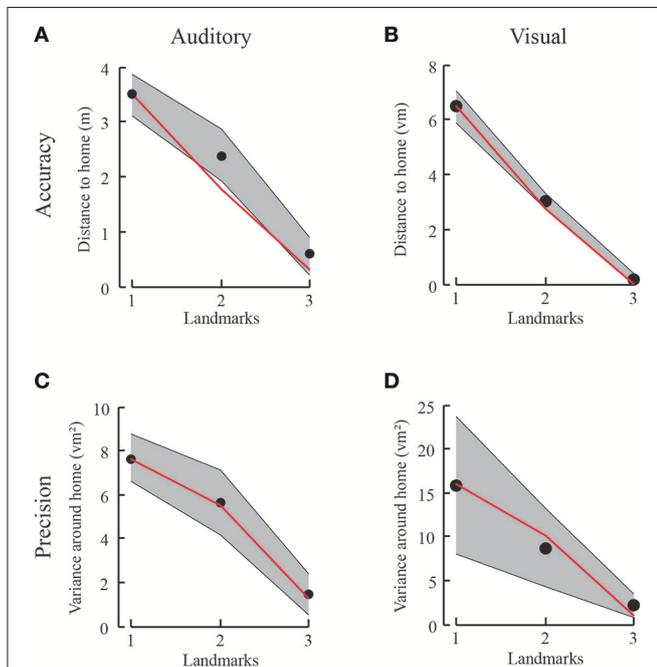
We test individually for all participants and both auditory and visual conditions whether the distance to the home location is decreasing with increasing landmark number. We compare the regression slope across the three conditions to the alternative hypothesis of having no trend [student's *t*-test,  $t_{(9)} = -31$ ,  $p < 0.001$ , slopes between  $-3.48$  and  $-2.47$  when fitting a linear regression model (Figure 5)]. We find a significant increase in accuracy, that is, the distance between the correct home location and the end of each trajectory is decreasing. This is also the case for the decrease in variance [student's *t*-test,

$t_{(9)} = -4$ ,  $p < 0.01$ , for all participants, slopes between  $-13.32$  and  $-3.51$ ]. Comparing accuracy and variance between experimental data and the "Gaussian Model" predictions show no significant difference for any landmark combination. Here we tested prediction and data for each condition: One-tailed paired-sample *t*-test, values between  $t_{(9)} = 0.37$ ,  $p = 0.71$  and  $t_{(9)} = -1.75$ ,  $p = 0.25$  for accuracy. One-tailed paired-sample *t*-test, values between  $t_{(9)} = -0.71$ ,  $p = 0.75$  and  $t_{(9)} = 0.92$ ,  $p = 0.19$  for variance. Power analyses with the given effects indicate that we would need over 100 participants to reach a significance level of 0.95 when comparing the behavioral accuracy with the prediction from the model, and over 2,000 participants for comparing behavioral and the model predicted precision. These analyses together support our claim of a non-significant difference between the model and the empirical data. Therefore, we conclude that the information from three landmarks is combined using individual landmark cues in a fashion indistinguishable from optimal considering that the decision is based on the maximum likelihood estimate. Since none of the deterministic models that e.g., rely on distance or direction information makes any prediction for the precision of the homing performance, it here makes no real sense to perform any model comparison with these models. Therefore, we decided to omit such a comparison.

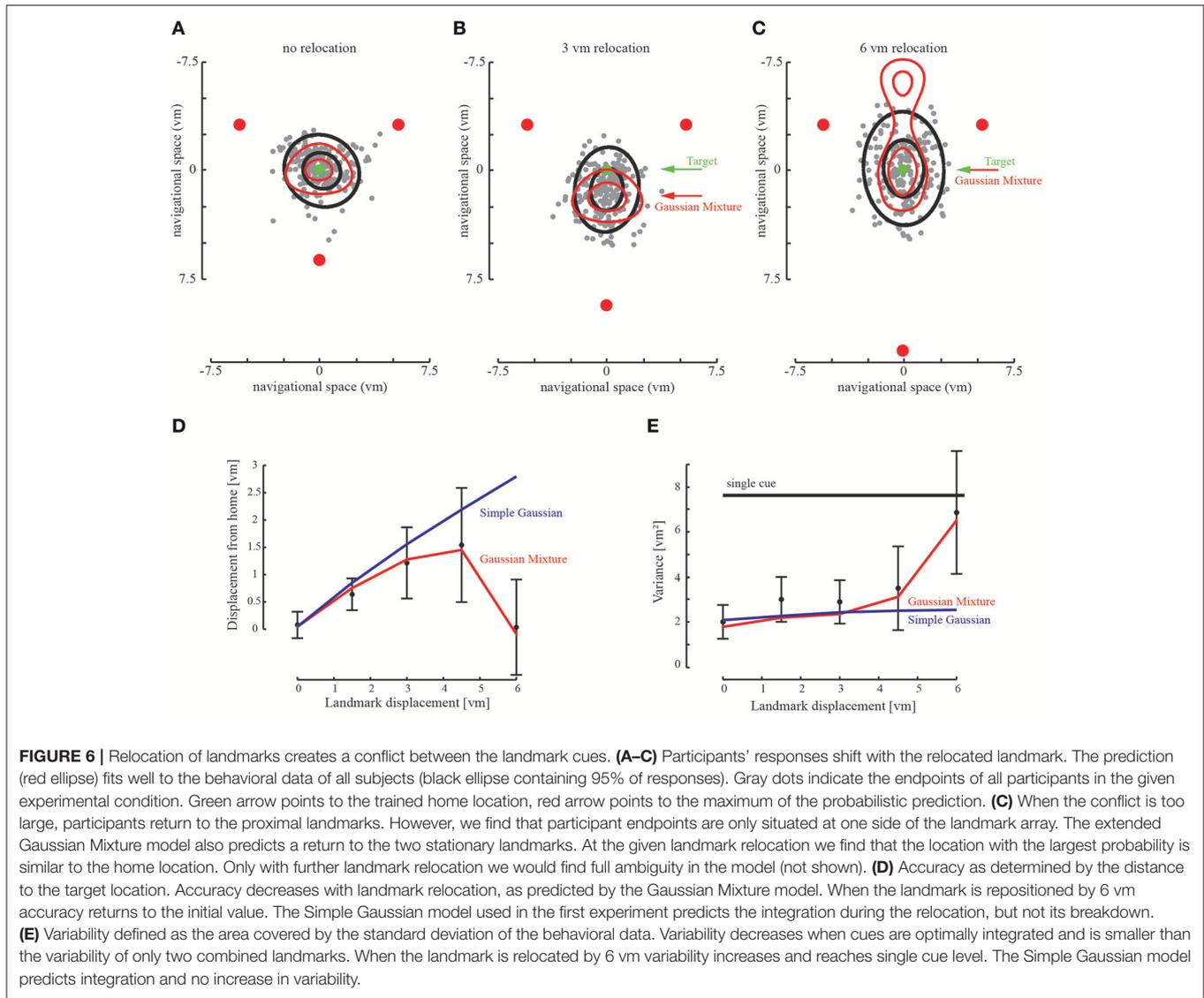
In the following experimental condition, we need to extend the simple "Gaussian" model to capture the breakdown of integration. This is done using the Mixture Model described above. To verify that this extended model does not change the results of these 1, 2, and 3-Landmark Conditions, we also tested the prediction of the extended Mixture Model against our results and found no significant difference when comparing the two model predictions [two-tailed paired-sample *t*-test  $t_{(2)} = 2$ ,  $p = 0.18$  for accuracy and  $t_{(2)} = 0.31$ ,  $p = 0.78$  for variance in all landmark conditions]. That is, for this "Number of Landmark" Experiment the additional parameters of the extended Mixture Model do not make a difference.

### Conflicts in Landmark Reliability

To further examine how spatial information from different landmarks is combined, we introduced changes in the visual landmark array between training of the home position (i.e., remembering the landmark configuration relative to home) and the respective test to find back home. Particularly, we were interested in how these changes affect the homing performance as this reveals how the structure of landmarks is integrated into a whole for navigation. To this end, we relocated one landmark outwards between training and test (Figures 6A,B). These changes were introduced very subtly and thus went unnoticed by the participant, which we confirmed by debriefing the participants after the experiment. We found that with small changes participants average homing performance is affected by the relocated landmark. That is, their locomotion end points are not centered on the learned home position but pulled toward the relocated landmark, which marks the new center of the rearranged three landmarks. Thus, this behavior with small relocations confirms that the landmark information is integrated into a whole spatial configuration with the new home



**FIGURE 5 |** Results from the first experiment with auditory (A,C) and visual (B,D) landmarks. (A,B) Accuracy and model prediction measured from participants' performance in the auditory (A) or visual (B) condition for each number of landmarks. Black dots depict the mean accuracy for a given condition with gray shaded standard deviation. The red line corresponds to the probabilistic prediction based on the results from only one landmark. (C,D) Variance in goal location performance for participants' behavior in the auditory (C) or visual (D) condition for each number of landmarks in black and model prediction in red. The variance decreases with increasing number of landmarks, just as predicted by the mixture of Gaussian probabilistic model.



position defined by the maximum likelihood position in its center.

Probably even more interesting, this behavior is reverted when the third landmark is relocated more than 5 virtual meters between training and test. With such large changes, we observed that participants reverted to use the two landmarks which remained at the initial location, ignoring the third deviant landmark (Figures 6C,D). This is a clear sign of break-down of integration of the spatial layout information and a vetoing of the third deviant landmark. To model these results, we used the same probabilistic approach described earlier based on maximum likelihood estimation, but this time using the mixture of Gaussians as the likelihood function (Mixture Model) instead of the simple Gaussian Model. The goal was again to predict participants' behavior in the three landmarks condition based on empirical data from the homing behavior in the one landmark

condition of the previous experiment without any additional free parameters.

When the third landmark is relocated by a small amount the Mixture Model predicts that the homing behavior of participants follows the deviant landmark and as such according to this model the homing performance should be determined by the new center of the landmark configuration after relocation. With small relocations, when the central Gaussians of the individual probability distributions still largely overlap, we find that the empirical data follow this prediction very closely.

When the third landmark is further relocated (beyond 5 virtual meters), the conflict between training and test of the available landmark cues is increased. According to the model, with larger relocations the overlap between the center Gaussians of the likelihood functions of the deviating landmark and the other two decreases (cf. Figure 3B). Instead the heavy tails of

the mixture of Gaussians will gain more influence and the product between the likelihood functions will become bimodal (Figure 3C). As a result, we find that the Mixture Model predicts a return to the two adjacent landmarks and not to the single third landmark (Figure 6C). But the two possible locations of the remaining two landmarks are not exactly equally likely as they have been in the Two-Landmark Condition. In contrast, the location on the side of the third landmark is emphasized by the prediction reaching a higher probability, because the heavy tails have an asymmetric influence on both sides of the two unchanged landmarks (Figure 6C). This prediction coincides very well with human behavior. The experimental data indicate that the two unchanged landmarks are not ambiguous when a third landmark is available, even when the conflict, that is, the relocation is large. Even when integration of the third landmark breaks down it is still used for disambiguation of the remaining landmarks.

To compare the model prediction to the empirical homing locations, we compute the difference between the prediction and the average participants' data. We repeat this for accuracy (the distance to the correct home location) as well as precision (the variance of the data, measured by the area of the ellipsoid containing 95% of the participant responses as depicted in Figure 6, or measured by the area of the prediction, covering 95% of the probability). This reveals no significant differences in accuracy [paired-sample  $t$ -test,  $t_{(4)} = 1.16$ ,  $p = 0.31$  on the hypothesis that the difference between model and prediction is 0] and variance [paired-sample  $t$ -test  $t_{(4)} = 2.36$ ,  $p = 0.07$  on the hypothesis that the difference between model and prediction is 0]. Thus, the Mixture Model seems a good description of the human behavioral data.

By contrast, when we model the same relocation data using the simple Gaussian Model, we find a significant difference to the participants' data both in accuracy [paired sample  $t$ -test  $t_{(4)} = -2.85$ ,  $p < 0.05$  on the hypothesis that the difference between model and prediction is 0] and variance [paired-sample  $t$ -test  $t_{(4)} = 4.43$ ,  $p < 0.01$  on the hypothesis that the difference between model and prediction is 0]. This clearly indicates that by contrast to the Mixture Model, the simple Gaussian Model cannot explain the human behavioral data and can thus be rejected. Again, since none of the deterministic models make any prediction for the precision of the homing performance, we decided to also omit such a model comparison here.

In conclusion, when integration breaks down the more reliable landmarks are preferred and the performance returns to the target location, resulting in increased accuracy and variance, as also predicted by the Mixture Model. This suggests that precision is sacrificed for accuracy (Figure 6E). Only when the disambiguating landmark is relocated far enough we expect a return to the two landmark cue probabilities, as the third landmark would not be considered as relevant anymore.

## DISCUSSION

We investigated human homing behavior with different configurations of auditory and visual landmarks. We found that both types of landmarks are used in a similar way for homing.

This suggests that the underlying navigational mechanism might be general and not specific for each sensory modality. Such hints at common mechanisms have been discussed for time, space, and quantity (e.g., Walsh, 2003) and many empirical studies find that cue combination works similar for different modalities (e.g., Trommershauser et al., 2011 for an overview).

## Sensory Processing for Directional Hearing and Seeing

How does the navigation performance depend on the number and type of landmarks? Overall, we found that in the auditory and visual homing tasks, the number of landmarks and thus the degree of ambiguity affects both the accuracy and the precision of human homing behavior. With an increasing number of landmarks, the spatial ambiguity is reduced and the participants seem to be taking advantage of this. However, at first glance there seemed to be one exception: the behavior in the auditory task with three identical landmarks showed that the ambiguity is higher compared to the visual task with three landmarks. As reason we have identified the segmentation problem in audition with identical sound sources, which again introduced spatial ambiguity. Making the sound source identifiable helped solving the segmentation problem and as such reduced ambiguity. In turn, presenting the sources which were individually identifiable increased precision and accuracy in the auditory homing task with three landmarks to a level comparable to visual homing with three landmarks. It might seem surprising that such a seemingly small change in the sound profiles of the auditory landmarks could improve the results to such a degree that it became comparable to the homing performance observed with visual landmarks. A likely reason might have been the differences in the acquisition and processing of spatial information between vision and audition (Alain et al., 2001, see Bregman, 1994 for a synopsis): Vision is spatiotopically organized and might allow for an easy segmentation of non-occluded objects in the visual field. As such, in general humans can discriminate spatially distributed visual landmarks even if they are identical in their appearance. On the contrary, audition is organized in frequency maps and one has to infer spatial location of the sound sources from the difference in sound profiles between the two ears. If two sound sources emit similar frequency content it becomes very difficult for the auditory system to divide the sound sources into individual objects. Instead individual objects with sounds that are very similar might be perceptually merged. Think of a pair of speakers playing two exact same sound signals: the perceived sound is located right in between both speakers. Solving this auditory ambiguity problem can be achieved by separating cues not only by location, but also by the characteristics of the sound profiles. Another difference between the two modalities is that auditory cues can be heard all the time, whereas visual objects can only be updated when they are in the visual field. Interestingly, our preliminary test revealed that participants were able to correctly count at least 3 auditory landmarks with identifiable, yet similar sounds. During the experimental testing, participants were unable to use this knowledge to locate the target location until we separated the cues also by the characteristics

of the sound profiles. In case of multiple cues participants could benefit from redundant information, when cues differ not only in their physical position but also in the profile of the sound envelope.

## Models for Landmark Navigation

We find that participants integrate the available information in a near to optimal fashion in terms of our probabilistic model, by reducing the variability of homing behavior. The behavioral data are very similar to the parameter-free predictions of the probabilistic models, which provide the benchmark for optimal performance in form of an ideal observer. That is, integration of an increasing number of landmarks is well predicted by the Gaussian Model, which is based on multiplying the individual likelihood functions as derived from the single landmark condition. Optimal integration also occurs when one landmark is relocated. However, it breaks down in favor of higher accuracy but less precision when the third landmark is displaced too far, i.e., when the conflict between the learned and the tested configuration becomes too large (Figures 6A–C). However, this breakdown is not described by the simple Gaussian Model. Both, integration of numerous landmarks and also its breakdown is well predicted by the probabilistic MLE integration model assuming a mixture of two Gaussians as the likelihood function (Figures 6D,E). This model parsimoniously solves the problem that arises at larger conflicts between signals where, in order to behave robustly, integration should break down (cf. Knill, 2007; Körding et al., 2007; a discussion of the slightly different models for the breakdown of integration can be found in Ernst and Di Luca, 2011 and in Ernst, 2012).

Many psychophysical studies provide evidence for optimal cue integration in human multisensory perception (e.g., review: Ernst, 2006). The situation is less clear in the context of navigation, even though there is growing evidence that navigating animals also combine cues based on Bayesian principles (Cheng et al., 2007). Studies on ant navigation suggest that different systems are in operation simultaneously integrating different sources of information (Collett, 2012; Legge et al., 2014; Wystrach et al., 2015; Fleischmann et al., 2016) and evidence from landmark navigation studies in pigeons (Sutton, 2002; Blaisdell and Cook, 2005; Cheng et al., 2006) or cue integration in monkeys (Fetsch et al., 2010; Dokka et al., 2015) also indicates close to optimal cue integration.

Still, when it comes to modeling navigation behavior it is often suggested that one navigational system, such as path-integration or snapshot matching are solely used for certain tasks and places during a journey. That is, they are not combined but one system is dominating the behavior at a time (Etienne et al., 1996). In consequence, there are several homing models that focus on “view-based” methods only. In these models an image, which is not necessarily a static snapshot (Cartwright and Collett, 1983, 1987; Collett and Rees, 1997; Zeil et al., 2003; Möller and Vardy, 2006; Graham and Cheng, 2009; Basten and Mallot, 2010; Dittmar, 2011), is taken at a certain position and stored in memory. During homing, this snapshot is constantly compared to the current view of the scenery (e.g., Stürzl et al., 2016). This approach can also be used in autonomous agents, which have

the computing power to do pixel by pixel comparisons (Zeil et al., 2003). These models often rely on geometric constraints, such as angles or distances between landmarks. Waller et al. (2000) tested whether humans use a combination of distance and bearing information to relocate to a target location. Consistent with our data from Experiment 1, he could show that participants mainly rely on distance information to the landmarks. It is noteworthy that one of the participants tested in Experiment 1, displays a behavior in the three visible landmark condition which would fit to the bearing hypothesis by Waller, leading to response shapes pointing to individual landmarks. This is not the case for any of that other 15 participants tested in this Experiment 1. Using the distance and bearing approach by Waller et al. (2000), however, is not enough to predict the breakdown of integration observed in Experiment 2, as it does not predict how to weigh information in case of discrepant information. A different approach from robotics navigation is the average landmark vector model (Lambrinos et al., 1998), which requires geocentric knowledge about the current position, rendering the model immune to landmark ambiguities.

In contrast, our model is based on empirical data, making it possible to determine the probability for localizing a given spatial position with a varying number of landmarks. We do not know how participants identify individual landmarks and for our model this is irrelevant since the only input is the empirical homing performance with one landmark. It is a model for cue combination only and the detailed analysis of characteristic homing errors allowed us to elucidate how humans combine multiple landmark cues and how to model this integration behavior. By contrast a recent study focuses on reconstruction-based hypotheses for a scene-matching, showing that it is possible to use scene information to reconstruct the spatial layout of a navigation setup (Pickup et al., 2013). A comparison between empirical homing data and likelihood maps generated by their homing algorithm, allows the authors to quantify the success of different reconstruction-based models. As in the present study, error distributions in Pickup et al. (2013) varied substantially with changes in scene layout, which is very similar to our study, but Pickup et al. did not vary the number of available landmarks or changed the spatial landmark setup between training and test as we did here. Interestingly, simultaneous localization and mapping (SLAM) algorithms from robotics that involve probabilistic approaches for cue combination, suffer from similar constraints as human navigation performance when the objects to be recognized are ambiguous (Cummins and Newman, 2008; Mullane et al., 2011).

Even though initial models exist on how to implement Bayesian optimal cue integration in population codes (Ma et al., 2006), we do not want to go as far as speculating how such navigational integration behavior might be implemented neuronally in the human brain (Krakauer et al., 2017).

## Cue Integration and Breakdown

Maximum Likelihood Estimates have been widely used to explain behavioral phenomena, as in visual-audio or visual-haptic perception (e.g., Ernst and Banks, 2002; Alais and Burr, 2004). It has also been applied to human locomotion

behavior (Butler et al., 2010). Thus, it seems reasonable that this mechanism could be used for navigation as well. This is supported by findings of Nardini et al. (2008), who showed that optimal landmark cue integration develops with age, such that children below age 10 behave suboptimally. The probabilistic model determines the certainty of using a single landmark cue (i.e., the relevant information the landmark provides with respect to the target location—here: home) and assumes that all further combinations of the same cue are based on multiplications of the single-landmark likelihood distributions. Our Gaussian mixture model is determined by only four parameters: (a) The distance estimate of the target to the directionally ambiguous landmark, (b) the variance of this estimate, (c) the variance of the second Gaussian distribution and (d) the ratio between the two underlying Gaussian distributions (cf methods: modeling). It does not only fit the empirical data for one cue, but also for other combinations, describing the spatial position, as well as its variability.

Another feature of cue integration is that it eventually breaks down, such that observers rely on single-cue estimates instead of the integrated estimate when information between cues becomes largely discrepant. The probabilistic model described here can predict this breakdown by modeling the likelihood provided by each landmark as a mixture of two Gaussian distributions—one with a variance corresponding to the uncertainty in the target-landmark distance estimate and the other one with a very large variance acting as a pedestal, both centered at the target-landmark distance around the landmark.

Quite interestingly, the probabilistic approach we describe here requires an estimate about the quality of information from each landmark. That is, it requires an estimate of the shape of the probability distribution (the variances of the Gaussians and their relative contributions). As experimenters, we can estimate the distribution by fitting the Gaussian distribution to our single landmark data and then predict the performance using multiple landmarks. How the participants acquire these estimates about the landmark uncertainty while navigating the natural world remains unclear. However, there is quite some evidence that participants are able to acquire this information online while performing cue integration tasks (e.g., Ernst and Banks, 2002).

We assume that cue integration and its breakdown are important in natural navigation, where landmarks could be potentially relocated, since other people might have manipulated the environment, while we were away. Using probabilistic models in such cases seems reasonable since we can determine with some probability whether a particular landmark has changed. Such probabilistic models have been applied very successfully in other domains investigating human multisensory perception and integration behavior (Ernst and Banks, 2002; Alais and Burr, 2004; Butler et al., 2010; Fetsch et al., 2010). In this experiment, the participants seem to be using the third landmark for disambiguation, increasing the accuracy, thus balancing precision and accuracy in for homing. This is an optimal solution for such a homing task, in which the goal is to reach the target as accurately and as precisely as possible.

Even though landmark ambiguity may be much less of a problem in a natural scene due to the fact that natural landmark information is less redundant and repetitive than the information provided in our experiment. This reduced cue experiment serves as a proof that humans use the available information in a probabilistic fashion and making optimal decisions about the use of this information for navigation. In a similar way probabilistic information may be used in more natural tasks for example when a potential landmark object occurs too often it most likely becomes negligible as a cue for navigation. As an example consider a tree which is a reliable landmark when encountered as a single object in the desert. However, when you encounter the same tree in the forest, probabilistically it will provide less useful information as a landmark for navigation. Taken together, using such redundant landmarks for empirical tests helps us to determine the underlying mechanisms, which are otherwise hard to uncover. It also seems likely that humans use redundant strategies, such as landmark navigation in combination with path-integration, and homing algorithms to aid navigation in different environments, in which reliable landmarks can be occluded by other objects or where they are sparsely distributed.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the American Psychological Association (APA) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ethics committee of Bielefeld University.

## AUTHOR CONTRIBUTIONS

SJ, ME, and NB designed research. JF and ME designed and built the setup. SJ performed behavioral research. SJ and NB analyzed the data. SJ, ME, and NB wrote the paper. All authors contributed with comments and suggestions to the improvement of the manuscript

## FUNDING

This research was supported by the Cluster of Excellence Cognitive Interaction Technology “CITEC” (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

## ACKNOWLEDGMENTS

We thank Corinna Osterbrink and Daniel Detzmeier for their help in conducting the experiments.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnbeh.2017.00132/full#supplementary-material>

## REFERENCES

- Alain, C., Arnott, S. R., Hevenor, S., Graham, S., and Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12301–12306. doi: 10.1073/pnas.211209098
- Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. doi: 10.1016/j.cub.2004.01.029
- Angelaki, D. E., and Cullen, K. E. (2008). Vestibular system: the many facets of a multimodal sense. *Annu. Rev. Neurosci.* 31, 125–150. doi: 10.1146/annurev.neuro.31.060407.125555
- Baddeley, B., Graham, P., Husbands, P., and Philipides, A. (2012). A model of ant route navigation driven by scene familiarity. *PLoS Comput. Biol.* 8:e1002336. doi: 10.1371/journal.pcbi.1002336
- Basten, K., and Mallot, H. A. (2010). Simulated visual homing in desert ant natural environments: efficiency of skyline cues. *Biol. Cybern.* 102, 413–425. doi: 10.1007/s00422-010-0375-9
- Blaisdell, A. P., and Cook, R. G. (2005). Integration of spatial maps in pigeons. *Anim. Cogn.* 8, 7–16. doi: 10.1007/s10071-004-0223-1
- Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT press.
- Butler, J. S., Smith, S. T., Campos, J. L., and Bühlhoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *J. Vis.* 10, 23–23. doi: 10.1167/10.11.23
- Cartwright, B. A., and Collett, T. S. (1983). Landmark learning in bees. *J. Comp. Physiol.* 151, 521–543. doi: 10.1007/BF00605469
- Cartwright, B. A., and Collett, T. S. (1987). Landmark maps for honeybees. *Biol. Cybern.* 57, 85–93. doi: 10.1007/BF00318718
- Cheng, K., Shettleworth, S. J., Huttenlocher, J., and Rieser, J. J. (2007). Bayesian integration of spatial information. *Psychol. Bull.* 133, 625–637. doi: 10.1037/0033-2909.133.4.625
- Cheng, K., Spetch, M. L., Kelly, D. M., and Bringman, V. P. (2006). Small-scale spatial cognition in pigeons. *Behav. Process.* 72, 115–127. doi: 10.1016/j.beproc.2005.11.018
- Cheung, A., and Vickerstaff, R. (2010). Finding the way with a noisy brain. *PLoS Comp Biol.* 6:e1000992. doi: 10.1371/journal.pcbi.1000992
- Collett, M. (2012). How navigational guidance systems are combined in a desert ant. *Curr. Biol.* 22, 927–932. doi: 10.1016/j.cub.2012.03.049
- Collett, M., Chittka, L., and Collett, T. S. (2013). Spatial memory in insect navigation. *Curr. Biol.* 23, 789–800. doi: 10.1016/j.cub.2013.07.020
- Collett, T. S., and Rees, J. A. (1997). View-based navigation in Hymenoptera: multiple strategies of landmark guidance in the approach to a feeder. *J. Comp. Physiol. A* 181, 47–58. doi: 10.1007/s003590050092
- Cummins, M., and Newman, P. (2008). FAB-MAP: probabilistic localization and mapping in the space of appearance. *Int. J. Rob. Res.* 27, 647–665. doi: 10.1177/0278364908090961
- Dittmar, L. (2011). Static and dynamic snapshots for goal localization in insects? *Commun. Integr. Biol.* 4, 17–20. doi: 10.4161/cib.13763
- Dittmar, L., Stürzl, W., Jetzschke, S., Mertens, M., and Boeddeker, N. (2014). Out of the box: how bees orient in an ambiguous environment. *Anim. Behav.* 89, 13–21. doi: 10.1016/j.anbehav.2013.12.007
- Dokka, K., DeAngelis, G. C., and Angelaki, D. E. (2015). Multisensory integration of visual and vestibular signals improves heading discrimination in the presence of a moving object. *J. Neurosci.* 35, 13599–13607. doi: 10.1523/JNEUROSCI.2267-15.2015
- Ernst, M. O. (2006). "A Bayesian view on multimodal cue integration," in *Human Body Perception From The Inside Out*, eds G. Knoblich, I. M. Thornton, M. Grosjean, and M. Shiffrar (Oxford: Oxford University Press), 105–131.
- Ernst, M. O. (2012). "Optimal multisensory integration: assumptions and limits," in *The New Handbook of Multisensory Processes*, ed B. E. Stein (Cambridge, MA: MIT Press), 1084–1124.
- Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433. doi: 10.1038/415429a
- Ernst, M. O., and Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169. doi: 10.1016/j.tics.2004.02.002
- Ernst, M. O., and Di Luca, M. (2011). "Multisensory perception: from integration to remapping," in *Sensory Cue Integration*, eds J. Trommershaeuser, K. Koerding, and M. S. Landy (Oxford: Oxford University Press), 224–250.
- Etienne, A. S., and Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus* 14, 180–192. doi: 10.1002/hipo.10173
- Etienne, A. S., Maurer, R., and Séguinot, V. (1996). Path integration in mammals and its interaction with visual landmarks. *J. Exper. Biol.* 199, 201–209.
- Fetsch, C. R., DeAngelis, G. C., and Angelaki, D. E. (2010). Visual–vestibular cue integration for heading perception: applications of optimal cue integration theory. *Eur. J. Neurosci.* 31, 1721–1729. doi: 10.1111/j.1460-9568.2010.07207.x
- Fleischmann, P. N., Christian, M., Müller, V. L., Rössler, W., and Wehner, R. (2016). Ontogeny of learning walks and the acquisition of landmark information in desert ants, *Cataglyphis fortis*. *J. Exp. Biol.* 219, 3137–3145. doi: 10.1242/jeb.140459
- Gepshtein, S., Burge, J., Ernst, M. O., and Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *J. Vis.* 5, 7–7. doi: 10.1167/5.11.7
- Gillner, S., and Mallot, H. A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *J. Cogn. Neurosci.* 10, 445–463. doi: 10.1162/089892998562861
- Gillner, S., Weiß, A. M., and Mallot, H. A. (2008). Visual homing in the absence of feature-based landmark information. *Cognition* 109, 105–122. doi: 10.1016/j.cognition.2008.07.018
- Graham, P., and Cheng, K. (2009). Ants use the panoramic skyline as a visual cue during navigation. *Curr. Biol.* 19, 935–937. doi: 10.1016/j.cub.2009.08.015
- Jetzschke, S., Ernst, M. O., Moscatelli, A., and Boeddeker, N. (2016). Going round the bend: Persistent personal biases in walked angles. *Neurosci. Lett.* 617, 72–75. doi: 10.1016/j.neulet.2016.01.026
- Kearns, M. J., Warren, W. H., Duchon, A. P., and Tarr, M. J. (2002). Path integration from optic flow and body senses in a homing task. *Perception* 31, 349–374. doi: 10.1068/p3311
- Knill, D. C. (2007). Robust cue integration: a Bayesian model and evidence from cue-conflict studies with stereoscopic and figure cues to slant. *J. Vis.* 7, 5–5. doi: 10.1167/7.7.5
- Körding, K. (2007). Decision theory: what "Should" the nervous system do? *Science* 318, 606–610. doi: 10.1126/science.1142998
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., and Tenenbaum, J. B. (2007). Causal inference in multisensory perception. *PLoS ONE* 2:e943. doi: 10.1371/journal.pone.0000943
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marín, A., MacIver, M. A., and Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist Bias. *Neuron* 93, 480–490. doi: 10.1016/j.neuron.2016.12.041
- Lambrinos, D., Möller, R., Pfeifer, R., and Wehner, R. (1998). "Landmark navigation without snapshots: the average landmark vector model," in *Proceeding of Neurobiology Conferenece Göttingen*, eds N. Elsner and R. Wehner (Stuttgart: Georg Thieme Verlag).
- Legge, E. L., Wystrach, A., Spetch, M. L., and Cheng, K. (2014). Combining sky and earth: desert ants (*Melophorus bagoti*) show weighted integration of celestial and terrestrial cues. *J. Exp. Biol.* 217, 4159–4166. doi: 10.1242/jeb.107862
- Lent, D. D., Graham, P., and Collett, T. S. (2013). Visual scene perception in navigating wood ants. *Curr. Biol.* 23, 684–690. doi: 10.1016/j.cub.2013.03.016
- Ma, W. J., Beck, J. M., Latham, P. E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438. doi: 10.1038/nn1790
- Marlinsky, V. V. (1999). Vestibular and vestibulo-proprioceptive perception of motion in the horizontal plane in blindfolded man. Estimations of linear displacement. *Neuroscience* 90, 389–394. doi: 10.1016/S0306-4522(98)00450-3
- Mittelstaedt, M.-L., and Mittelstaedt, H. (2001). Idiopathic navigation in humans: estimation of path length. *Exp. Brain Res.* 139, 318–332. doi: 10.1007/s002210100735
- Möller, R., and Vardy, A. (2006). Local visual homing by matched-filter descent in image distances. *Biol. Cybern.* 95, 413–430. doi: 10.1007/s00422-006-0095-3
- Mullane, J., Vo, B.-N., Adams, M. D., and Vo, B. T. (2011). A random-finite-set approach to Bayesian SLAM. *IEEE Trans. Robot.* 27, 268–282. doi: 10.1109/TRO.2010.2101370
- Nardini, M., Jones, P., Bedford, R., and Braddick, O. (2008). Development of cue integration in human navigation. *Curr. Biol.* 18, 689–693. doi: 10.1016/j.cub.2008.04.021
- Philippides, A., Baddeley, B., Cheng, K., and Graham, P. (2011). How might ants use panoramic views for route navigation? *J. Exper. Biol.* 214, 445–451. doi: 10.1242/jeb.046755

- Pickup, L. C., Fitzgibbon, A. W., and Glennerster, A. (2013). Modelling human visual navigation using multi-view scene reconstruction. *Biol. Cybern.* 107, 449–464. doi: 10.1007/s00422-013-0558-2
- Souman, J. L., Frissen, I., Sreenivasa, M. N., and Ernst, M. O. (2009). Walking straight into circles. *Curr. Biol.* 19, 1538–1542. doi: 10.1016/j.cub.2009.07.053
- Stürzl, W., Gria, I., Mair, E., Narendra, A., and Zeil, J. (2015). Three-dimensional models of natural environments and the mapping of navigational information. *J. Comp. Physiol. A* 201, 563–584. doi: 10.1007/s00359-015-1002-y
- Stürzl, W., Zeil, J., Boeddeker, N., and Hemmi, J. M. (2016). How wasps acquire and use views for homing. *Curr. Biol.* 26, 470–482. doi: 10.1016/j.cub.2015.12.052
- Sutton, J. E. (2002). Multiple-landmark piloting in pigeons (*Columba livia*): landmark configuration as a discriminative cue. *J. Comp. Psychol.* 116:391. doi: 10.1037/0735-7036.116.4.391
- Trommershauser, J., Kording, K., and Landy, M. S. (eds.). (2011). *Sensory Cue Integration*. Oxford: Oxford University Press.
- Waller, D., Loomis, J. M., Golledge, R. G., and Beall, A. C. (2000). Place learning in humans: the role of distance and direction information. *Spat. Cogn. Comput.* 2, 333–354. doi: 10.1023/A:1015514424931
- Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002
- Wehner, R. (2003). Desert ant navigation: how miniature brains solve complex tasks. *J. Comp. Physiol. A* 189, 579–588. doi: 10.1007/s00359-003-0431-1
- Wilhelm, P. (2011). *IT-gestützte Leistungsmessung im Sport: System und Anwendung*. PhD-Dissertation. University Library Bielefeld, Hochschulschriften.
- Wystrach, A., Mangan, M., and Webb, B. (2015). Optimal cue integration in ants. *Proc. R. Soc. B* 282:20151484. doi: 10.1098/rspb.2015.1484
- Zeil, J. (2012). Visual homing: an insect perspective. *Curr. Opin. Neurobiol.* 22, 285–293. doi: 10.1016/j.conb.2011.12.008
- Zeil, J., Hofmann, M. I., and Chahl, J. S. (2003). Catchment areas of panoramic snapshots in outdoor scenes. *JOSA A* 20, 450–469. doi: 10.1364/JOSAA.20.000450
- Zhao, M., and Warren, W. H. (2015). How you get there from here interaction of visual landmarks and path integration in human navigation. *Psychol. Sci.* 26, 915–924. doi: 10.1177/0956797615574952

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Jetzschke, Ernst, Froehlich and Boeddeker. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.