Not seeing the forest for the trees: Combination of path integration and landmark cues in human virtual navigation

Jonas Scherer[1*¶], Martin M. Müller[1¶], Patrick Unterbrink[2], Sina Meier[1], Martin Egelhaaf[1], Olivier J. N. Bertrand[1&], Norbert Boeddeker[2&],

1 Department of Neurobiology, Bielefeld University, Germany

2 Department of Cognitive Neuroscience, Bielefeld University, Germany

¶These authors contributed equally to this work.

& These authors also contributed equally to this work.

* jonas.scherer@uni-bielefeld.de

Abstract

Our brain combines sensory inputs from various sources by dynamically weighting cues according to their reliability and relevance for a given task. In our daily lives, one such task is to navigate familiar but also unknown environments, as we move from place to place. Two pertinent cues in navigation are the spatial arrangement of landmarks, and the continuous path integration of travelled distances and changes in direction. Although several modelling studies have shown that Bayesian integration of cues provides a good explanation for navigation in environments with up to three objects at the population level, it remains unclear how navigation of individuals is affected by more complex environments. To investigate how humans process and combine cues of varying reliability about landmarks and path integration in complex environments, we conducted virtual reality experiments. We varied the number of landmarks to study navigation in environments ranging from an open steppe to a dense forest, thus going beyond environments with three landmarks that have been studied in the past. We analysed spatial behaviour at both the population and individual level with linear regression models and developed a computational model, based on maximum likelihood

estimation (MLE), to infer the underlying dynamic combination of cues. With more than three landmarks, individual differences between participants in the use of cues are striking. For some, the addition of landmarks does not worsen their performance, whereas for others it seems to impair their use of landmark information. It appears that navigation success in complex environments depends on the ability to identify the correct clearing around the goal location, suggesting that some participants may not be able to see the forest for the trees while others do.

Author summary

Imagine you wake up in the morning in the middle of a forest. It is chilly, and you have no phone or GPS connection. Leaving your backpack behind, you meander through the terrain in quest for firewood and eventually want to return to your belongings. How do you navigate? We investigated this exact question using a first-person computer game, behavioural data and computational modelling. Naturally, your strategy depends on the features that the landscape offers: deserts are different from forests. In principle, all people can use surrounding objects (landmarks) and keep track of the path they walked (path integration). We found that while some people are good navigators in all different scenarios, from open steppes to dense vegetation, others can't see the forest for the trees.

1 Introduction

Spatial navigation is one of the most crucial behavioural competencies of many animals, including humans. Reliable navigation in diverse spatial tasks is based on the encoding of spatial environments to form internal representations such as cognitive maps or graphs [1–5]. Navigators need to combine different information sources to determine their own location, as well as that of other relevant places. The cognitive processes underlying spatial navigation abilities have been the focus of much research, with two major sources of spatial information being identified as the basis of spatial learning: landmarks [6–9] and path integration (PI) [10–12]. Both are used by non-human and human navigators alike [13].

The use of landmarks most often refers to the usage of external environmental

features to identify and locate places of interest in a navigator's surroundings [7, 9, 14]. Landmarks can be used for guidance towards a goal ("turn left at the gas station") as well as for self-localisation ("I see the Eiffel Tower, so I must be in Paris").

Path integration (PI), or "dead reckoning", refers to the continuous integration of self-motion information (translation and rotation, gathered for example from visual and proprioceptive cues) [15, 16] for the purpose of navigation.

Cue integration in human spatial navigation Combining cues can be a crucial strategy to improve navigation success since the information provided by each individual cue comes with its own limitations. PI accumulates errors, potentially decreasing its usefulness over time, while landmarks might be visible only from some part of the environment or hard to identify in visually ambiguous situations [17]. One effect of landmark ambiguity was demonstrated by [9], who could show that homing performance, i.e. the ability to return back to one's starting position, improved when human navigators were able to make use of multiple local landmarks close to their goal location. The combination of landmark cues could be shown to allow participants to navigate more accurately and precisely when all landmarks indicated the "true" goal location [9]. However, this integration of cues also made them susceptible to being misled, when one of the landmarks was covertly displaced. The participants combined the estimates of their goal location derived from the manipulated landmark with that derived from the other "true" cues and headed towards an intermediate location. This means, that they did not perceive the mismatch between the cues and continued to integrate them.

In cue integration, the relative importance of available sources of information depends on a variety of factors, including availability, task context, environmental structure, and cue reliability. The combination of information sources with different relative importance can be represented well in a Bayesian framework of optimality [18, 19]. In such a framework, each source of information is weighted based on its perceived reliability and can be combined with prior knowledge about certain features of the environment [20]. Since cue reliabilities are directly reflected in Bayesian cue integration, the framework has been extensively used in human spatial navigation research to explain errors in several spatial navigation tasks [6, 7, 21].

The application of a Bayesian framework of optimality to spatial navigation focuses

on two major aspects: First, better performance can be achieved by combining different information sources, as every single cue and its internal processing is subject to noise and errors. Second, cues may be integrated by weighting single cue estimates in proportion to their reliability. As a consequence of this weighting, the combined estimate of multiple cues will be more precise [6], while its overall accuracy will depend most strongly on the accuracy of the estimate derived from the most reliable cue.

Here, **accuracy** refers to the average magnitude of errors made by a navigator trying to solve a spatial task, while **precision** refers to the standard deviation in errors made over multiple attempts. Together, these two error measures form the basis of analysis of several experiments, which validate the prediction of optimal cue integration models and show that humans mostly combine cues in a statistically optimal manner, but are also prone to systematic misjudgements under some conditions [6, 7, 9, 22].

Using homing tasks to investigate cue combination One of the most frequent and important spatial tasks to solve for any navigator is returning to a previously visited location, like an important food source or their home. Success in this task, which is often called "homing", has been shown to depend on the combination of spatial information sources using landmarks and PI [6, 23]. While some animals like desert ants have shown remarkable abilities in PI [13, 24, 25], humans are considered to have only coarse PI and rather focus on available visual features in solving spatial tasks [12, 23]. Accordingly, "homing" tasks have received much attention in the spatial cognition literature to disentangle the combination of information from landmarks and PI in human navigation (e.g. [6, 7, 9, 20, 26]). One of the most well-known homing paradigms in spatial navigation research is the triangle completion task [6, 7, 27, 28]). In a triangle completion task, the participant is actively or passively guided along two sides of a triangle, and at the end of this outbound path asked to either actively return to, or passively point at the starting location [29–31]. Comparing systematic errors in homing or pointing for participants in different conditions (e.g. with or without landmarks [7], longer or shorter paths, or different turning angles [15, 28], or for healthy individuals vs. those suffering from vestibular disorders [32]) allows to draw conclusions about how these factors might have influenced behaviour. Manipulating the reliability of landmark cues in different environments in a triangle completion task is one vital approach to

investigate the use of cues from landmarks and PI in Bayesian cue integration models, as cue weighting therein depends on cue reliability that corresponds to performance accuracy and precision.

Influence of environmental complexity Recent experimental and theoretical studies have started to explore the relation between these systematic errors in cue combination and the characteristics of the environment within which a behaviour was observed. It was shown that the reliability of spatial information can be re-evaluated by navigators depending on spatial and temporal variation of features in an environment [20]. [23] reported that participants often misjudge the reliability of visual cues in a homing task and even continue to use landmark cues for visual guidance when objects were displaced and put in conflict with the internal PI of the participants. However, the studies by [20] and by [23] both also report that if participants successfully identify objects as spatially unstable (i.e. appearing at different locations throughout the experiment), this knowledge is quickly incorporated into the participant's navigation strategy, resulting in reduced trust in landmark-based guidance and increased reliance on PI. These effects of environmental context also seem to be expressed differently in different people, leading to individual differences in short-term learning [33] and the relative reliance on different spatial cues [34].

The findings by [9], [20], and [34] lead us to an important observation: If the use and combination of spatial cues for navigation depends on a) the makeup of the environment and b) the nature of the task at hand, it becomes necessary to more explicitly consider these factors in designing our experiments, if we want to deepen our understanding of the mechanisms that govern (human) spatial cognition. In other words, for a given task, we can categorise different environments based on how easy it will be to solve that type of task in the respective environment. This in turn will depend on different aspects of the makeup of that environment, such as the number, distribution and distinguishability of spatial features in it. We will call the totality of these features the "complexity" of that environment. Building on this, we can ask how changing specific features contributing to the complexity of an environment will affect the ability of human navigators to solve a spatial task and also which strategies they might employ to do it.

As an example, consider a meadow with a single tree on it. If we need to memorise

and return to a location close to that tree, we can assume that the tree will serve as a useful landmark, providing us with information about the relative distance between the goal location and itself. Now consider we add a large number of similar-looking trees to this meadow. Geometrically, the information provided by the original tree did not change. However, when navigating in this new, more complex environment we might on the one hand make use of new, emergent cues, such as spatial configurations of groups of trees. On the other hand, using any given tree or configuration of trees at all might be limited to the degree we can identify the tree(s) in question. This illustrates how cue use can be closely tied to the complexity of a spatial environment.

However, so far, most experimental studies concerned with cue combination and cue preference questions have been conducted in relatively simple environmental settings and have not systematically addressed the question of complexity-induced changes in cue use.

How does homing performance of human navigators depend on the visual complexity of the environment? We aim to answer this question by systematically varying the degree of spatial complexity in a spatial homing task. Our study is guided by two hypotheses when comparing homing behaviour in environments with different numbers of ambiguous landmark objects.

First, if navigators are able to identify individual objects (e.g. based on the spatial configuration of object groups or by disambiguating landmark positions based on PI information), homing errors should decrease as more objects are added, up to a point where the goal location can be completely constrained based on the spatial information derived from landmark cues. This is the situation described by [9], who could show that homing performance was best when participants could fully triangulate the goal from the three present landmark cues.

Second, if objects become ambiguous to the navigator, as could be the case if more objects are added beyond what is needed to geometrically constrain the goal location, homing performance should decrease as more and more ambiguous objects in the environment increase the chances for the navigator to be lead astray. This limits the usefulness of the landmarks cues for homing and in extreme cases, the navigator might rely solely on PI to find the goal.

However, when taking into account results like those of [34], who report large inter-individual differences in cue use even in relatively simple environments, we expect that reliance on the different cues should vary between participants in more complex environments, maybe even to a greater degree than in simple environments. We therefore expect, that homing performance is subject to inter-individual variability and will explicitly include this factor in our analysis.

To systematically assess the effects of environmental complexity on homing behaviour, we have performed a large-scale virtual reality (VR) experiment on desktop computers. The use of VR tools is of special interest for this type of experiment since it allows for the real-time presentation and manipulation of complex, interactive environments. We made use of these capabilities to design a homing task, in which participants were first led away from a given location in the virtual environment and then had to return there. Crucially, this environment differed between experimental conditions, containing different numbers of visually ambiguous trees surrounding the goal location, creating different degrees of clutter. We take these different degrees of clutter to be our proxy for the abstract idea of "environmental complexity" in this study. The experimental setup outlined above allowed us to target the question of how homing performance depends on the visual complexity of an environment by systematically introducing additional, ambiguous objects.

Note that while the trees might serve as landmarks, we cannot directly determine how participants derive spatial information from the trees and will be referring to them only as "objects" throughout the text.

2 Results

Healthy participants (N=23) completed a triangle completion homing task in virtual reality. In a triangle completion task, the participant is guided away from a goal location along two legs of a triangle and then asked to return to the goal location, thus completing the triangle [6, 7, 27, 28]. The characteristic errors made over several repetitions of this task, i.e. the deviations between the true goal location and the locations actually navigated towards by the participant, can then be used to quantify

the homing performance of the participant. In our virtual version of the task, that participants performed remotely on their desktop computers, they were displaced from a campfire that indicated their goal location, by asking them to collect two piles of firewood, which formed waypoints along the outbound legs of the triangle. This path was the same across all environmental conditions (Fig. 1). The start of the return trip was located at a distance of 25 vm to the goal location. Neither the piles of wood nor the campfire was visible during the return, to prevent their use as spatial cues. Only the number of identical tree objects in the surround was systematically varied between zero and 99 (zero, one, two, three, ten, 99). Conditions with higher object counts always contained the objects of conditions with lower object counts as well (Fig. 1). To create optic flow for visual estimation of self-motion we provided identical grass tufts on the ground that were randomly arranged for each trial. We quantify homing performance as position error, which is the Euclidean distance between the actual goal location and the endpoint of the trajectory performed by the participant.

Fig 1. Triangle completion task and environmental complexities across conditions. Environmental complexity is varied in a triangle completion task. (a) Participants are guided away from a goal location along two sides of a triangle. The goal location is indicated in our VR environment by a campfire, and waypoints by piles of firewood, that participants are asked to collect because their campfire went out. Waypoint locations are not changed and participants in all repetitions and conditions walk the same outbound path and are asked to return to their goal location, the campfire, at the end (example trajectory). Environmental complexity is varied by altering the number of available objects between zero, one, two, three, ten, and 99. The conditions with higher numbers of objects contain also the objects from conditions with a lower number of objects. Scale is given in virtual meters (vm), that correspond to actual meters in the VR. Subfigures b, c, and d show screenshots from the VR for conditions with zero, three, and 99 objects, respectively, at the second waypoint in the direction of the goal location.

2.1 General effect of different degrees of clutter

The distribution of trajectory endpoints across all conditions reveals that the number of objects (trees) substantially impacts the position error (Fig. 2). Both error measures, accuracy and precision, improve when one, two, or three objects are added around the goal. When no objects are available, trajectory endpoints are spread widely around the goal position with a median error of almost 10 virtual meters (vm), $\pm$ 5 vm (standard deviation) (see Figures 3 and 4). Virtual meters resemble actual meters in the VR, based on the scaling of visible 3D assets (grass tufts, trees, wood planks, campfire).

When individual objects are added to the scene, accuracy, i.e. the median position error, decreases to a minimum of little more than $2\,vm$ for trials in which the goal location was constrained by three surrounding objects and precision also increases with a median standard deviation of position error as small as $2\,vm$.

However, if further objects are added beyond the initial three, a noticeable trend emerges: for numerous participants, the position error once again rises as the object count reaches ten or 99. This trend is accompanied by a considerable increase in the standard deviation of position errors among different participants (as indicated by the marker colour in Fig. 2 and 4). Consequently, this leads to an overall decrease in precision, bringing it to a level similar to what is observed for environments devoid of any objects. Nevertheless, for several participants, the accuracy and precision stay roughly at the level observed for the three object condition, with one participant even showing the highest precision in the 99 object condition (Fig. 4, bottom). Hence, the number of objects influences performance of individual participants differently (for a quantitative assessment see below).

While we describe the patterns of errors in ascending order from few to many objects present in the visual scene, conditions were presented in random order to the participants during the experiment.

Fig 2. Homing performance of one example participant increases with growing number of objects. Trajectory endpoints of single trials are depicted by red dots. The participant is plotted in the same colour as in all figures. Thereby colour is varied from dark red to light yellow from corresponding low to high median position error of that participant in the 99 object condition (see Fig. 4 top right). Outbound and homing trajectories are indicated by light grey lines. Cutout does not show complete object constellations, please see Fig. 1 for a more zoomed-out view of the environment.

Fig 3. Homing performance at population level increases up to three, and then decreases up to 99 objects. Trajectory endpoints of single trials are depicted as dots. Individual participants are plotted in the same colour in all figures. Thereby colour is varied from dark red to light yellow from corresponding low to high median position error of that participant in the 99 object condition (see Fig. 4 top right). Cutout does not show complete object constellations, please see Fig. 1 for a more zoomed-out view of the environment.

Fig 4. Position error accuracy (top) and precision (bottom) on population level are lowest with three objects. Boxplots are based on median position error in virtual meters (vm) (top) or standard deviation (SD) (bottom) of N=23 single participants with n=24 repetitions within respective conditions (see dots). Participants are plotted in the same colour as in all figures. Thereby colour is varied from dark red to light yellow from corresponding low to high median position error of that participant in the 99 object condition (see Fig. 4 top right). Boxplot notches indicate 95% confidence intervals around the median. Green and purple horizontal bars at the bottom indicate low and high clutter effects (see section "Separating effects of low and high clutter" for details)

2.2 Disentangling the effect of different degrees of clutter

Regression models are commonly used to analyse the relationship between two or more variables to make predictions and identify trends in data. Especially powerful are linear mixed effects models, that include fixed and random effects. Fixed effects refer to a study's factors of interest, while random effects describe randomly sampled factors, such as participant identity. Taking both effect types into account, results in more accurate and reliable estimates of complex relationships in the data.

Based on the distribution of position errors across the conditions, we fit linear mixed effects models to identify the underlying mechanisms that lead to the observed changes in performance. We hypothesised (see introduction section "How does homing performance of human navigators depend on the visual complexity of the environment?") and observe a clear effect of the condition (number of objects) and see an influence of participant identity (see Fig. 3 and 4). Hence, we aim for a model to quantify the influence of both of these factors on position error accuracy and precision. Linear mixed models assume a normal distribution of residuals of the response variable, which in our case is the median position error (accuracy), or its standard deviation (precision). Both the accuracy and precision measures are zero-bounded, and thus their residuals in a regression analysis cannot be expected to be normally distributed. Thus, all further analyses were performed using the log-transformed position error and its log-transformed standard deviation respectively. The predictor variable *condition* (number of objects) was always modelled as ordinal, while *participant identity* was modelled as categorical.

An initial, simple model using only the experimental condition (number of objects) as an ordinal predictor already showed significant predictive power for both accuracy (Tab. 1 model 1) and precision (Tab. 1 model 4) of goal estimates. To investigate

possible individual differences in performance across conditions, we added participant identity as a random predictor (Tab. 1 models 2 and 4), which significantly improved the model for accuracy, and for precision (likelihood ratio test: accuracy: $\chi^2 \approx 21.85$, logLik1=-112.5, logLik2=-122.4, p<0.0001; precision: $\chi^2 \approx 20.49$, logLik1=-202, logLik2=-212.2, p<0.0001).

The residuals in error scores for both models are characterised by a degree of heteroscedasticity, which is caused mainly by the large variance in the 99 object condition (see Fig. 4). Accordingly, residuals for errors are homoscedastic across conditions when the 99 object condition is excluded. However, as moderate heteroscedasticity is not considered to be a major influence on linear mixed models [35], we consider the model family described above as adequate to describe the observed data. Thus, we base our analysis of accuracy on a regression model of the following form:

$$\ln(\mathit{position\,error}) \sim \mathit{ordinal\,condition} + (1 \mid \mathit{participant}) \tag{1}$$

and for the analysis of precision in this form:

$$\ln(\sigma_{\mathit{position\,error}}) \sim \mathit{ordinal\,condition} + (1 \mid \mathit{participant}) \tag{2}$$

with σ being the observed standard deviation of the position error. While the model explains approximately $R^2 \approx 59\%$ of variance in our median position errors, 40% are attributed to our fixed condition effect. For position error variance approximately $R^2 \approx 55\%$ of variance is explained, 36% are attributed to the condition effect. Note that there is no way to calculate R^2-values for mixed effects models, and the values reported above are approximations based on an implementation of [36] with refinements by [37]. The model suggests a linear as well as a quadratic component in the relationship between condition and position error for accuracy (Tab. 1 model 2) and precision (Tab. 1 model 5) that we investigate in more detail below. We observe a trend of reduced position error over sessions (see appendix "S2 Session order effect"), however, including session number as an additional fixed effect in the regression model results in only a slight improvement of fit. As the effect of order of sessions is small and not the primary focus of this study we do not consider it in further analyses.

Table 1. Performance of linear regression models with and without the random effect of participant. Model 1 predicts the position error from the condition (number of objects). Model 2 additionally includes a random participant factor. Model 3 restates model 2 but uses forward difference coding. Models 4 to 6 follow the same structure as 1 to 3, but predict position error variance instead. In all models condition is ordinal and polynomials from 1st to 5th degree are fitted. est.: estimate; s.e.: standard error; t: t value; d.f.:degrees of freedom; p: p-value

$model\ 1 : \ln(position\ error) \sim condition\ ordinal$
$AIC = 258.9, BIC = 279.3$

	est.	s.e.	t	d.f.	p
intercept	1.34	0.05	26.25	132	<0.001
condition ^{1}	-0.79	0.13	-6.30	132	<0.001
condition ^{2}	0.84	0.13	6.69	132	<0.001
condition ^{3}	0.08	0.13	0.63	132	0.53
condition ^{4}	-0.14	0.13	-1.09	132	0.28
condition ^{5}	-0.12	0.13	-0.93	132	0.35

$model\ 2 : \ln(position\ error) \sim condition\ ordinal + (1|participant)$
$AIC = 239, BIC = 262.4$

	est.	s.e.	t	d.f.	p
intercept	1.34	0.08	16.71	23	<0.001
condition ^{1}	-0.79	0.10	-7.79	115	<0.001
condition ^{2}	0.84	0.10	8.27	115	<0.001
condition ^{3}	0.08	0.10	0.78	115	0.44
condition ^{4}	-0.14	0.10	-1.34	115	0.18
condition ^{5}	-0.12	0.10	-1.15	115	0.25

$model\ 3 : \ln(position\ error) \sim condition\ ordinal + (1|participant)$
forward difference coding; $AIC = 239.0, BIC = 262.42$

	est.	s.e.	t	d.f.	p
intercept	1.34	0.08	16.35	24.05	<0.001
1 object	-0.61	0.15	-4.15	120.23	<0.001
2 objects	-0.50	0.15	-6.12	120.23	<0.001
3 objects	-0.38	0.15	-3.98	120.23	0.01
10 objects	0.31	0.15	3.42	120.23	0.04
99 objects	0.28	0.15	3.44	120.23	0.06

Table 1. continued.

$model\ 4 : \ln(\sigma_{position\ error}) \sim condition\ ordinal$
$AIC = 247.2, BIC = 267.7$

	est.	s.e.	t	d.f.	p
intercept	1.00	0.05	20.49	132	<0.001
condition ^{1}	-0.22	0.12	-1.86	132	0.07
condition ^{2}	1.00	0.12	8.33	132	<0.001
condition ^{3}	0.06	0.12	0.49	132	0.63
condition ^{4}	0.02	0.12	0.19	132	0.85
condition ^{5}	0.01	0.12	0.05	132	0.96

$model\ 5 : \ln(\sigma_{position\ error}) \sim condition\ ordinal + (1|participant)$
$AIC = 228.7, BIC = 252.1$

	est.	s.e.	t	d.f.	p
intercept	1.00	0.08	12.90	24.05	<0.001
condition ^{1}	-0.22	0.10	-2.23	120.23	0.03
condition ^{2}	1.00	0.10	9.99	120.23	<0.001
condition ^{3}	0.06	0.10	0.59	120.23	0.56
condition ^{4}	0.02	0.10	0.22	120.23	0.82
condition ^{5}	0.01	0.10	0.06	120.23	0.96

$model\ 6 : \ln(\sigma_{position\ error}) \sim condition\ ordinal + (1|participant)$
forward difference coding; $AIC = 228.68, BIC = 252.1$

	est.	s.e.	t	d.f.	p
intercept	1.00	0.08	12.90	24.05	<0.001
1 object	-0.67	0.14	-4.74	120.23	<0.001
2 objects	-0.38	0.14	-2.67	120.23	0.01
3 objects	-0.08	0.14	-0.57	120.23	0.57
10 objects	0.23	0.14	1.66	120.23	0.1
99 objects	0.67	0.14	4.75	534.05	<0.001

2.3 Separating effects of low and high clutter

We set out to investigate the effects of environmental complexity on navigation performance. In our experiments, complexity was manipulated by changing the number of objects present in the virtual environment. Based on the patterns of performance described above, we will group our experimental conditions into low clutter effects for trials with zero to three objects (green bar in Fig. 4 and 5, and green distributions in Fig. 6), and high clutter effects for trials with three to 99 objects (purple bar in Fig. 4 and 5, and purple distributions in Fig. 6) respectively. We motivate this separation by means of two approaches below.

Fig 5. Quadratic fits on the log-transformed position error (top) or log-transformed position error standard deviation (SD) (bottom) with ordinal condition show a local minimum (red vertical line) close to the three object condition. Fits include data from N=23 participants with n=24 repetitions per condition. Ordinal conditions from zero to five correspond to conditions with zero, one, two, three, ten, and 99 objects respectively. Quadratic regressions fit the data significantly better than a linear regression and the local minimum at three objects is validated in a linear regression model with forward difference coding (see Tab. 1). Individual participants are plotted in the same colour in all figures. Thereby colour is varied from dark red to light yellow from corresponding low to high median position error of that participant in the 99 object condition (see Fig. 4 top right). Green and purple horizontal bars at the bottom indicate low and high clutter effects (see section "Separating effects of low and high clutter" for details)

Fig 6. Performance improvement in accuracy and precision of position error from zero to three objects and performance impairment from three to 99 objects. Green and purple curves indicate kernel-density estimations fitted to the difference in median position error (left) or position error standard deviation (SD) (right) between zero and three objects (green), or three and 99 objects (purple), respectively. For a statistical comparison of skewness, kurtosis, and difference of distributions see section "Quantifying effects of low and high clutter". Boxplots summarise adjacent data points, that indicate differences for single participants. Boxplot notches indicate 95% confidence intervals around the median. All but one participant performed better with three than with zero objects. Only some participants performed worse with 99 objects, while others remained equally good. Participants are plotted in the same colour as in all figures. Thereby colour is varied from dark red to light yellow from corresponding low to high median position error of that participant in the 99 object condition (see Fig. 4 top right).

1) Linear vs quadratic fit As described above, we observe a significant linear and a significant quadratic component in our linear mixed effect regression models for both accuracy and precision (see section "Disentangling the effect of different degrees of clutter"). A simple linear fit does not well reflect the visually observed pattern of position error decrease from zero to three objects and increase from three to 99 objects (see Fig. 4 and 5). The significant quadratic component matches the observation better, which is confirmed by a comparison of the linear and the quadratic component of our linear mixed models above. It shows the quadratic function yielding a better fit to our data (for accuracy: linear: AIC≈293, BIC≈301.8; quadratic: AIC≈255.4, BIC≈267.1; likelihood ratio test: χ^2 ≈39.63, logLik1=-143.5, logLik2=123.7, p<0.0001; and for precision AIC≈297.7, BIC≈306.5; quadratic: AIC≈241.5, BIC≈253.2; likelihood ratio test: χ^2 ≈58.23, logLik1=-145.8, logLik2=-116.7, p<0.0001). An evaluation of the quadratic fits reveals theoretical minima for position error (3.19) and its standard deviation (2.66) close to the three object condition, identifying this condition as the one with the smallest homing errors (see Fig. 5). This speaks in favour of two separate effects with an error decrease for low clutter conditions (zero to three objects), and an error increase for high clutter conditions (three to 99 objects).

2) Forward difference contrast coding Alternatively to analysing the linear and quadratic components of our linear mixed model, we can re-state the model using forward difference contrast coding. In forward difference contrast coding, "the mean of the dependent variable for one level of the categorical variable is compared to the mean of the dependent variable for the next level" [38]. The underlying coding matrix allows

for comparing the central tendency of the dependent variable of each class with the subsequent class on the axis of the independent variable. In our case, this means that the zero object condition is compared to the one object condition, which in turn is compared to the two object condition, and so on. For our data, this approach results in significant negative slopes for the conditions with one, two and three objects, and in significant positive slopes for ten and 99 objects, for both accuracy (Tab. 1 model 3) and precision (Tab. 1 model 6). Note that for precision the estimate from the two to three object condition shows only a trend, indicating no significant improvement in precision. Again, this analysis speaks in favour of two separate effects for low clutter and high clutter conditions.

2.4 Quantifying effects of low and high clutter

With the three object condition separating low and high clutter effects, we ask which of the factors (number of objects and individual differences) can better explain the changes in homing performance for more and less cluttered environments.

To quantify these effects, we analyse the median performance change in accuracy and precision from the zero object to the three object condition and compare it to the change from the three object to the 99 object condition for each individual participant. Note that this analysis was conducted without log-transforming position error because it is not subject to assumptions of linear regression models, and offers easier interpretability of the results.

When comparing the two distributions of performance changes in the low clutter and high clutter environments (see Fig. 6) we see that, as expected, the position error and its standard deviation decrease as objects are added in the "low clutter" (zero to three objects) environments. Conversely, in "high clutter" (three to 99 objects) environments, we find in the kernel density distribution for accuracy one local maximum at $0\,vm$ and a second lower one around $15\,vm$ (Fig. 6 left), and for precision at $2\,vm$ and around $10\,vm$, respectively (Fig. 6 right). The first peak indicates no considerable change in position error for most participants, while the second one depicts participants, for whom 99 objects lead to an impairment of homing performance. While the median difference in position error between zero and three objects does not significantly differ from a

normal distribution (Shapiro-Wilk test: $W \approx 0.974$, $p \approx 0.793$), the distribution of differences between three and 99 objects does differ significantly from normality (Shapiro-Wilk test: $W \approx 0.735$, $p < 0.0001$).

This difference of distributions is also reflected in a significant Wilcoxon signed-rank test result as well as in differences in skew and kurtosis (zero vs three objects: $K \approx -0.67$, $S \approx 0.04$; three vs 99 objects: $K \approx 1.49$, $S \approx 1.62$; Wilcoxon signed-rank test: $Z = 0$, $p < 0.0001$, $d \approx 2.16$). For the difference in standard deviation of the position error between zero and three objects the distribution does differ significantly from a normal distribution (Shapiro-Wilk test: $W \approx 0.902$, $p \approx 0.027$), while it does not differ significantly from normal for the difference between three and 99 objects (Shapiro-Wilk test: $W \approx 0.949$, $p \approx 0.277$). The difference of distributions is likewise reflected in a significant Wilcoxon signed-rank test as well as in differences in skew and kurtosis (zero vs three objects: $K \approx 1.72$, $S \approx -1.27$; three vs 99 objects: $K \approx -0.12$, $S \approx 0.49$; Wilcoxon signed-rank test: $Z = 1.0$, $p < 0.0001$, $d \approx 2.33$).

We therefore conclude that the number of objects is the main driver of participant homing performance in low clutter conditions (zero to three objects), while for the high clutter conditions (three to 99 objects) the individual behaviour of the participants plays a larger role in determining their success.

2.5 Relation between accuracy and precision

We have analysed median homing errors (accuracy) and the standard deviation of errors (precision) and found matching patterns for environments with different degrees of clutter for both measures. This is supplemented by direct correlations of the two measures in all conditions (see Fig. 7), which shows that higher accuracy is usually associated with higher precision, at least on the population level. However, looking at the performance of single participants reveals inter-individual differences. Participants exhibit different degrees of accuracy and precision, with some being accurate but imprecise (Fig. 7a and e), and others being inaccurate but precise (Fig. 7b and c). These participants weaken the precision-accuracy correlation at the population level. The strongest correlation was found in the zero object condition, where 44% of variance in accuracy is explained by precision.

Fig 7. Position error precision and accuracy are significantly correlated in all conditions. There are large inter-individual differences with exemplary participants (a-f) performing (a and e) accurately and imprecisely, (b and c) inaccurately and precisely, (d) inaccurately and imprecise, or (f) accurately and precisely. Data points indicate median position error and its standard deviation (SD) for each of N=23 participants with n=24 repetitions per condition. Individual participants are plotted in the same colour in all figures. Thereby colour is varied from dark red to light yellow from corresponding low to high median position error of that participant in the 99 object condition (see Fig. 4 top right). Grey areas indicate 95% confidence intervals.

2.6 Modelling cue interactions using MLE

Maximum likelihood estimation (MLE) is an important statistical tool for the analysis of navigation strategies [6]. MLE assumes that navigators combine information from different sources in a way that maximises the likelihood of the observed behaviour. To employ MLE as a tool, one must first formulate a mathematical model which describes the way the information sources in question could be used. MLE then seeks to find the set of parameters that best explain the observed data given the predefined model. Then, in a second step, one can compare the overall likelihood of different model variants, which are formal ways of describing qualitatively different uses of spatial information. This procedure can provide quantitative evidence for the qualitative question of how cues are used by an observed navigator.

For our study, we want to model the use of PI and landmark cues, to gain insights into how the two sources of spatial information interact in the different environmental conditions. We compose our model of a **distance** and a **direction** estimate that can be thought of as the norm and angle of a vector, which represents a complete **position** estimate. For landmark information, inspired by [9], we model distance information as Gaussian probability density functions fitted on the distance estimates made by participants relative to the landmark cues (see section "Maximum Likelihood Estimation modelling" for more details). For PI-based distance information, we extend the proposed model and fit a Gaussian on distance estimates relative to the start of the return trip (see Fig. 8A). To derive a full position estimate from the PI system, we add a directional component to the model. This directional component takes the form of a circular normal, or von Mises, probability density function, fitted to directional estimates relative to the start of the return trip (Fig. 8B). This approach to encoding directional estimates has been introduced by [39] and employed successfully to model

human navigation behaviour by [7]. By either setting or fitting different values to the parameters of these three model components (landmark-based distance estimate, PI-based distance and direction estimate), we can now quantitatively express different hypotheses about how the landmarks and PI are used by the participants.

We summarise the model components in Tab. 2 and visualise the spatial combination of functions in Fig. 8, which shows the result of performing MLE along both the distance and the angular domains. It can be clearly seen, how the PI-driven angular estimate provides the spatial constraint needed to limit position estimates from the full ring given by the landmark-driven distance function alone to the area actually targeted by participants.

In the zero object condition, there is only one source of distance and direction information, namely the PI system (see Fig. 8 top row). However, each added object provides another source of distance information relative to itself, which we can incorporate into our model. We therefore create models using the distance and direction information provided by the PI system and combine them with distance-based models for each object. Specifically, each of the three objects present in the one to three object conditions had a goal distance of $10\,vm$. These objects surround the goal location and allow for very precise homing in all conditions in which they are present, if identified correctly among the other objects, as we describe above (see section "Introduction").

We create distance models for these objects and combine them with the distance and direction models from the PI condition and investigate their goodness of fit, expressed as the log-likelihood of the respective model. These model comparisons show that model variants including the information from PI always outperform those disregarding the PI information when trying to predict participant behaviour (best models for each condition shown in Fig. 8 middle and bottom row, full comparison of model fits in appendix S3 Table). This clearly indicates that participants combine information provided by their PI system with information provided by the objects to improve their homing performance. It is especially intriguing, that the prediction areas produced by these models very closely match the observed participant behaviour without needing to fit the parameters of the landmark models (see Fig. 8, panels E-G), which are instead based on the function mean set at the geometrically correct distance of $10\,vm$. This kind of match indicates that 1) the assumptions made by the models represent the

actual navigation strategies employed by the participants to a reasonable degree and 2) that in the one to three landmark conditions the participants' goal estimate is, on average, identical to the geometrically correct goal location.

For the high-clutter conditions (ten, 99 objects, Fig. 8, panels H-I), a full integration approach, as described above, is not sensible, neither from a conceptual nor from a mathematical perspective. From a conceptual standpoint, it is unlikely that participants explicitly memorise distances to such large numbers of objects and combine the information across the whole environment. Mathematically, the integration of such a large number of cues, which involves the multiplication of probabilities, results in likelihoods approaching negative infinity, precluding any meaningful comparison operations among variants of models within this class.

An alternative way to encode the information provided by the objects is an "alternation" approach, in which a navigator switches between the individual objects used for homing from trial to trial. Mathematically, this can be represented by an addition of the individual object's predictions (as opposed to a multiplication for the integration model). This model class is conceptually more appropriate for the ten and 99 object conditions, since we observe a broad spread of responses across the environment in this condition. Such a pattern matches the prediction of cue alternation models, which lack the variance reduction over single cue conditions predicted by integration models [6]. For our analysis, we evaluate integration and alternation models for each condition and compare the goodness of fit of variants using different model parameters. Specifically, we compare models fitted on the results of the zero object condition, which represents homing performance driven purely by PI, with those fitted on the data of each of the other conditions (see Tab. 2 for an overview and appendix S3 Table for detailed comparisons). If the use of PI was unchanged across conditions, these model variants should not differ in goodness of fit. However, we find that the zero object models are always outperformed by the models fitted on the data for the given condition when evaluated on the data of the given condition. This indicates that the presence of objects does not only add further information in the form of a second type of cue but also improves the quality of the information provided by the PI system. This assessment is further supported by the fact that when considering the PI components of our models, parameter estimates qualitatively differ between the landmark and

no-landmark conditions. For the average PI-based distance estimate, we observe a degree of *overshooting* on the population level for the zero object condition, while a degree of *undershooting* is observed for the conditions in which landmarks were present (Fig 9 top left panel). Intriguingly, the overshoot in distance estimates returns for the 99 object condition, matching the results of the analysis of position errors presented above, which indicate that several participants behave very similarly in the zero object and 99 object conditions. We find similar results for the PI-based direction estimates, however, here two objects seem to be required to overcome the directional bias observed in the no-landmark condition (Fig 9 bottom left panel). When investigating the spread of estimates in both distance and direction domains (Fig 9 right panels), we find that results closely match those observed in the analysis of position errors, with precision increasing (spread in estimates decreasing) as up to three landmarks are added and then decreasing again as further objects are added.

Fig 8. Overview of MLE modelling results. Panels show participant goal estimates (black dots) for different experimental conditions, overlayed on top of maximum likelihood model predictions for the respective condition. Heatmap colour signifies the predicted likelihood of targeting a given area, with darker colours indicating greater likelihood. Panel A introduces the ring-shaped distance estimate model, based on fitting a normal distribution to the distance component of the goal estimates. Heatmap shows the overall best model fit (maximum likelihood estimate) for this model type for the pooled data of all N=23 participants for the zero object condition. Panel B introduces the fan-shaped direction estimate model, based on fitting a von Mises (circular normal) distribution to the directional component of the goal estimates. Heatmap shows the overall best model fit (maximum likelihood estimate) for this model type for the pooled data of all participants for the zero object condition. Panel C shows the prediction resulting from an integration of both model components ($ring \times fan$) for the same data. Panels D-F show the resulting prediction from an integration model combining a full PI model with an additional ring model for each object fitted at $10\,vm$ distance ($ring \times fan \times \prod ring$). Panels G-H show the resulting prediction from a model combining a full PI model with an alternation of ring models for each object fitted at $10\,vm$ distance ($ring \times fan \times \sum ring$).

Fig 9. Maximum likelihood parameter estimates for the path integration model. Top row shows maximum likelihood estimates with bootstrapped 95% confidence intervals for the parameters of the distance models fitted on the pooled data of all N=23 participants (μ gives the average expected distance, σ the standard deviation of expected distances), bottom row shows parameters for the direction models (μ gives average expected direction, κ is the concentration paramater of the underlying von Mises distribution and has been transformed to circular standard deviation for easier readability). See Eq. 3 and Eq. 4 for the full description of the model functions. Fitted parameters are shown for each experimental condition separately, with overlapping 95% confidence intervals indicating that parameter estimates are similar between conditions. Dashed red lines indicate geometric ground truth where applicable.

Table 2. Overview of maximum likelihood estimation model components. "g" denotes distance models , "f" denotes direction models. "LM" and "PI" denote whether models were based on the landmark information or PI estimate respectively. "$*$" denotes that models parameters were maximum likelihood estimates, either for the current condition or for the 0-object-condition ($0\,objs$) if specified. "$true-dist$" denotes that landmark models were fitted using the true distance to the goal for the 3 central objects surrounding the goal, rather than the MLE for the given data. "$\prod$" denotes integration of individual landmark models, while "$\sum$" denotes alternation. Top subdivision lists basic model components, middle subdivision gives the form of the full models used, bottom subdivision gives the best performing model variant for the different conditions.

model	**description**
g_{PI}	PI-based distance model component
f_{PI}	PI-based direction model component
g_{LM}	LM-based distance model component
$g_{PI} \times f_{PI} \times \prod g_{LM}$	integration model example
$g_{PI} \times f_{PI} \times \sum g_{LM}$	alternation model example
$g^*_{PI} \times f^*_{PI}$	best fit 0 objects
$g^*_{PI} \times f^*_{PI} \times \prod g^{true-dist}_{LM}$	best fit 1-3 objects
$g^*_{PI} \times f^*_{PI} \times \sum g^{true-dist}_{LM}$	best fit 10-99 objects

3 Discussion

In our VR experiment, participants show clear changes in homing performance across environments with varying visual complexity, based on different numbers of objects available. We analysed the average magnitude of homing errors (accuracy), as well as the spread in errors (precision) for each environment. When no objects are present and participants can thus only rely on path integration (PI) alone, both accuracy and precision are poor, and individual performance differs considerably. Both measures, accuracy and precision, greatly improve above the level observed for the pure PI condition when one to three objects are added around the goal (Fig. 4). To our knowledge, this study is the first to investigate the effect of environments with more

than three objects systematically. For those high clutter environments (namely ten or 99 in our experiment), we observe less uniform responses with one share of participants for which both accuracy and precision stay mostly unchanged and another share for which either one or both of these error measures increase considerably, indicating large difficulties to successfully navigate in such cluttered environments (Fig. 4 and Fig. 6).

An analysis using linear mixed effect models, as well as maximum likelihood estimation (MLE) leads us to conclude that homing performance is driven by the number of objects in low clutter environments (zero to three objects), and mainly determined by the individual preferences in cue use of the participants in high clutter environments (three to 99 objects). These individual differences in the influence of environmental complexity on accuracy and precision (Fig. 4 and Fig. 6) are also indicated by the patterns observed in correlations between accuracy and precision, where higher accuracy is usually associated with higher precision but with variance across participants (Fig. 7).

3.1 Low clutter: PI and geometric reasoning dominate homing performance

All participants improved in homing performance with up to three objects added around the goal. This can be explained on the basis of object constellation geometry. The first three objects were placed in a triangular arrangement around the goal location. Thus, each additional object further geometrically constrained the region in which the goal could lie. This is because, with only a single object present, the only information available from the object itself is its distance to the goal, which can be memorised by the participant. As predicted by a simple landmark-based homing model by [9], this leads to a ring-shaped area in which the target could be located based on the object-derived information. We observe goal estimates to be located in this area. As further objects are added, the area becomes constrained to the intersection of the ring-shaped area, leading to a fully constrained goal as soon as three objects are present. In other words, to triangulate a specific point in a three-dimensional space, three distinct references are necessary. Based on this geometric argument, we hypothesised homing performance to increase with each object up to the third one (see introduction

section "How does homing performance of human navigators depend on the visual complexity of the environment?"). This hypothesis is validated by our findings.

However, in our version of the task, participants walked around freely in the virtual environment and were not displaced or relocated. Therefore, in addition to the object cues, participants could make use of their PI, charged on the outbound path of each trial. Furthermore, the walked triangle itself provides a geometric constraint. Due to these factors, the response distributions observed are directionally limited based on the approach direction.

Overall, for all the low clutter conditions (zero to three objects), the fact that responses were limited to a smaller area than the one observed for the zero object condition shows that participants integrated the different sources of information in the Bayesian sense, instead of alternating between them, which would have resulted in a larger spread of responses, rather than a reduced one [18, 40] (see Fig. 8C-F).

To quantitatively express this combination of spatial cues we combine a model meant to predict distance estimates based on landmarks [9] with a model meant to predict directional estimates [7, 39]. Together, these two models allow us to create predictions from both the PI system, as well as any landmark present. We could show that these predictions match the observed spatial pattern of responses, and thus expand on the findings by [9] for one, two and three landmark objects (see also Fig. 8 middle row). In the condition with one object, this is illustrated by goal estimates being located in only a part of the predicted ring-shaped distribution. In the two object condition, rather than a bimodal distribution of responses, we see a more uniform spread between the predicted two locations, with a clear bias towards the "true" goal location, rather than its mirror (see Fig. 3 and Fig. 8 middle row). Our MLE models indicate, that this improvement over the "pure" landmark situation described by [9] can be attributed to the presence of distance information from PI. Finally, with three objects, the goal location is fully spatially constrained and within a Bayesian framework of optimal cue integration, we expect any remaining errors to be attributed to the limitations in cue quality and imprecisions in the encoding of spatial information or execution of movements. Accordingly, we expect a tight distribution of goal estimates around the true goal. For this condition, we actually observe that some of the provided goal estimates are slightly worse than expected for a fully triangulated goal (see Figs. 3

and appendix "S1 Full trajectory figures"). Here again, our MLE models provide additional insight, as we could show that these goal estimates lie within the prediction area of the directional component of our PI model (see Fig. 8). The observed patterns thus indicate that participants integrated object and PI information even when the PI estimate was biased towards underestimating the distance to be travelled.

3.2 High clutter: individual preferences in cue use dominate homing performance

For our high clutter conditions (ten and 99 objects) it is hard to disentangle the use of PI and object-based guidance. We cannot simply use the geometric considerations above as large inter-individual differences indicate that it is not clear how individuals use the available objects for navigation (Fig. 6).

As outlined by [41] individual navigation abilities might be determined by a variety of factors ranging from how environmental and self-motion cues are perceived and combined, to what cue qualities are personally favoured. Potentially relevant factors may include sex, age, and differences in motion, hippocampal processing, and executive functions [41]. [42] also find evidence that individuals store environmental information in spatial tasks with different degrees of detail and argue for individually different complexities in spatial memory and navigation ability. These individual navigation preferences are supported by the findings of [34] who distinguish two groups of navigators, one group that weights available cues of visual and auditory modality similarly (and combines them optimally in the Bayesian sense), and another group that weights visual cues more strongly. They consider their observation of a preference for visual landmarks in one group to indicate a reset of the directional component of the PI, if landmarks are judged to be reliable by the navigator. Depending on which objects in a condition are considered reliable, different objects might dominate the visual perception of the surroundings for different participants, and if accompanied by high cue weighting, reset PI. Considering that certain trees in our VR are close to the outbound and homing path, possibly a tree's distance to the avatar and the magnitude of optic flow generated by it, might decide about its dominance in cue integration.

From their experimental results in a cue conflict task between visual PI and unique

landmarks, [7] suggest that participants use a hybrid model of cue integration. This assessment was later supported by [11] in a cue conflict task with proprioceptive and visual sensory inputs. In such a hybrid model, body-based and visual cues were combined only when the estimates of PI and landmark guidance were similar. If cues are dissimilar, they will compete for dominance. The switch between combination or competition happens according to the subjective probability that both sensory inputs are correct and refer to the same environmental feature. In our experiment, this is related to the position of goal estimates derived from different cues and the magnitude of difference between them. Previously, [7] have concluded that landmarks can reset the orientation of PI but cannot influence its response variance.

This does not match the results obtained from our MLE modelling, where the standard deviation of both distance and direction estimates are reduced when objects are present (Fig. 9). Thus in our experiment, when PI and object cues are perceived as conflicting, this does not seem to cause a complete reset of either one of them, but rather a different form of combination, which yields at least some improvement of precision to the location estimates made by participants. One possible explanation is that the inclusion of prior information provided by individual causal inference is added in the cue integration process, which depicts how likely sensory inputs originate from the same source depending on their perceived similarity [43, 44].

If the cues (PI and objects) are combined optimally in a Bayesian sense, we would not expect the variance to increase in high clutter because more available cues sharpen the distribution of goal estimation, unless reliability of the cues is reduced. We cannot clearly determine if the quality of PI or landmark guidance is reduced in clutter, but we argue that with more objects in the scene, the optic flow and therefore the information for PI increases with no identifiable downside. This suggests that, at least for some participants, the reliability of objects in high clutter is reduced. Whether objects in our different conditions are considered as reliable landmarks can depend on a variety of factors, which we summarise below.

3.3 The forest or the trees?

In a visually complex, cluttered environment, people might memorise single objects, constellations of different numbers of objects, clearings, or even the edge of the clutter. Because we do not know what features of the visual surroundings are remembered by participants and also cannot differentiate between different variants of cue use in our experiment, we consider the trees in our experimental design as only "objects" and not "landmarks", taking into account [45] who highlights that the term "landmark" is often used in an ambiguous manner. In order to avoid an ambiguous definition of a landmark, recent research established objective identification of salient buildings in a VR city, using a graph-theoretical model approach based on a participant's view fixations and saccades in corresponding eye tracking data [14]. However, how exactly a certain navigator uses such categorised landmarks in any specific environment still remains unclear.

One possible variant of cue use would be to consider the entirety of available objects as one landmark cue without precisely determining its sensory processing. The quality and reliability of a cue influence its importance and weighting in the process of cue integration. Hence, the more useful the objects are as a source of spatial information, the higher people should weigh object guidance and the better they should perform. Thus, performance across conditions in our experiment reveals that object guidance is less useful, when no or too many ambiguous objects are available, with one, two, and ten object conditions representing in-between cases. Cue quality of the objects is likely caused by processing effort to disambiguate single trees, calculate geometric relations, and encode the goal location from it. The 99 object condition represents a special condition, as participants separate into those who keep good performance and those who get worse, likely based on whether participants are able to disambiguate object cues. This hypothesis is supported by the observation, that a regression model including all conditions shows significant heteroscedasticity, that is not present when the 99 object condition is excluded. When comparing the median position error in the three object condition and the 99 object condition on the individual level, we observe deviation from a normal distribution mainly driven by positive skew due to a considerable amount of participants in our sample that show distinctly higher median position errors than others (Fig. 6).

We have been able to partially capture the effects of differing abilities of participants to employ objects for guidance by using a cue alternation instead of an integration approach. A model treating every object like one of the three objects immediately surrounding the goal and randomly picking one of them every trial yields at least a rough estimate of population behaviour. Such a model cannot capture when and how individual objects are used as cues, but it can be a good first approximation, which further underlines that different participants perceive the objects differently, possibly as individual trees, or as the entirety of a forest.

3.4 PI and object guidance interplay

The observations of [7] and [11], that highly reliable cues can reset less reliable ones, imply for our experiment that when the object cues are judged to be highly reliable, they can reset the orientation of PI, while they cannot do so when they are judged to be unreliable. Thus, continued high performance in highly cluttered environments could be the result of successful attempts by the participants to identify specific constellations of objects and thus in our experiment to arrive at the correct forest clearing. When specific constellations of trees are identified, they result in a reliable object cue that can reset inaccurate PI. Good PI, however, helps to find the correct direction and distance and makes it easier to look for the right object constellations. In contrast, participants who cannot disambiguate between the different object constellations might steer towards another, incorrect clearing and goal location, or even disregard all objects and fall back on PI alone. This last strategy should result in a large spread in goal estimates, as can be seen from the zero object condition (Fig. 2 and Fig. 1). Interestingly, from the perspective of cue integration, the interplay of guidance mechanisms (landmarks and PI) might actually lead to worse performance for some participants. If they walk in a wrong direction due to imprecise PI, they might mistake incorrect, but spatially similar, object constellations for their goal location. As apparent from the comparison between performance accuracy and precision between the zero and 99 object condition for single participants (different coloured dots in Fig. 4), we cannot disentangle if it is PI or landmark guidance that makes participants perform well or badly in high clutter. In other words, people who show good PI in the zero object condition are not

necessarily performing well in the 99 object condition, or vice versa. We argue that PI performance in the zero object condition as it stands is not a good predictor for performance in conditions with objects. The MLE models show that PI estimates are updated as objects are added to the environment, as indicated by the change from overshoot to undershoot in the population distance estimates. This kind of qualitative change in the output of the PI system might indicate that the additional spatial information provided by the objects is not only used to gradually improve the estimate provided by the PI system but to actively reset it, as was also proposed by [7].

We can also draw conclusions about PI and object guidance and how they might influence each other from the comparison of accuracy and precision in the different conditions. We consider participants who combine high accuracy in homing with low precision (bottom right quadrants and subfigures a and e in Fig. 7) who end up roughly in the correct location, and participants who show low accuracy with high precision (upper left quadrants and subfigures b and c in Fig. 7) who target an incorrect location consistently. Interestingly, low accuracy combined with high precision occurs mainly in conditions with zero, one, and two objects. In the zero object condition, this would imply that the participant's PI consistently led them to a wrong place. [46] demonstrated that PI performance is partly based on a correlation between the ability to track rotational self-motion and grey matter volume in the cerebellum. Later, [47] were also able to show significant correlations in the hippocampus, entorhinal cortex, and thalamus between structural volume and individual first-person navigational accuracy, which might underlie some of the differences we observe.

When a participant's PI is repeatedly guiding them into a direction that is not the goal direction, we consider their PI biased (see also population trends in MLE models, Fig. 9). Such a biased PI can also be seen to influence a participant's goal estimation in other conditions in which objects are present. In the one object condition, a biased PI can guide participants to the wrong part of the ring-shaped target area around the only present object (subfigure b in Fig. 7). Thus, combining PI information with that derived from the object can cause larger position errors than PI alone, depending on the directional bias of the PI. In the two object condition, the intersection of "rings" around the two objects indicating the correct goal distance leads to two possible goal locations. Here, cue integration with a sufficiently biased PI can lead participants to choose the

incorrect one repeatedly (subfigure c in Fig. 7). One might also expect inaccurate, but precise participants in the high clutter conditions, as we have just described for the other conditions. However, we do not observe such a pattern for those conditions. This implies that participants who struggle in the high clutter conditions, do not confuse single objects, object constellations, or clearings, but rather cannot successfully integrate the information provided by the objects with their PI system and form a consistent image of the forest. Eventually, they end up in a number of wrong places. High accuracy along with low precision can mostly be found in the 99 object condition (bottom right quadrant in Fig. 7). One might expect that participants who show this pattern of performance have problems using the information provided by the objects and fall back on their PI system to solve the task, resulting in error patterns similar to those observed for the zero object condition. However, on closer inspection the participants in question seem to be able to home towards the correct goal with high precision in most trials but target incorrect locations in individual trials, leading to the overall pattern of low precision when analysing the totality of all trials. We see that PI, which is the only cue present in the zero object condition, can also influence target estimation in the other conditions, both positively and negatively. On the one hand, it can help participants select the correct of multiple possible goal locations indicated by objects. On the other hand, a biased PI can also have an opposing effect. The individual weighting of PI and object guidance does not allow us to make direct predictions in our experiment from the pure PI condition with zero objects to high clutter environments with ten or 99 objects.

3.5 Conclusion and outlook

In this study, we demonstrate that navigation performance varies in low and high clutter: In low clutter environments, the number of objects and geometric reasoning drive performance, while in high clutter environments, individual differences in cue use are decisive. An MLE model incorporating spatial information from PI and objects reveals that behaviour in low clutter is well described by cue integration. In high clutter, our data reveals that individual participants find different solutions to the task. For some, adding objects does not impair their performance, for others this seems to impair their use of object information. While we are able to outline a variety of

reciprocal influences between spatial information from PI and objects, we cannot ultimately elucidate what exact features of objects in clutter of different degrees are used by individual participants, or what exactly determines cue weighting when information is perceived as conflicting.

To further investigate cue weighting of PI guidance and object guidance, our future study designs will incorporate targeted cue conflict trials. To better understand visual processing and the use of object features, we will analyse individual-level eye tracking data. These experimental approaches may be well-suited to reveal why some participants do not see the forest for the trees.

4 Methods

4.1 Task description

This study's paradigm is a triangle completion task, in which the participant is guided away from a goal location along two legs of a triangle and then asked to return to their goal location, thus trying to complete the triangle as accurately as possible. By quantifying the errors made by participants, this task assesses the participant's homing performance. We tested participants in triangle completion tasks in different conditions that differed in the number of available tree objects in the surroundings. The number was systematically varied between zero and 99 (zero, one, two, three, ten, 99) identical, rotationally symmetric trees to present different degrees of visual clutter and thus different degrees of visual ambiguity of the possible object cues. Conditions with higher object counts always contained the objects of conditions with lower object counts as well (Fig. 1). The condition with zero objects was included to provide baseline performance for a situation in which only PI information was available for navigation, while the one to three object conditions allowed us to compare our results to those of [9], who used similar spatial configurations. Finally, the ten and 99 object conditions provided us with environments of intermediate (ten objects) and high (99 objects) degrees of visual clutter.

4.2 Experimental procedure

25 healthy participants (aged 19-30 years, 9 self-identified as female, 16 self-identified as male) took part in the experiment. Participants were informed about the general procedure of the experiment but not about its specific purpose or the nature of the different conditions. All participants gave their written informed consent before participating and monetary compensation (7 Euros per hour). The experiments were approved by the Bielefeld University Ethics Committee and conducted in accordance with the guidelines of the Deutsche Gesellschaft für Psychologie e.V. (DGPs), which correspond to the guidelines of the American Psychological Association (APA). The experiment was carried out over six sessions (one training + five main sessions), split evenly over two experimental days, which were not more than 1 calendar day apart. The experiment was provided in the form of a stand-alone Windows application, created using the virtual navigation toolbox for the Unity3D game engine and was executed remotely on participants' desktop home computers [48]. A supervisor was present to instruct and assist participants according to a standardised protocol during the initial training session via video call. After successful training, participants were able to complete the remaining experimental sessions without further supervision. However, participants were instructed to take regular breaks in between experimental sessions of at least 5-10 minutes. Each main session consisted of six repetitions of each of the six experimental conditions (number of objects) in a randomised order, leading to a total of 36 trials per session and adding up to n=24 repetitions per condition overall for each participant over the whole experiment. Each session took roughly 30-45 minutes to complete, leading to a total experiment duration of 3-4.5 hours per participant.

4.3 Trial structure

During a single trial of the experiment, the participant was first guided to the goal location, marked by a stylised campfire in the virtual environment. The avatar was controlled using the up/down and the left/right arrow keys, respectively, for translation and rotation. After reaching the campfire goal location, they were tasked with collecting pieces of firewood, with the wood appearing in sequence at the two corners of the first two legs of the triangle path, being 35.4 and 25 vm long (see Fig. 1). This was termed

the outbound phase of the trial. During this initial phase, a floating arrow in front of participants indicated the correct direction towards the next pile of wood, but the participants were otherwise free to move as they liked. It was ensured that no trees were placed on the outbound, or direct return path, so that participants did not have to maneuver around them. To ensure an overall straight path along the legs of the outbound path, the participant was confronted with a time limit (45 seconds) to collect all the firewood. The remaining time was always visible in the form of a countdown bar at the top of the screen. Once the participant had completed the outbound phase, they were asked to return to the campfire (i.e. the goal location), once again while being presented with a countdown timer (45 seconds). Neither the piles of wood nor the campfire were visible during the return, to prevent their use as spatial cues. Identical grass tufts on the ground were randomly arranged for each trial to enable visual estimation of self-motion by optic flow. A sun-like light source provided global, uni-directional illumination from directly above so that no directional information could be derived from shadows in the environment. The avatar moved at 2.5 virtual meters per second. During each trial, the position and rotation of the player's avatar in the virtual environment were recorded for every rendered frame. All data collected from the participants and during the experiment was handled according to the General Data Protection Regulation (GDPR) of the European Union.

4.4 Analysis

All analyses, except for the linear regression analysis which was done in custom R [49] scripts in version 4.1.3, were conducted in custom Python [50] scripts in version 3.9.12.

4.4.1 Error measure

The primary error measure for our analysis is the position error, which is the Euclidean distance between the intended goal location and the endpoint of the actually walked trajectory. We calculated median position errors for each participant and condition as our accuracy measure, and standard deviation of position error as our precision measure (the lower the standard deviation the higher the precision).

4.4.2 Power analysis for sample size estimation

For the design of this study, we analysed data from a previous pilot study with a virtual triangle completion task and different numbers of trees [48]. and did power analyses for repetitions per condition and for number of participants using the one-sample t-test function in G*Power 3.1 [51].

We planned to collect data from N=20 participants with n=24 repetitions per condition each. Given that planned sample size, power $\beta = 95\%$, and significance level $\alpha = 5\%$, the observed $2.78\,vm$ median standard deviation of position errors within participants and within conditions of the pilot study allows to detect differences in position of $\geq 2.15\,vm$ in a two-sided one-sample t-test. Likewise, for the observed $4.02\,vm$ standard deviation of participants' median position error within conditions, we can identify differences in position of $\geq 3.5\,vm$

In this study's data (N=23, n=24), following the same calculations, we can resolve with the same power, significance level, and t-test, differences of $\geq 3.2\,vm$ within participants and within conditions, and $\geq 2.25\,vm$ between participants and within conditions, both of which are sufficient to support the reported conclusions.

4.4.3 Exclusion criteria

The experiment consisted of training and five main test sessions with initially 25 participants. Six of the participants did not complete the last main session, and thus we decided to include only four main sessions to acquire a balanced dataset and include those six participants that otherwise would have been rejected. Two additional participants were excluded from all analyses, one of which did not finish the experiment in the given time, and one who indicated in the follow-up questionnaire that they were using the computer display's rim as an external cue to better estimate the goal location. We remained with 23 participants completing 24 trials per condition (23.7±0.71 after NaN removal) within four main sessions. From the raw data, we excluded trials for which data recording yielded NaN values, which was the case in less than seven trials for any participant. Also, we detected trial incidences where participants rotated counter-clockwise at the homing start, a behaviour that likely changes the estimation of the rotational aspect in their PI. However, we did not consider those instances as

incorrect task execution and maintained them. For validation purposes, we also ran our analysis including all available sessions and rejecting data symmetry which yields the same qualitative results with only minor deviations in descriptive statistics. The same holds true when throwing out trials, where participants rotate counter-clockwise at the homing start position. Data cleaning can be retraced in detail with the supplementary material.

4.4.4 Linear mixed effects models

Based on the distribution of position errors across the conditions, we fitted linear mixed effect models to identify the underlying mechanisms that lead to the observed changes in performance. We fitted a model predicting the log-transformed position error (and separately its standard deviation) from the number of objects as a fixed effect and the participant identity as a random effect. We present two approaches, one based on a quadratic regression and one based on forward difference contrast coding, to justify splitting the dataset into two separate effects, an effect of low clutter and an effect of high clutter. We quantify both effects by comparing differences in median position error and in its standard deviation between the respective conditions. Additionally, to understand how precision and accuracy relate to each other, we correlated them separately for each condition, and display results of exemplary participants performing at different degrees of accuracy and precision.

We also provide a secondary analysis using a linear mixed effects model on single-trial data (see section "Data availability" for the supplemental notebooks). The alternative analysis reproduces the qualitative results presented in the main text.

4.4.5 Maximum Likelihood Estimation modelling

We created MLE models to investigate the use of PI and landmark information. Our models are partially based on those proposed by [9], which encode the information provided by landmark cues using a Gaussian probability density function fitted on the distance estimates made by participants relative to the landmark:

$$g(d) = c \times \mathrm{e}^{\frac{-(d-\mu_d)^2}{2\sigma^2}} \tag{3}$$

with d being the metric distance measure with regard to the landmark, σ being the standard deviation of d, μ_d its median and $c = \frac{1}{\sigma\sqrt{2\pi}}$ being the scaling factor.

Since the PI system not only provides distance but also directional information, we expanded the model to include such a directional component in the form of a circular normal, or von Mises, probability density function:

$$f(x) = c \times \mathrm{e}^{\kappa \cos(x-\mu_x)} \tag{4}$$

with x being angular estimates in the interval $[-\pi, \pi]$, μ_x being the median of the angular estimates, $\kappa \approx \frac{1}{std(x)}$ being the concentration parameter and $c = \frac{1}{2\pi I_0(\kappa)}$ being the scaling factor, where $I_0(\kappa)$ is the modified Bessel function of order zero (see also [39] and [7] for an introduction to modelling directional information using circular functions).

We could then derive maximum likelihood estimates for Gaussian (distance) functions centred on every object location, as well as the start of the return trip, which could be combined with a von Mises (direction) function also fitted on the start of the return.

Combinations of distance estimates could be calculated by either summing $(\sum_{i=1}^{n} g_i(d))$, or multiplying $(\prod_{i=1}^{n} g_i(d))$, the probabilities derived by sampling the functions for each of the n landmarks. Here, a summation produces a cue **alternation** model, which assumes individual landmarks are picked at random with selection probabilities based on their reliability [52, 53], while a multiplication produces an **integration** model, which assumes each cue is used in a given trial weighted according to its reliability [6, 7]. For our study, the reliability of all landmarks was assumed to be equal, with the weight derived from the observed reliability in the one object condition [9].

For the full model, these combined distance estimates were then always combined with a single direction estimate ($f(x)$) centred on the start of the return. This combination was always of the integration form ($L_{combined} = f(x) \times g_{combined}(d)$) because a model assuming alternation between using only distance or direction information during a given trial is not plausible given the observed behaviour.

To determine goodness of fit, we tested different model variants against the observed data. Specifically, we always determined the MLE for both distance and direction models for the zero object condition, which gave us a model predicting where

participants should walk if they followed their PI exclusively. This model was then also evaluated on the data of the other conditions and compared with the variant of the same model fitted on the respective condition. These models were then combined with landmark-based distance models as described above. Parameters for the landmark model were always set to reflect the geometric ground truth of the distances of the three central objects around the goal, and the spread of estimates observed for the one object condition specifically ($\mu_d = 10$, $\sigma = \tilde{\sigma}(d_1)$, with d_1 being the distance estimates observed for the first landmark).

Model quality was evaluated by calculating the log-likelihood of each model variant ($\ell = \sum \log(L_{combined})$). Since model variants were always of the same class, i.e. being composed of the same number and type of functions, model log-likelihoods can be compared directly. Furthermore, we used a bootstrapping procedure to derive confidence intervals for all MLEs, which allows for direct visual comparison of parameters across conditions (see Fig. 9).

Finally, we also calculated Akaike and Bayesian information criteria (AIC, BIC) for all models as additional measures for comparison (see appendix "S3 Table" below). Model rankings remained unchanged regardless of the method of comparison.

5 Data availability

The study dataset that supports the findings of this study along with Jupyter Notebooks containing the full analysis pipeline are openly available. Data is available at `https://pub.uni-bielefeld.de/record/2982046`. Analysis code with accompanying instructions for execution is available on GitLab at `https://gitlab.ub.uni-bielefeld.de/jonas.scherer/not_seeing_the_forest_for_the_trees_combination_of_path_integration_and_landmark_cues_in_human_virtual_navigation`. Implementation details of the Unity3D project and an executable Unity project of this study's task are available on GitLab at `https://gitlab.ub.uni-bielefeld.de/virtual_navigation_tools/unity_vnt_showcase_triangle_completion`.

Supporting information

S1 Full trajectory figures. Here we provide figures showing full trajectory data that depict all trials from Figure 3. Most trajectories are straight when no or only few objects are present but for some participants they become more complex and curved with ten and 99 objects.

Fig. S1. Trajectory endpoints of single trials are depicted as dots. Individual N=23 participants are plotted in the same colour in all figures. Thereby colour is varied from dark red to light yellow from corresponding low to high median position error of that participant in the 99 object condition (see Fig. 4 top right). Cutout does not show complete object constellations, please see Fig. 1 for a more zoomed-out view of the environment.

S2 Session order effect. In our analysis, we observe a trend of reduced position error over sessions. This is unsurprising, as placement of objects and shape of the outbound path do not vary across trials and conditions. Solely the number of objects from a given set of objects with fixed positions differs between conditions. However, including session number as an additional fixed effect in a regression model results in only slight improvement of data fit. As the order effect is small and not the primary focus of this study we do not consider it in further analyses.

To deal with possible order effects in participant performance in our linear regression analysis, we average across trials and sessions using median position errors and position error standard deviation (scatter plot in Fig. 4). Please note that we also provide a linear mixed effects model on single-trial data in the supplement.

S3 Table. MLE model summaries.

Fig. S2. Order effect within the different conditions on position error accuracy (top) and precision (bottom). N=23 participants completed four sessions with six repetitions (n=24) in each condition.

Table 3. Summary table of MLE model fits for the zero object condition. For an overview of model components see Table 2. Model fits within each subdivision can be compared directly, since they belong the the same model family.

model ID	log Likelihood	BIC	AIC
g^*_{PI}	-1902.68	3817.96	3809.35
f^{flat}_{PI}	-1003.48	2019.57	2010.96
f^*_{PI}	-304.07	620.75	612.15
$g^*_{PI} \times f^*_{PI}$	-2206.75	4438.71	4421.50
$g^*_{PI} \times f^{flat}_{PI}$	-2906.16	5837.52	5820.31

Table 4. Summary table of MLE model fits for the one object condition. For an overview of model components see Table 2. Model fits within each subdivision can be compared directly, since they belong the the same model family.

model ID	log Likelihood	BIC	AIC
$g^{*\,0objs}_{PI}$	-1767.60	3547.81	3539.20
g^*_{PI}	-1590.35	3193.31	3184.69
f^{flat}_{PI}	-1007.16	2026.93	2018.31
$f^{*\,0objs}_{PI}$	-78.40	169.41	160.80
f^*_{PI}	185.07	-357.53	-366.14
$\prod g^{true-dist}_{LM}$	-1344.11	2700.82	2692.21
$g^*_{PI} \times f^*_{PI} \times \prod g^{true-dist}_{LM}$	-2749.38	5536.60	5510.92
$g^*_{PI} \times f^{*\,0objs}_{PI} \times \prod g^{true-dist}_{LM}$	-3012.85	6063.55	6037.86
$g^*_{PI} \times f^{flat}_{PI} \times \prod g^{true-dist}_{LM}$	-3941.61	7921.06	7895.37
$g^{*\,0objs}_{PI} \times f^*_{PI} \times \prod g^{true-dist}_{LM}$	-2926.63	5891.10	5865.42
$g^{*\,0objs}_{PI} \times f^{*\,0objs}_{PI} \times \prod g^{true-dist}_{LM}$	-3190.11	6418.05	6392.37
$g^{*\,0objs}_{PI} \times f^{flat}_{PI} \times \prod g^{true-dist}_{LM}$	-4118.86	8275.56	8249.88

Table 5. Summary table of MLE model fits for the two object condition. For an overview of model components see Table 2. Model fits within each subdivision can be compared directly, since they belong the the same model family.

model ID	log Likelihood	BIC	AIC
$g^{*\,0objs}_{PI}$	-1731.15	3474.92	3466.31
g^*_{PI}	-1525.18	3062.99	3054.37
f^{flat}_{PI}	-1008.99	2030.61	2021.99
$f^{*\,0objs}_{PI}$	-65.62	143.85	135.23
f^*_{PI}	476.24	-939.85	-948.47
$\prod g^{true-dist}_{LM}$	-2612.62	5250.48	5233.25
$g^*_{PI} \times f^*_{PI} \times \prod g^{true-dist}_{LM}$	-3661.57	7373.61	7339.41
$g^*_{PI} \times f^{*\,0objs}_{PI} \times \prod g^{true-dist}_{LM}$	-4203.42	8457.31	8423.11
$g^*_{PI} \times f^{flat}_{PI} \times \prod g^{true-dist}_{LM}$	-5146.80	10344.07	10309.87
$g^{*\,0objs}_{PI} \times f^*_{PI} \times \prod g^{true-dist}_{LM}$	-3867.54	7785.55	7751.35
$g^{*\,0objs}_{PI} \times f^{*\,0objs}_{PI} \times \prod g^{true-dist}_{LM}$	-4409.39	8869.25	8835.05
$g^{*\,0objs}_{PI} \times f^{flat}_{PI} \times \prod g^{true-dist}_{LM}$	-5352.77	10756.00	10721.81

Table 6. Summary table of MLE model fits for the three object condition. For an overview of model components see Table 2. Model fits within each subdivision can be compared directly, since they belong the the same model family.

model ID	log Likelihood	BIC	AIC
$g_{PI}^{*\,0objs}$	-1691.17	3394.95	3386.35
g_{PI}^{*}	-1403.12	2818.84	2810.23
f_{PI}^{flat}	-1005.32	2023.25	2014.64
$f_{PI}^{*\,0objs}$	-63.55	139.70	131.09
f_{PI}^{*}	559.06	-1105.50	-1114.11
$\prod g_{LM}^{true-dist}$	-4033.60	8105.04	8079.37
$g_{PI}^{*} \times f_{PI}^{*} \times \prod g_{LM}^{true-dist}$	-4877.66	9818.37	9775.74
$g_{PI}^{*} \times f_{PI}^{*\,0objs} \times \prod g_{LM}^{true-dist}$	-5500.27	11063.58	11020.95
$g_{PI}^{*} \times f_{PI}^{flat} \times \prod g_{LM}^{true-dist}$	-6442.04	12947.12	12904.49
$g_{PI}^{*\,0objs} \times f_{PI}^{*} \times \prod g_{LM}^{true-dist}$	-5165.72	10394.49	10351.85
$g_{PI}^{*\,0objs} \times f_{PI}^{*\,0objs} \times \prod g_{LM}^{true-dist}$	-5788.32	11639.69	11597.06
$g_{PI}^{*\,0objs} \times f_{PI}^{flat} \times \prod g_{LM}^{true-dist}$	-6730.10	13523.24	13480.60

Table 7. Summary table of MLE model fits for the ten object condition. For an overview of model components see Table 2. Model fits within each subdivision can be compared directly, since they belong the the same model family.

model ID	log Likelihood	BIC	AIC
$g_{PI}^{*\,0objs}$	-1732.74	3478.10	3469.48
g_{PI}^{*}	-1551.43	3115.49	3106.86
f_{PI}^{flat}	-1012.67	2037.96	2029.34
$f_{PI}^{*\,0objs}$	-62.93	138.48	129.85
f_{PI}^{*}	359.59	-706.56	-715.18
$\prod g_{LM}^{true-dist}$	-41593.09	83312.42	83227.77
$\sum g_{LM}^{true-dist}$	-423.04	972.32	887.67
$g_{PI}^{*} \times f_{PI}^{*} \times \sum g_{LM}^{true-dist}$	-1614.88	3381.25	3280.05
$g_{PI}^{*} \times f_{PI}^{*\,0objs} \times \sum g_{LM}^{true-dist}$	-2037.40	4226.28	4125.08
$g_{PI}^{*} \times f_{PI}^{flat} \times \sum g_{LM}^{true-dist}$	-2987.15	6125.77	6024.57
$g_{PI}^{*\,0objs} \times f_{PI}^{*} \times \sum g_{LM}^{true-dist}$	-1796.19	3743.86	3642.66
$g_{PI}^{*\,0objs} \times f_{PI}^{*\,0objs} \times \sum g_{LM}^{true-dist}$	-2218.71	4588.90	4487.70
$g_{PI}^{*\,0objs} \times f_{PI}^{flat} \times \sum g_{LM}^{true-dist}$	-3168.45	6488.38	6387.18

Table 8. Summary table of MLE model fits for the 99 object condition. For an overview of model components see Table 2. Model fits within each subdivision can be compared directly, since they belong the the same model family.

model ID	log Likelihood	BIC	AIC
$g_{PI}^{*\,0objs}$	-1732.01	3476.57	3468.02
g_{PI}^{*}	-1696.02	3404.59	3396.04
f_{PI}^{flat}	-974.07	1960.70	1952.15
$f_{PI}^{*\,0objs}$	-231.06	474.66	466.12
f_{PI}^{*}	-215.49	443.53	434.99
$\prod g_{LM}^{true-dist}$	-inf	inf	inf
$\sum g_{LM}^{true-dist}$	56.78	1128.46	520.51
$g_{PI}^{*} \times f_{PI}^{*} \times \sum g_{LM}^{true-dist}$	-1854.73	4976.59	4364.27
$g_{PI}^{*} \times f_{PI}^{*\,0objs} \times \sum g_{LM}^{true-dist}$	-1870.30	5007.71	4395.39
$g_{PI}^{*} \times f_{PI}^{flat} \times \sum g_{LM}^{true-dist}$	-2613.31	6493.75	5881.43
$g_{PI}^{*\,0objs} \times f_{PI}^{*} \times \sum g_{LM}^{true-dist}$	-1890.72	5048.56	4436.24
$g_{PI}^{*\,0objs} \times f_{PI}^{*\,0objs} \times \sum g_{LM}^{true-dist}$	-1906.29	5079.69	4467.37
$g_{PI}^{*\,0objs} \times f_{PI}^{flat} \times \sum g_{LM}^{true-dist}$	-2649.30	6565.73	5953.41

6 Acknowledgments

We thank Tim Schmoll and the Stats Club at Bielefeld University for valuable input for the statistical analysis of this study.

References

1. Gallistel C, Cramer AE. Computations on metric maps in mammals: getting oriented and choosing a multi-destination route. The Journal of Experimental Biology. 1996;199(1):211–217.

2. Moser EI, Moser MB, McNaughton BL. Spatial representation in the hippocampal formation: a history. Nature neuroscience. 2017;20(11):1448–1464.

3. Peer M, Brunec IK, Newcombe NS, Epstein RA. Structuring knowledge with cognitive maps and cognitive graphs. Trends in cognitive sciences. 2021;25(1):37–54.

4. Warren WH. Non-euclidean navigation. Journal of Experimental Biology. 2019;222(Suppl_1):jeb187971.

5. Meilinger T. The network of reference frames theory: A synthesis of graphs and cognitive maps. In: Spatial Cognition VI. Learning, Reasoning, and Talking about Space: International Conference Spatial Cognition 2008, Freiburg, Germany, September 15-19, 2008. Proceedings 6. Springer; 2008. p. 344–360.

6. Chen X, McNamara TP, Kelly JW, Wolbers T. Cue combination in human spatial navigation. Cognitive Psychology. 2017;95:105–144.

7. Zhao M, Warren WH. How you get there from here: Interaction of visual landmarks and path integration in human navigation. Psychological science. 2015;26(6):915–924.

8. Mallot HA, Lancier S. Place recognition from distant landmarks: human performance and maximum likelihood model. Biological cybernetics. 2018;112(4):291–303.

9. Jetzschke S, Ernst MO, Froehlich J, Boeddeker N. Finding home: Landmark ambiguity in human navigation. Frontiers in behavioral neuroscience. 2017;11:132.

10. Jetzschke S, Ernst MO, Moscatelli A, Boeddeker N. Going round the bend: Persistent personal biases in walked angles. Neuroscience letters. 2016;617:72–75.

11. Harootonian SK, Ekstrom AD, Wilson RC. Combination and competition between path integration and landmark navigation in the estimation of heading direction. PLoS computational biology. 2022;18(2):e1009222.

12. Chrastil ER, Warren WH. Executing the homebound path is a major source of error in homing by path integration. Journal of Experimental Psychology: Human Perception and Performance. 2021;47(1):13.

13. Etienne AS, Jeffery KJ. Path integration in mammals. Hippocampus. 2004;14(2):180–192.

14. Walter JL, Essmann L, König SU, König P. Finding landmarks-An investigation of viewing behavior during spatial navigation in VR using a graph-theoretical analysis approach. PLoS Computational Biology. 2022;18(6):e1009485.

15. Wiener JM, Mallot HA. Path complexity does not impair visual path integration. Spatial cognition and computation. 2006;6(4):333–346.

16. Chrastil ER, Nicora GL, Huang A. Vision and proprioception make equal contributions to path integration in a novel homing task. Cognition. 2019;192:103998.

17. Hoinville T, Wehner R. Optimal multiguidance integration in insect navigation. Proceedings of the National Academy of Sciences. 2018;115(11):2824–2829.

18. Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. Nature. 2002;415(6870):429–433.

19. Alais D, Burr D. Cue combination within a Bayesian framework. Multisensory processes: The auditory perspective. 2019; p. 9–31.

20. Roy C, Wiebusch D, Botsch M, Ernst MO. Did it move? Humans use spatio-temporal landmark permanency efficiently for navigation. Journal of Experimental Psychology: General. 2022;.

21. McNamara TP, Chen X. Bayesian decision theory and navigation. Psychonomic Bulletin & Review. 2022; p. 1–32.

22. Sjolund LA, Kelly JW, McNamara TP. Optimal combination of environmental cues and path integration during navigation. Memory & Cognition. 2018;46:89–99.

23. Zhao M, Warren WH. Environmental stability modulates the role of path integration in human navigation. Cognition. 2015;142:96–109.

24. Müller M, Wehner R. Path integration in desert ants, Cataglyphis fortis. Proceedings of the National Academy of Sciences. 1988;85(14):5287–5290.

25. Wystrach A, Mangan M, Webb B. Optimal cue integration in ants. Proceedings of the Royal Society B: Biological Sciences. 2015;282(1816):20151484.

26. Widdowson C, Wang RF. Human navigation in curved spaces. Cognition. 2022;218:104923.

27. Loomis JM, Klatzky RL, Golledge RG, Cicinelli JG, Pellegrino JW, Fry PA. Nonvisual navigation by blind and sighted: assessment of path integration ability. Journal of Experimental Psychology: General. 1993;122(1):73.

28. Harootonian SK, Wilson RC, Hejtmánek L, Ziskin EM, Ekstrom AD. Path integration in large-scale space and with novel geometries: Comparing vector addition and encoding-error models. PLoS computational biology. 2020;16(5):e1007489.

29. Kearns MJ, Warren WH, Duchon AP, Tarr MJ. Path integration from optic flow and body senses in a homing task. Perception. 2002;31(3):349–374.

30. Glasauer S, Amorim MA, Viaud-Delmon I, Berthoz A. Differential effects of labyrinthine dysfunction on distance and direction during blindfolded walking of a triangular path. Experimental Brain Research. 2002;145:489–497.

31. Wolbers T, Wiener JM, Mallot HA, Büchel C. Differential recruitment of the hippocampus, medial prefrontal cortex, and the human motion complex during path integration in humans. Journal of Neuroscience. 2007;27(35):9408–9416.

32. Xie Y, Bigelow RT, Frankenthaler SF, Studenski SA, Moffat SD, Agrawal Y. Vestibular loss in older adults is associated with impaired spatial navigation: data from the triangle completion task. Frontiers in neurology. 2017;8:173.

33. Glasauer S, Shi Z. Individual beliefs about temporal continuity explain variation of perceptual biases. Scientific Reports. 2022;12(1):10746.

34. Zanchi S, Cuturi LF, Sandini G, Gori M. Interindividual differences influence multisensory processing during spatial navigation. Journal of Experimental Psychology: Human Perception and Performance. 2022;48(2):174.

35. Schielzeth H, Dingemanse NJ, Nakagawa S, Westneat DF, Allegue H, Teplitsky C, et al. Robustness of linear mixed-effects models to violations of distributional assumptions. Methods in ecology and evolution. 2020;11(9):1141–1152.

36. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in ecology and evolution. 2013;4(2):133–142.

37. Johnson PC. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. Methods in ecology and evolution. 2014;5(9):944–946.

38. UCLA Statistical Consulting Group. R Library Contrast Coding Systems For Categorical Variables; 2021. `https://stats.oarc.ucla.edu/r/library/r-library-contrast-coding-systems-for-categorical-variables/`.

39. Murray RF, Morgenstern Y. Cue combination on the circle and the sphere. Journal of vision. 2010;10(11):15–15.

40. Rohde M, van Dam LC, Ernst MO. Statistically optimal multisensory cue integration: A practical tutorial. Multisensory research. 2016;29(4-5):279–317.

41. Wolbers T, Hegarty M. What determines our navigational abilities? Trends in cognitive sciences. 2010;14(3):138–146.

42. Weisberg SM, Newcombe NS. Cognitive maps: Some people make them, some people struggle. Current directions in psychological science. 2018;27(4):220–226.

43. Körding KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. PLoS one. 2007;2(9):e943.

44. Kayser C, Shams L. Multisensory causal inference in the brain. PLoS biology. 2015;13(2):e1002075.

45. Montello DR. Landmarks are exaggerated. KI-Künstliche Intelligenz. 2017;31:193–197.

46. Chrastil ER, Sherrill KR, Aselcioglu I, Hasselmo ME, Stern CE. Individual differences in human path integration abilities correlate with gray matter volume in retrosplenial cortex, hippocampus, and medial prefrontal cortex. Eneuro. 2017;4(2).

47. Sherrill KR, Chrastil ER, Aselcioglu I, Hasselmo ME, Stern CE. Structural differences in hippocampal and entorhinal gray matter volume support individual differences in first person navigational ability. Neuroscience. 2018;380:123–131.

48. Müller M, Scherer J, Unterbrink P, Bertrand OJN, Egelhaaf M, Boeddeker N. The Virtual Navigation Toolbox: Providing tools for virtualnavigation experiments. PLoS One. 2023;Forthcoming.

49. R Core Team. R: A Language and Environment for Statistical Computing; 2022. Available from: `https://www.R-project.org/`.

50. Python Software Foundation. The Python Language Reference; 2020. `https://docs.python.org/3.9/reference/index.html`.

51. Faul F, Erdfelder E, Lang AG, Buchner A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods. 2007;39(2):175–191.

52. Nardini M, Jones P, Bedford R, Braddick O. Development of cue integration in human navigation. Current Biology. 2008;18(9):689–693. doi:10.1016/j.cub.2008.04.021.

53. Goeke CM, Planera S, Finger H, König P. Bayesian alternation during tactile augmentation. Frontiers in Behavioral Neuroscience. 2016;10. doi:10.3389/fnbeh.2016.00187.

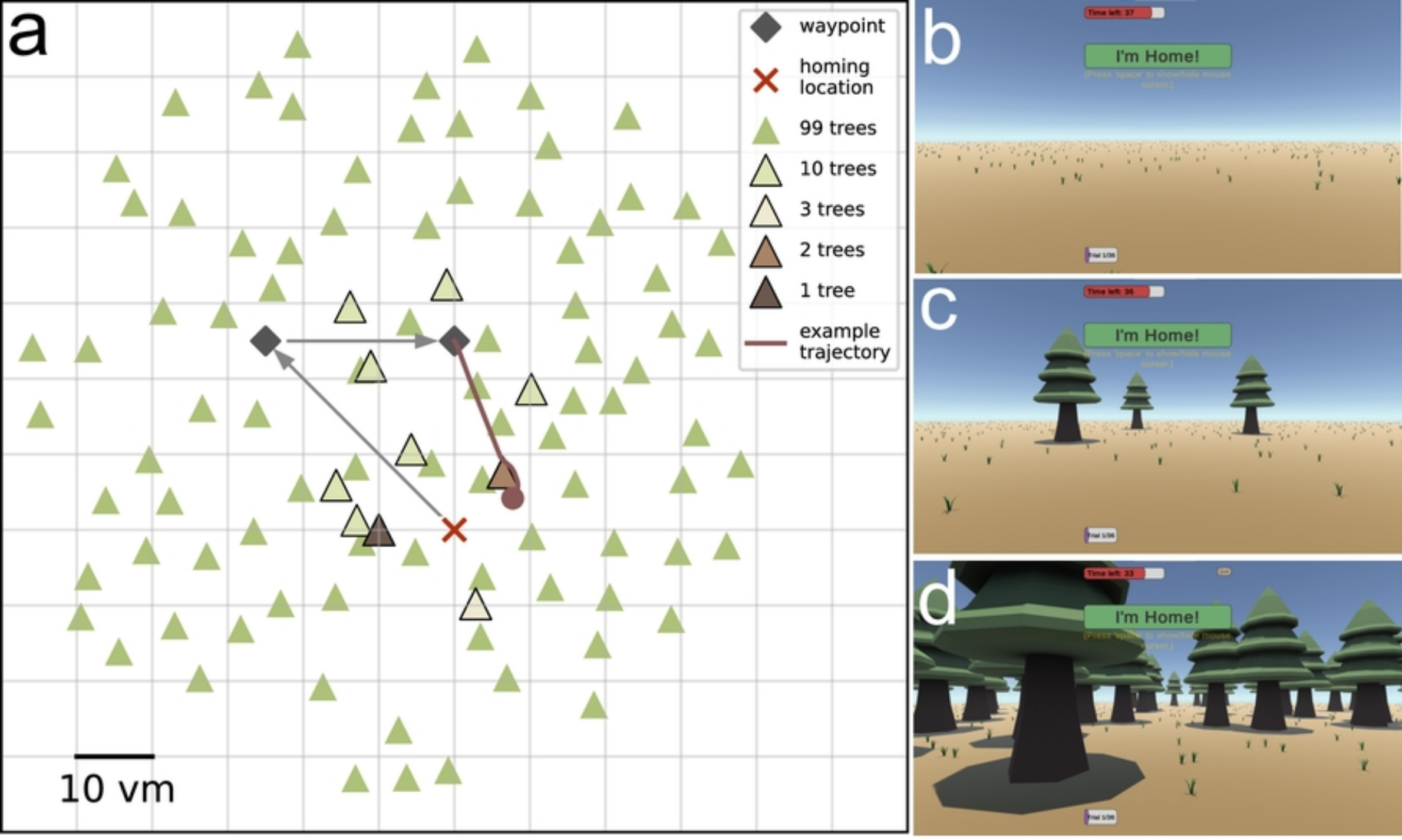
a
waypoint
homing location
99 trees
10 trees
3 trees
2 trees
1 tree
example trajectory
10 vm
b
I'm Home!
c
I'm Home!
d
I'm Home!

Figure 1

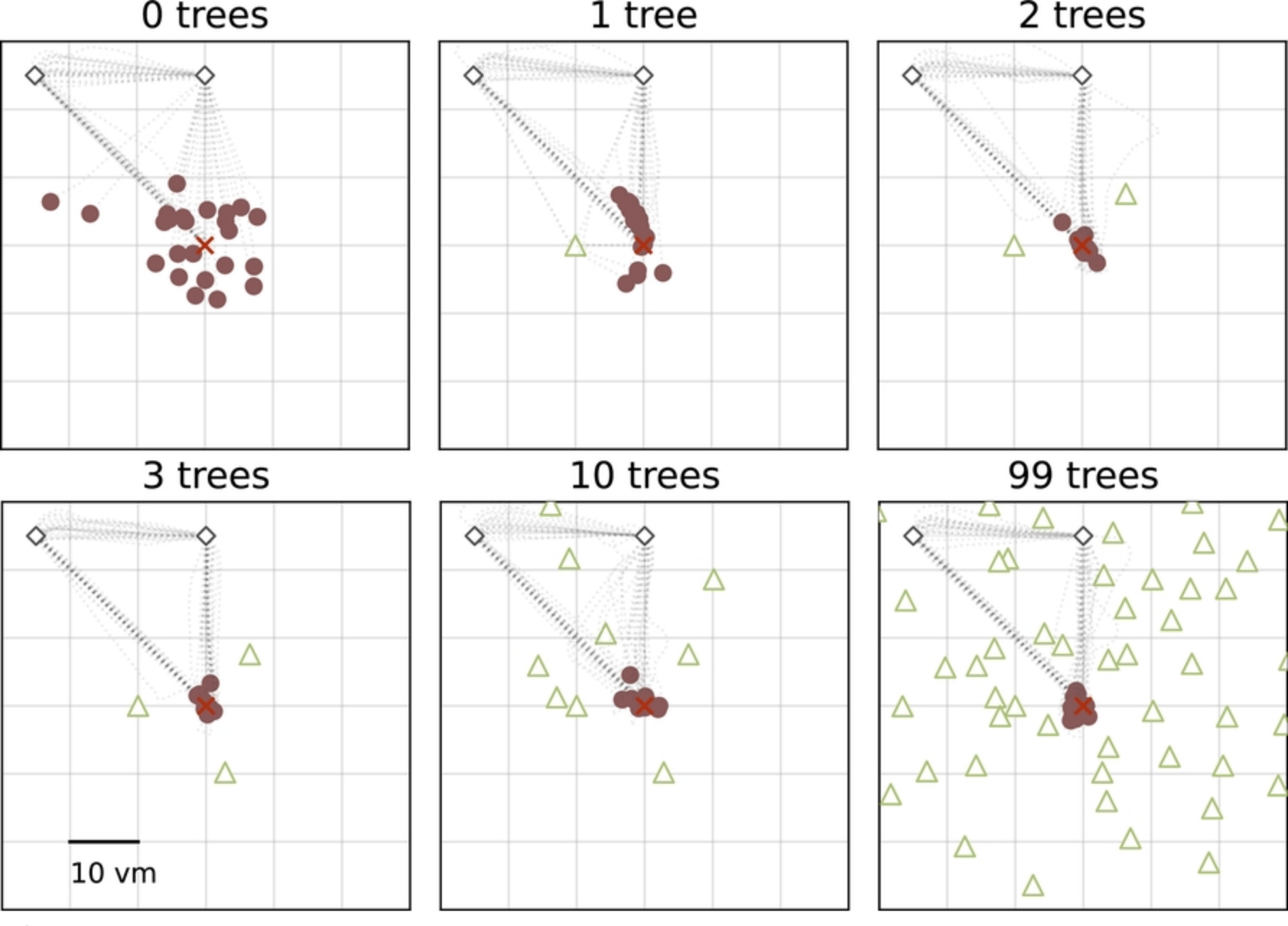

Figure 2

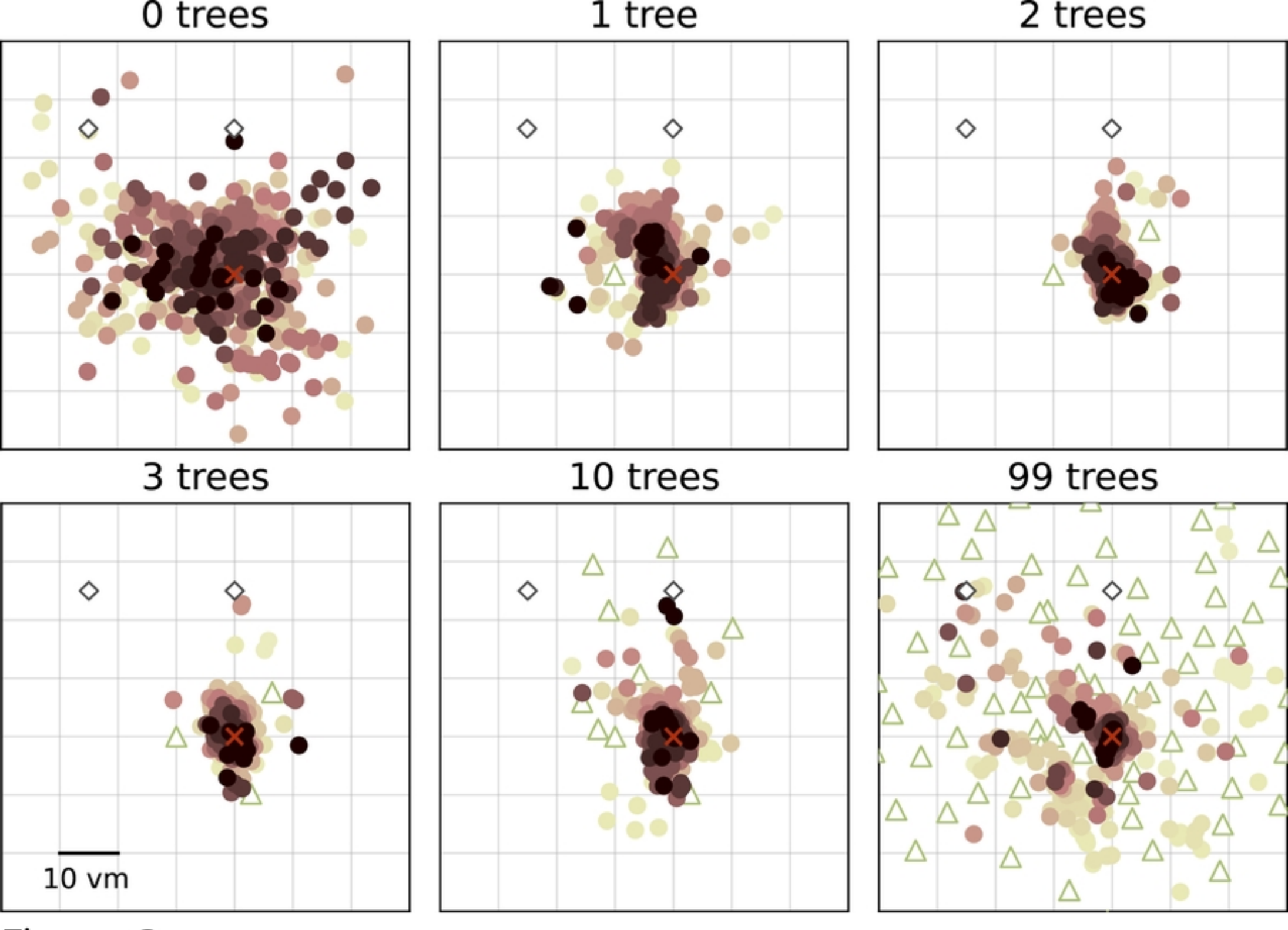

Figure 3

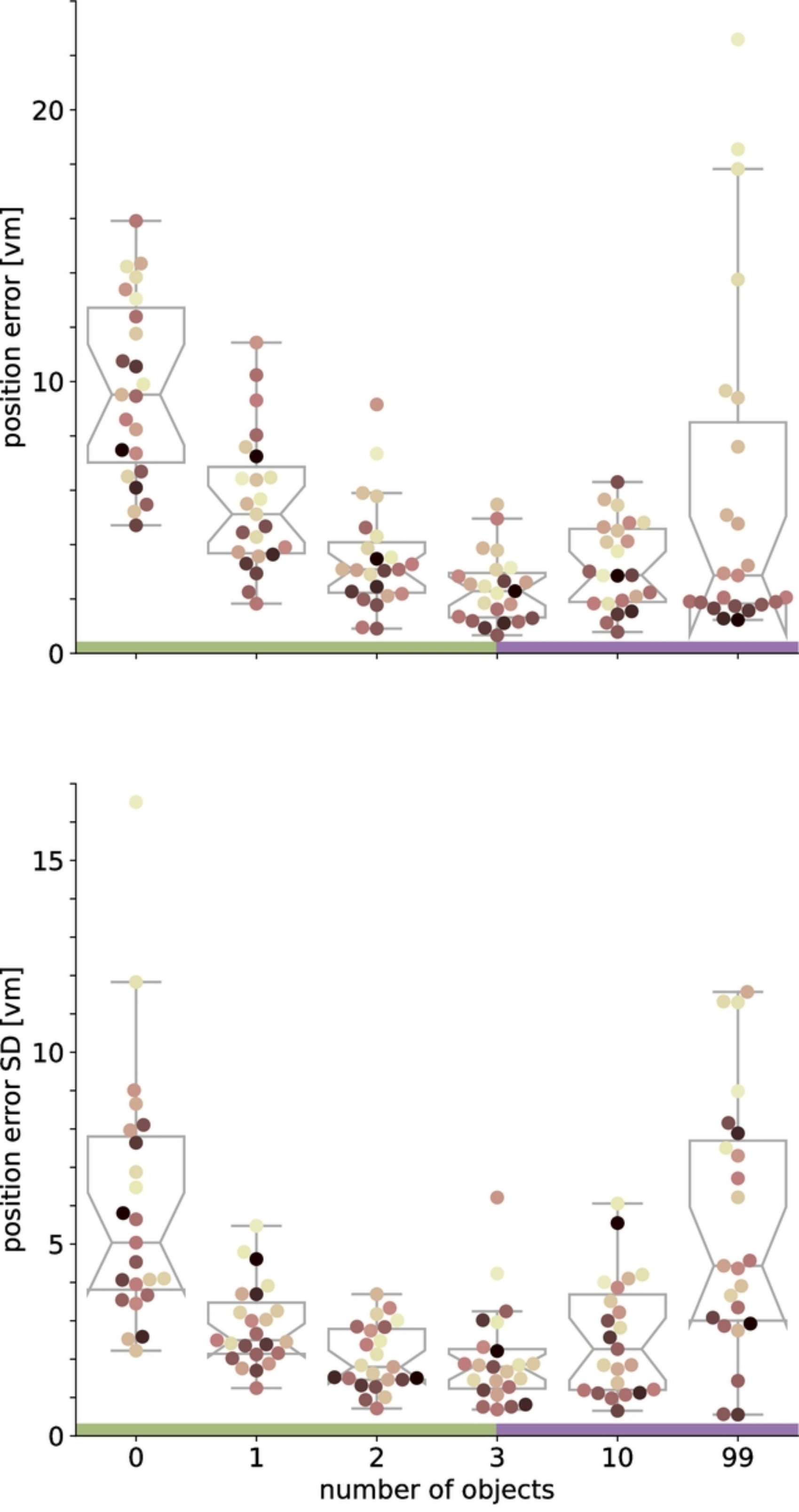
position error [vm]
0
10
20
position error SD [vm]
0
5
10
15
0
1
2
3
10
99
number of objects

Figure 4

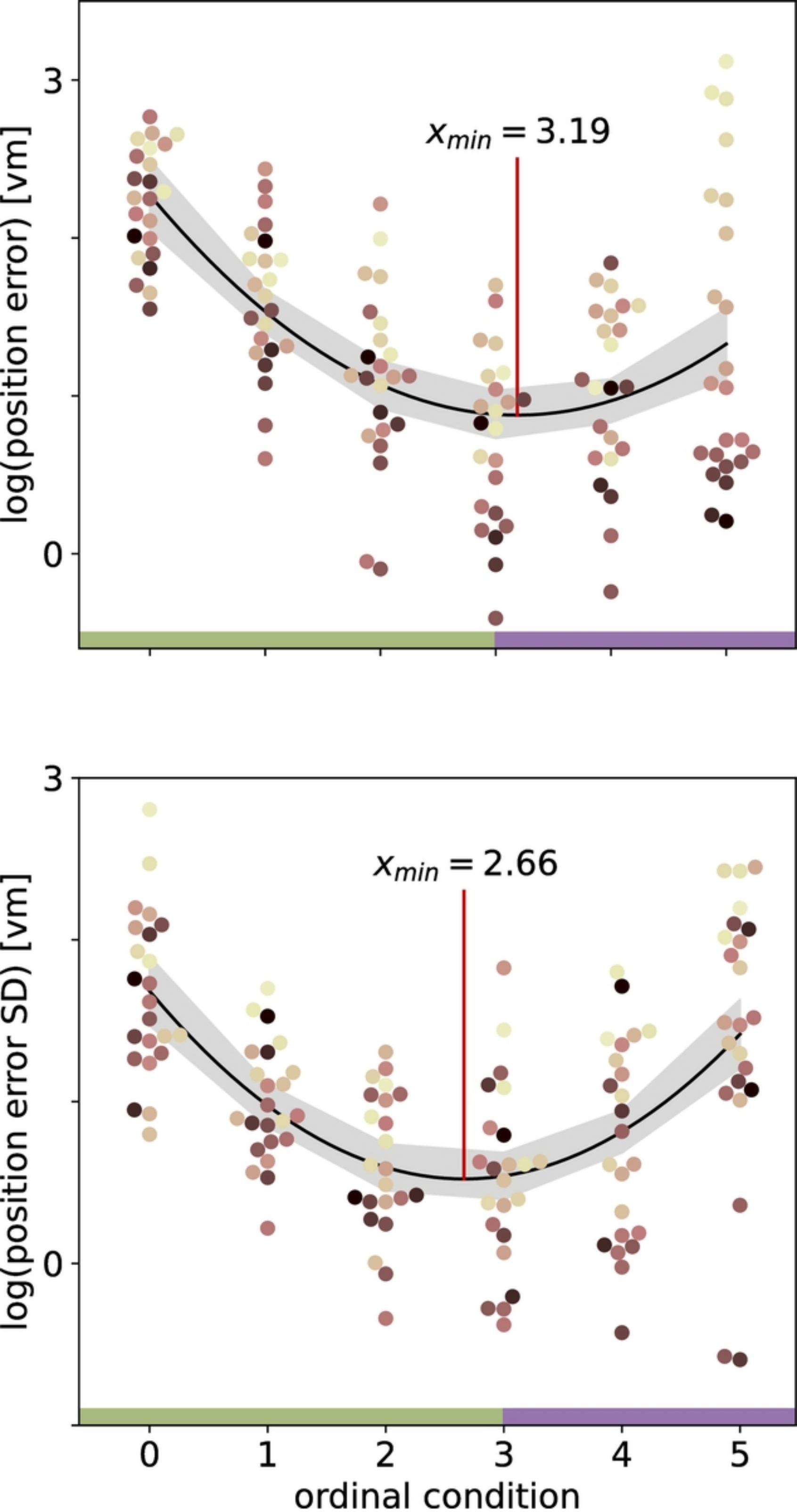
$x_{min} = 3.19$
log(position error) [vm]
3
0
$x_{min} = 2.66$
log(position error SD) [vm]
3
0
0
1
2
3
4
5
ordinal condition

Figure 5

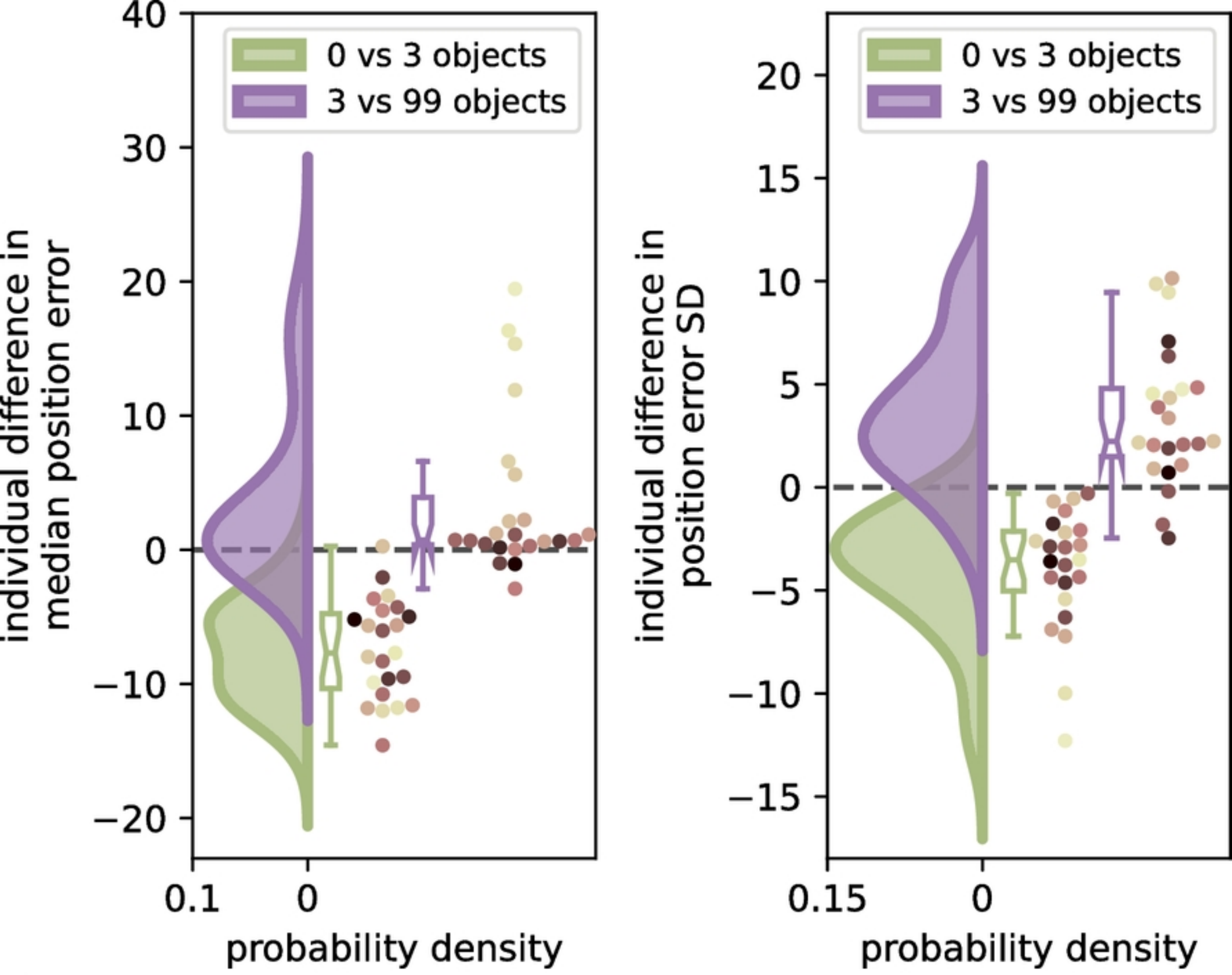
0 vs 3 objects
3 vs 99 objects
individual difference in median position error
40
30
20
10
0
−10
−20
0.1
0
probability density
0 vs 3 objects
3 vs 99 objects
individual difference in position error SD
20
15
10
5
0
−5
−10
−15
0.15
0
probability density

Figure 6

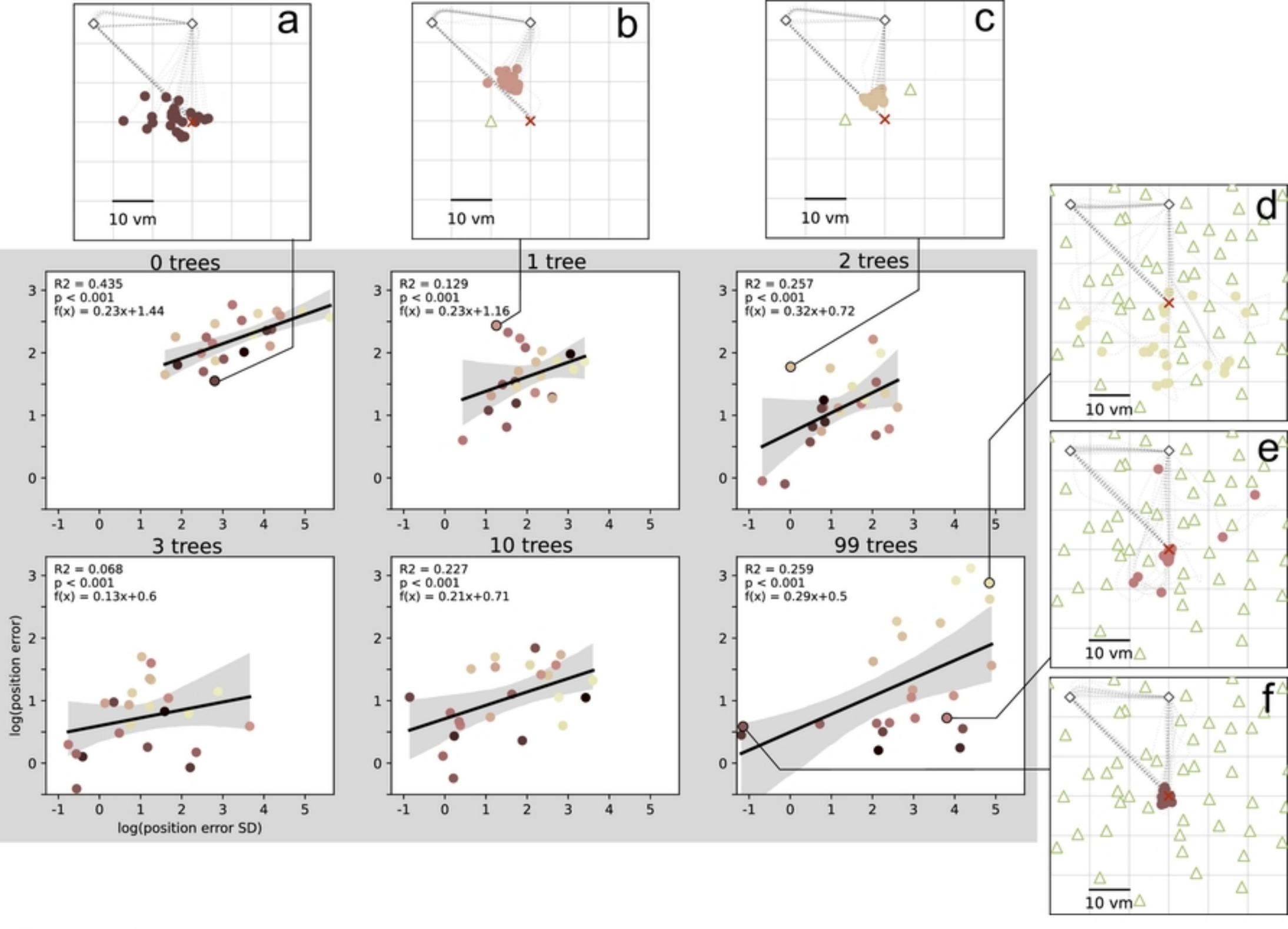
a
b
c
d
e
f
10 vm
0 trees
1 tree
2 trees
3 trees
10 trees
99 trees
R2 = 0.435
p < 0.001
f(x) = 0.23x+1.44
R2 = 0.129
p < 0.001
f(x) = 0.23x+1.16
R2 = 0.257
p < 0.001
f(x) = 0.32x+0.72
R2 = 0.068
p < 0.001
f(x) = 0.13x+0.6
R2 = 0.227
p < 0.001
f(x) = 0.21x+0.71
R2 = 0.259
p < 0.001
f(x) = 0.29x+0.5
log(position error)
log(position error SD)

Figure 7

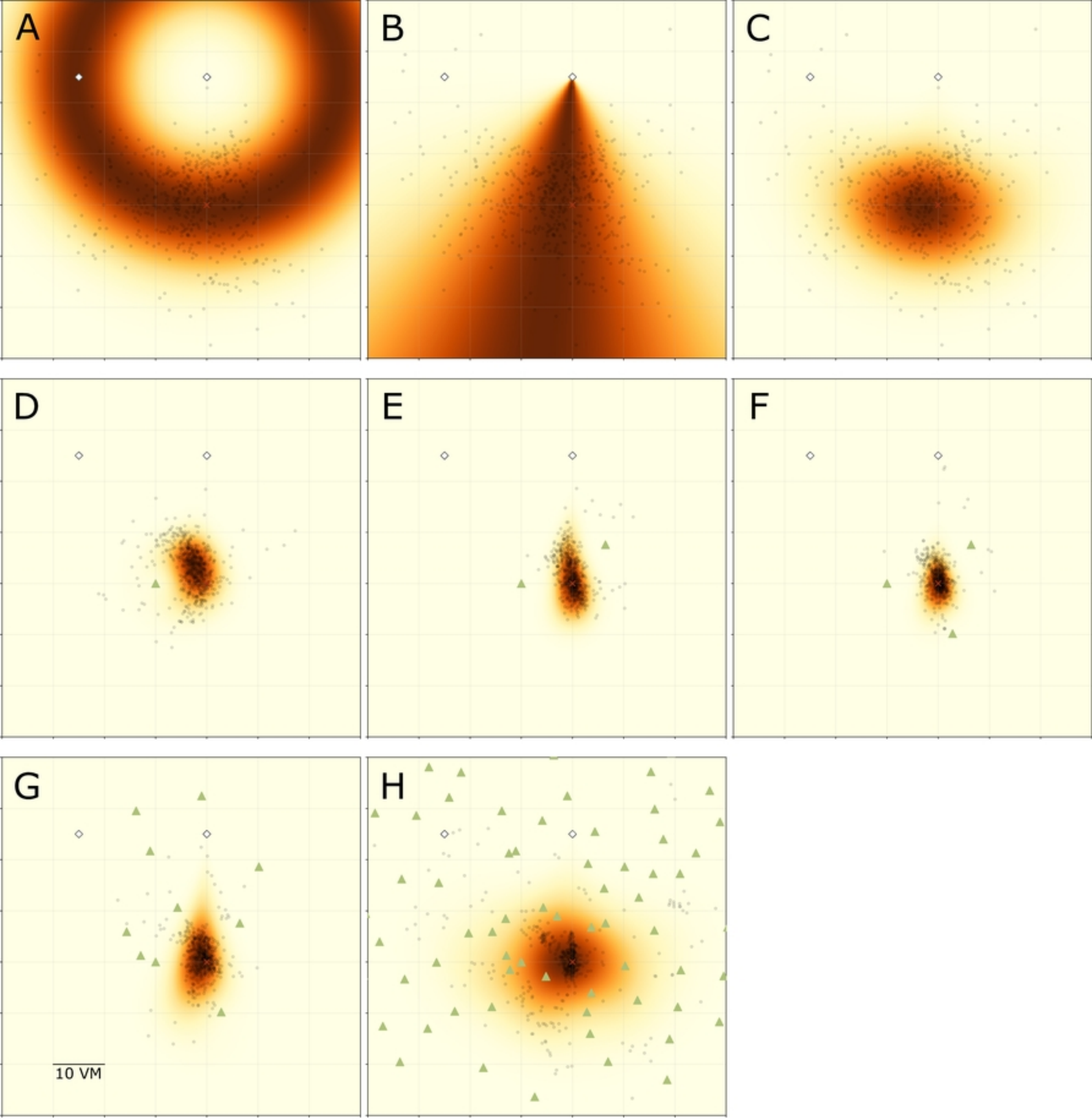

Figure 8

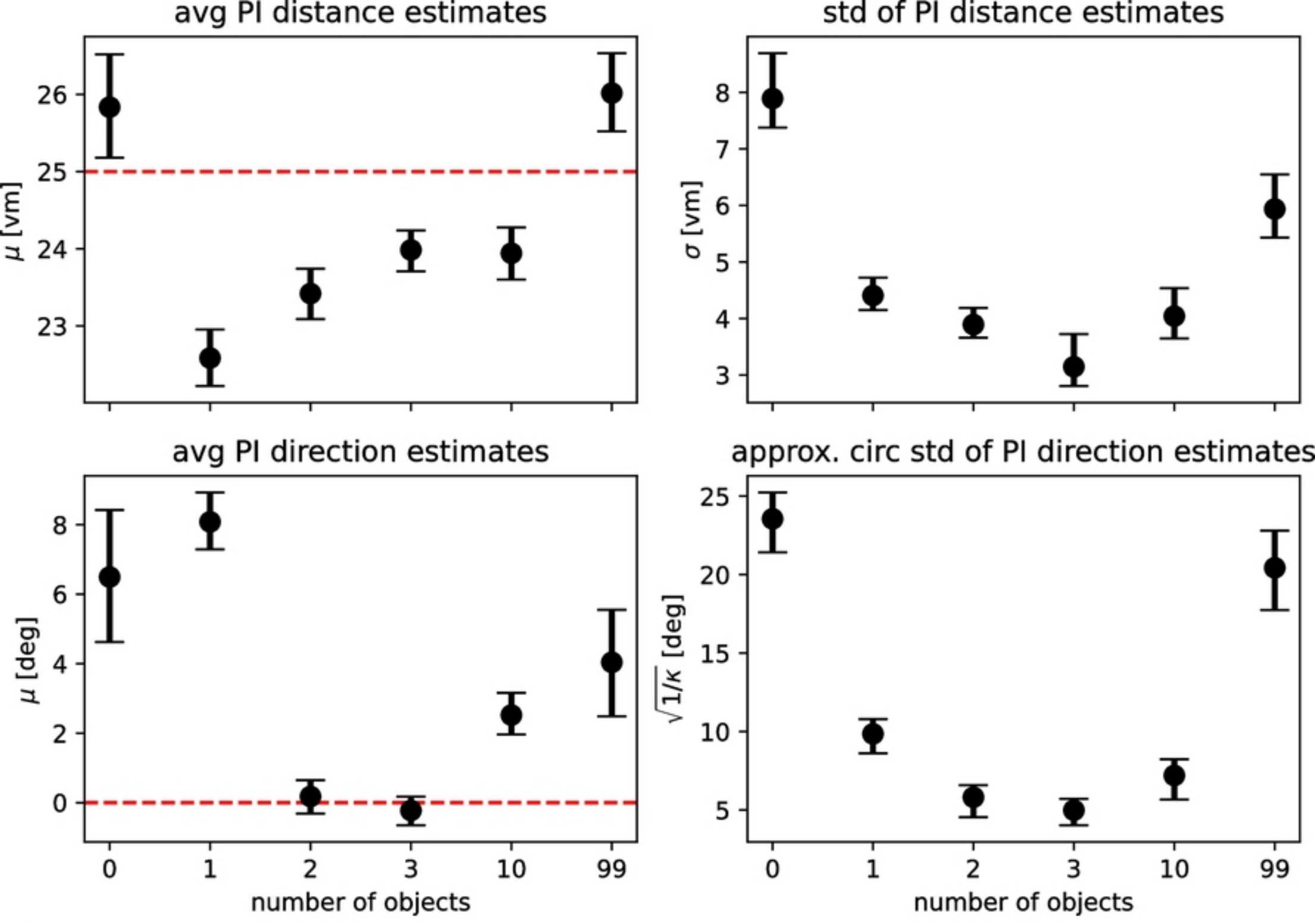
avg PI distance estimates
μ [vm]
26
25
24
23
std of PI distance estimates
σ [vm]
8
7
6
5
4
3
avg PI direction estimates
μ [deg]
8
6
4
2
0
approx. circ std of PI direction estimates
√1/κ [deg]
25
20
15
10
5
0
1
2
3
10
99
number of objects
number of objects

Figure 9

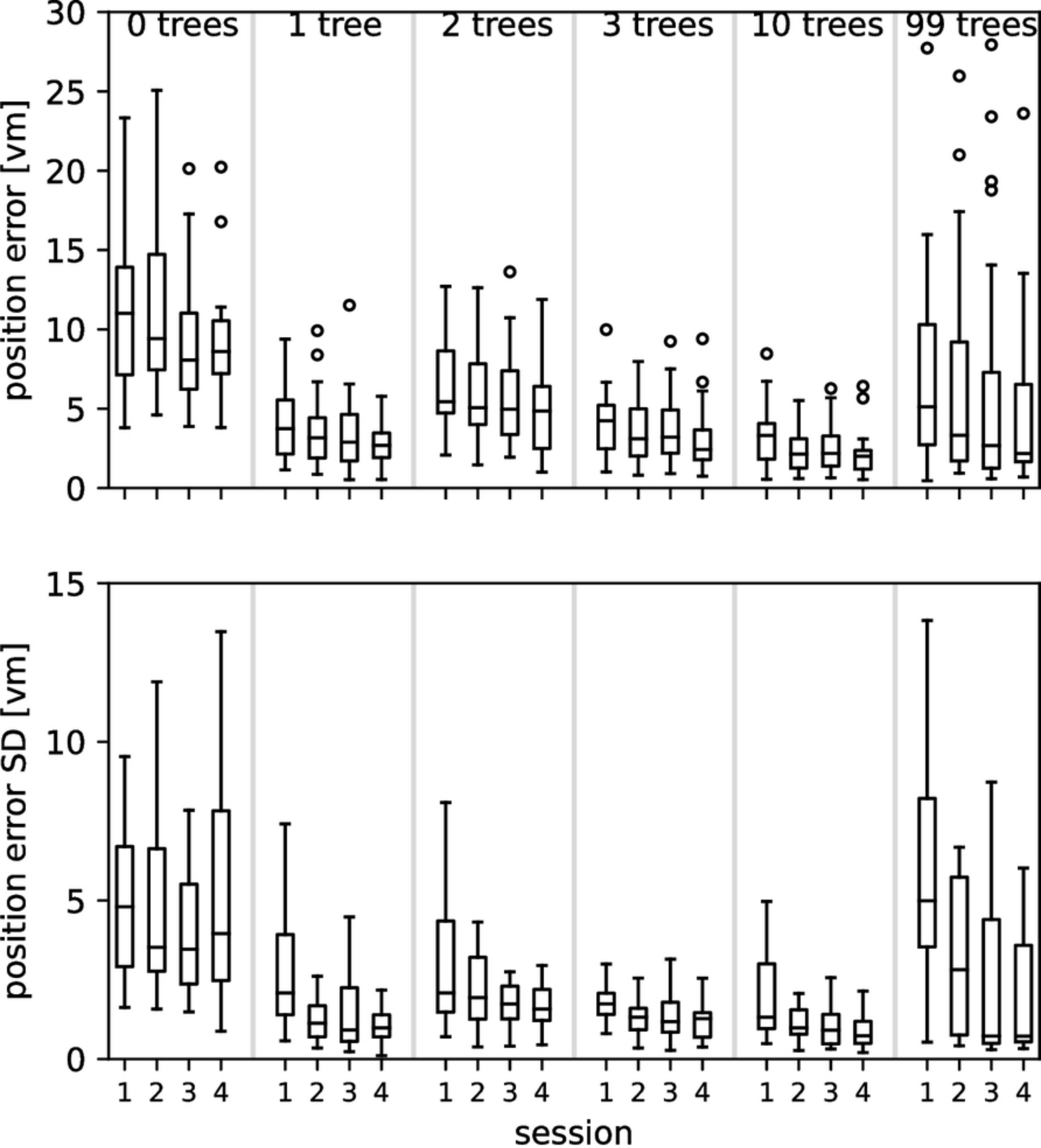
0 trees
1 tree
2 trees
3 trees
10 trees
99 trees
position error [vm]
position error SD [vm]
session
0
5
10
15
20
25
30
1 2 3 4

Figure S2

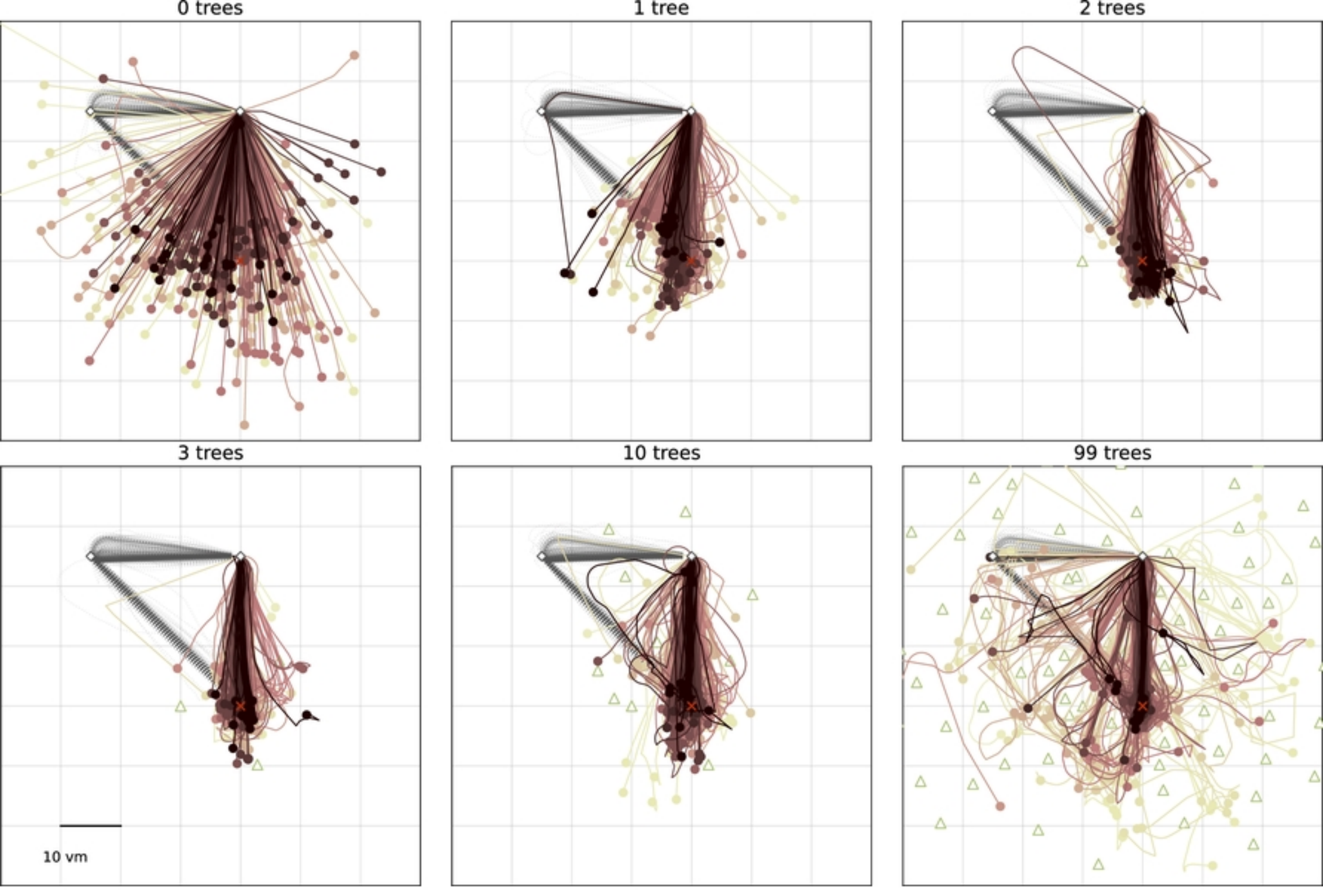

Figure S1