GIVEN THE BIRDS, WHERE IS THE FLOCK?

Visual estimation of the location of collections of points

Keiji Ota[1,2,3,4*] Qihan Wu[5] Pascal Mamassian[6] Laurence T Maloney[3,4]

[1] Centre for Brain and Behaviour, Department of Psychology, Queen Mary University of London

[2] Institute of Cognitive Neuroscience, University College London

[3] Department of Psychology, New York University, New York, United States

[4] Center for Neural Science, New York University, New York, United States

[5] Department of Psychological and Brain Sciences, Johns Hopkins University

[6] Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France

* Corresponding author

Keywords

Location estimation, Gestalt, perceptual organization, clustering, non-Gaussian distributions

Abstract

Many visual tasks resemble statistical judgments, estimating the center of a 'cloud' of identical items for example. In statistical form, the observer sees a sample drawn from an invisible probability density function (PDF) : the observer attempts to estimate the location of the PDF. He receives larger rewards for lower variance. Optimal minimum variance estimators are different for different distributions. We evaluated performance for three location families (Gaussian, Laplacian, and Uniform). How does human perceptual organization 'stack up' against optimal perceptual estimation? We found that observers adopted distinct estimators for distinct distribution families, each approximately minimizing variance for its family. To explain observers' ability to adapt to different distributional families, we propose that observers adopt a single estimator based on clusters in the sample. This Visual Cluster Model accounted for performance across distributions, suggesting that perceptual grouping first aggregates 'parts' into clusters and only then into a 'whole'.

[148 words]

Significance statement

Imagine a flock of birds. We can predict where additional birds might appear by estimating the probability density function (PDF) of the birds. We compared performance in estimating the location of PDFs drawn from Gaussian, Laplacian or Uniform location families, rewarding lower variance. The optimal estimators for the three families are different. Observers approximately minimized variance for each condition and used different estimators for different families. How could they do so? We propose that observers use a single estimator based on sample clusters: perceptual grouping first aggregates ‘parts’ into clusters and only then into a ‘whole’.

[96 Words]

Introduction

The human visual system segments scenes into visual objects (Koffka, 1935/2013; Wagemans, Elder, et al., 2012; Wagemans, Feldman, et al., 2012) adding visual structure to what are initially only patterns of light across the retinas. Such 'wholes made of parts' have their own spatial locations in 1D, 2D or 3D which the observer can be asked to estimate. If the observer is asked, for example, to saccade to a 2D cloud of points, the saccade can be interpreted as an estimate of the center of the cloud (McGowan et al., 1998; Melcher & Kowler, 1999). *The first question is, how is this location estimate computed from the individual points?* Birds in flight form a flock in 3D (Figure 1A) but typically no one bird marks the center of the flock. There is no obvious answer to the problem as posed.

We focus on a statistical version of the task where the 'parts' form a 1D sample drawn from a 1D statistical distribution and – importantly – we can compare human performance to normative. Different families of pdfs will serve as generators

Not every perceptual task can be modeled as a statistical task where 'parts' are a sample drawn from a partially-known PDF, but considering such tasks allows us to use powerful statistical tools in a special case.

Other Gestalt phenomena such as border ownership (D'Angelo et al., 2025), object segmentation (Chen et al., 2025; Jeurissen et al., 2024; Papale et al., 2024) and shape perception (Nielsen & Connor, 2024) and more are also part of 'middle vision' (Ullman, 1984), the stage in visual processing that follows encoding of initial visual information into retinotopic areas and links local features to the items in the environment that generated them (Groen et al., 2017; Peirce, 2015).

Terminology: The collection of points (‘*parts*’) form a *sample* drawn from a one-dimensional *probability density function* (PDF), the *generator*. In Figure 1B, the red curves, dashed and solid, are copies of the same PDF with the red, solid PDF being the PDF used to generate the sample. They represent a *location family*, a collection of PDFs differing only in location (e.g. Figure 1B). The colored dots are a 1D sample from the solid PDF. The location estimation problem is a problem in parametric point estimation (Lehmann, 1998; Mood et al., 1974) as summarized in Figure 1C.

We address three questions: (1) *How does the observer estimate the location of the PDF given only the sample and knowledge of the location family which the PDF belongs to?* (2) *How does changing the PDF (‘generator’) underlying the location family affect the location estimation – if it has any effect at all? And* (3) *How does human location estimation compare to uniform minimum variance unbiased estimation (UMVUE) for each of the distributions we consider?*

The three location families we consider are the Gaussian, the Laplacian and the Uniform, illustrated below (Figure 2). A possible outcome is that the observer uses the same estimator, possibly the Gaussian, for all three distributions. A second possibility is that the observer uses different estimators for the different location distributions but at least one of them is not the UMVUE of the corresponding distribution. A final possibility is that the observer chooses estimators close to the UMVUE for each of three location distributions, somehow adapting its estimation algorithm to the distribution (generator).

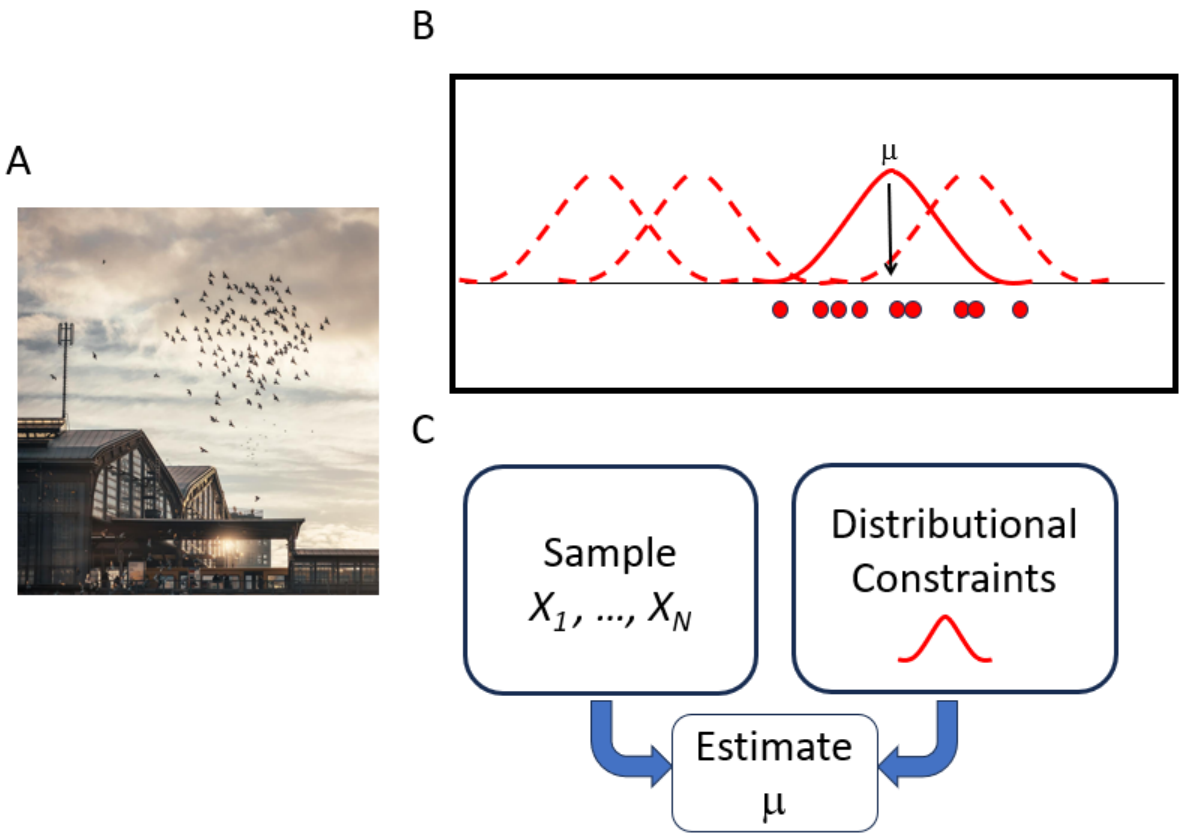


Figure 1: Estimating the Location of a Collection of Points. A. A Flock of Birds. A flock of birds in flight. **B. The one-dimensional Gaussian Location Task.** The dashed and solid red curves are examples of Gaussian probability density functions (PDFs) with unit variance for different values of the location parameter μ. All of the red curves together comprise a *Gaussian location family*. The solid red curve is the actual PDF from which the sample was drawn. The sample is composed of *N* positions and is plotted as red points along the horizontal axis. The observer sees only the sample $X_1, \ldots, X_N$ and is asked to estimate the value of the location parameter μ corresponding to the PDF that was used to generate the sample. **C. 1D Parametric Estimation.** The observer sees the individual points and knows that the points are a sample from a PDF in the Gaussian location family with unit variance. He estimates the unknown location of the generating PDF and knowledge of the PDF may influence his choice of estimator.

In the present study, we consider three different location-scale families of probability density functions (PDFs). One of these families is the Gaussian PDF, commonly used in visual modeling, notably cue combination models (Landy et al., 2007; Oruç et al., 2003). A second motivation for using Gaussian distributions in modeling visual judgments is that actual visual and motor error is close to Gaussian in a wide class of everyday tasks (Ota et al., 2019; Tanae et al., 2021; Trommershauser et al., 2008). Humans frequently encounter tasks involving Gaussian random variables and they may tend to apply estimation rules and other statistical procedures that are appropriate when distributions are Gaussian (Tassinari et al., 2006; Vilares et al., 2012). However, it is less well known how humans behave when distributions are non-Gaussian (see (Maloney & Thomas, 1991) for an exception).

Figure 1C summarizes what the observer in our experiment knows – the locations of points and knowledge of the distributional family from which they are drawn but not the precise PDF. *Parametric point estimation* methods (Lehmann, 1998; Mood et al., 1974) allow us to compute optimal estimators for each PDF family based on criteria such as minimum variance for a wide range of distributions.

If the PDF is Gaussian with fixed variance 1 and with unknown *location* parameter μ, the PDF is

$$\phi_\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}} . \quad (1)$$

Several examples of PDFs from a *Gaussian location family* (Eq. 1) are plotted as dashed red curves in Figure 1B. As μ is varied, the PDF moves left and right (hence the term

"location parameter"). The "true" PDF, the one from which the sample was actually drawn, is plotted as a solid red curve.

Again, the observer's task is – given only the sample and the Gaussian constraint – to estimate the unknown *location* μ of the true PDF. The estimator that the observer uses is a function of the sample and may depend as well on the distributional assumption (Figure 1C). The observer's estimate can be denoted $T_G^O(X_1, ..., X_n)$ or, if there is no ambiguity, just T_G^O. "O" stands for the Observer, "G" for the Gaussian assumption.

While the observer has two sources of information (*distributional constraints* and the *sample*) that inform their estimate $T_G^O(X_1, ..., X_n)$, it is usually the sample values that take center stage, with the Gaussian assumption in the background. But what use – if any – does the observer make of the distributional constraints? How are the parts combined to estimate the location of the whole when the distribution generating the parts is not Gaussian?

In Figure 2 we illustrate the Location Task for three choices of distributional family. Figure 2A corresponds to the Gaussian case that we just discussed (Eq. 1). Figure 2B illustrates a location task based on the Laplacian distribution[1], where the PDF is now

$$\varphi_\mu(x) = \frac{1}{2} e^{-|x-\mu|} \quad . \qquad (2)$$

The PDF is symmetric around the unknown location parameter μ. The observer's task is still to estimate μ. Finally, the Uniform Location Task is based on a uniform distribution (Figure 2C),

[1] The actual PDFs used were always scaled to have the same variance. We present them in simplified form in Eqs. 2 and 3. See Methods.

$$\begin{aligned} \varphi_\mu(x) &= 1 \qquad |x-\mu| < \frac{1}{2} \\ \varphi_\mu(x) &= 0 \qquad |x-\mu| \geq \frac{1}{2} \end{aligned} \quad . \tag{3}$$

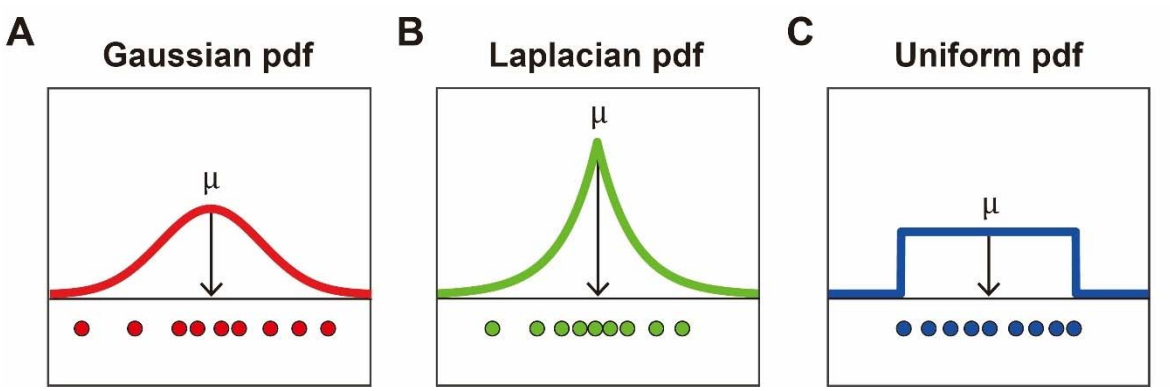


Figure 2. Location Tasks. A. Gaussian Location Task. B. Laplacian Location Task. C. Uniform Location Task. Each plot shows a probability function and a sample, color coded. The PDFs used in the experiment all had the same variance. See Methods and Materials.

We assessed human performance in the three location tasks for the distributions shown in Figure 2. In our experiment, the observer familiarized himself with the distributional families and the typical samples they generate. On every training trial, they saw a sample drawn from a PDF. The color of the points identified the PDF location family from which the PDF is drawn (as in Figure 2).

Uniformly Minimum Variance Unbiased Estimators (UMVUE). An estimator is any function of the sample, $T(X_1, \ldots, X_n)$ and, since an estimator is a function of random variables, it is itself a random variable with its own expected value $E[T]$ and variance $Var(T)$.

A common criterion for a "good" estimator is that it is *unbiased*, i.e. $E[T] = \mu$. A second common criterion is that the estimator should have the lowest possible variance. If an estimator satisfies both of these two criteria for all values of μ then it is a *Uniformly*

Minimum Variance Unbiased Estimator (UMVUE) or, less formally, the *minimum variance estimator.* Cue combination models use UMVUEs and compare these with estimators human observers adopt (Landy et al., 1995; Oruç et al., 2003; Trommershauser et al., 2011).

We can compute a measure of the *efficiency* of any unbiased estimator T for a task as the ratio of the variance of the UMVUE to that of the estimator. Since $Var(T^{UMVUE}) \leq Var(T)$, the *efficiency* of T

$$H = Var(T^{UMVUE})/Var(T) \tag{4}$$

is a value between 0 and 1.

L-Estimators. For convenience, we will focus on a set of estimators referred to as L-estimators (Lehmann, 1998). They are defined as weighted combinations of order statistics

$$T^{w}(X_1, \ldots, X_n) = w_1 X_{(1)} + \ldots + w_n X_{(n)} \tag{5}$$

where the sum of the weights is constrained to be 1. The order statistics (Mood et al., 1974) are just the sample values renumbered from left to right. We can represent any L-estimator by a plot of its weights (Figure 3).

The UMVUEs for the Three Location Tasks. Parametric estimation comprises a collection of mathematical techniques that allow us to identify optimal estimators including UMVUEs for different estimation problems with different distributional families (Lehmann, 1998; Mood et al., 1974). The unbiased estimator T^{+} minimizing variance for the Gaussian task illustrated in Figure 1A is just the arithmetic mean of the sample (Lehmann, 1998; Mood et al., 1974). It is repeated in Eq. 6, written in the form of an L-estimator

$$T^{+} = T^{+}(X_1, \ldots X_n) = \sum_{i=1}^{n} x_i \Big/ n = \sum_{i=1}^{n} \frac{1}{n} x_{(i)} . \qquad (6)$$

This estimator is unbiased $E[T^{+}] = \mu$ and it minimizes the variance $E[(T^{+} - \mu)^2]$.

Not every location task has an *UMVUE*. While the *UMVUE* for the Gaussian is the sample *mean* (Eq. 6), the Laplacian has no closed-form solution for its *UMVUE*. We computed the unbiased L-estimator that minimized variance by numerical methods (exhaustive search) and plot it in Figure 3B, the *UMVUE* among L-estimators denoted $UMVUE_L$. We use the term *UMVUE* to refer to either *UMVUE* or $UMVUE_L$ when it is clear from context which is intended.

The *UMVUE* for the uniform family is

$$T^{u}(X_1, \ldots, X_n) = \frac{X_{(1)} + X_{(n)}}{2} , \qquad (7)$$

the average of the minimum and maximum values in the sample (the *average of extremes*). We have three different *UMVUEs* each matched to the distribution family of a location estimation task (Fig 3). Each is unbiased.

Previous studies on human vision have typically used the Gaussian location family in modeling human performance. These studies have shown that human observers accurately estimate the center of a 2D cloud of points and thus adopt the correct UMVUE for a location estimation of a Gaussian distribution (Gan et al., 2023; Juni et al., 2010; McGowan et al., 1998; Melcher & Kowler, 1999; Rashid & Chubb, 2022; Sun et al., 2016, 2018). However, the centroid is not the minimum variance estimator for the locations of Laplacian and Uniform distributions. It has not been tested how a different location family of a statistical distribution systematically influences human location estimation based on the individual points.

We will investigate how the human visual system adapts its estimation rule $T^O(X_1,...,X_n)$ to location tasks based on each of the distributional families in Figure 2 – if it adapts it at all? Second, we will compare human performance to optimal unbiased estimators minimizing variance (UMVUEs) for each of the three location families in Figure 2.

One possible outcome is that observers may simply adopt the same estimator for all three location tasks, ignoring the distributional information. At the other extreme, the observer might adopt the correct UMVUE estimator for each of the three tasks in Figure 3.

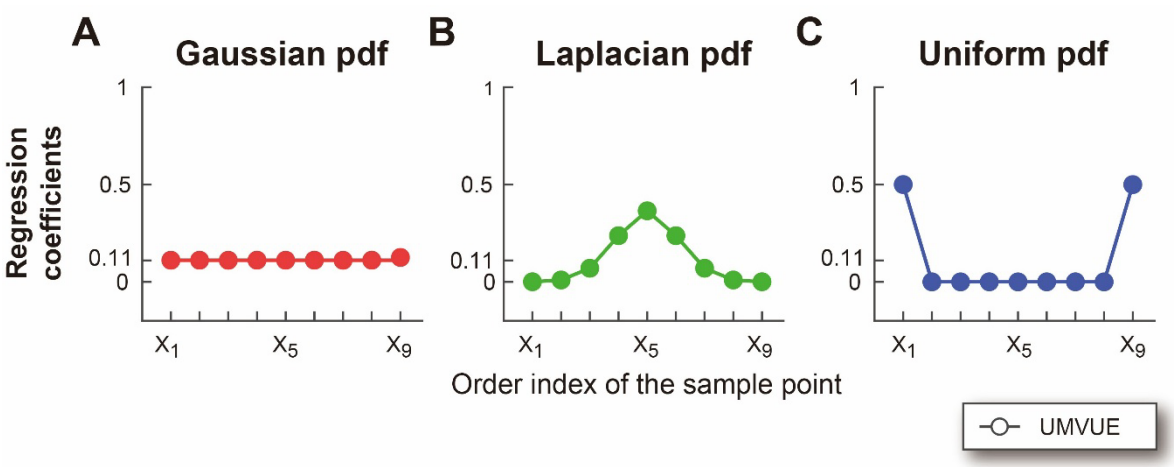


Figure 3: Uniformly Minimum Variance Unbiased Estimators (UMVUEs). A. Gaussian Location Task. B. Laplacian Location Task. The *UMVUE*$_L$ L-estimator estimated numerically (the Laplacian has no *UMVUE*). **C. Uniform Location Task.** Three UMVUE estimators are represented as L-estimators of weighted average of order statistics. The Y-axis denotes weights (i.e., regression coefficients) on each sample point ordered from left to right. UMVUE estimator for the Gaussian location task is the arithmetic mean of the sample. The UMVUE for the Laplacian location task has the largest weight on the median. The UMVUE for the Uniform location task is the average of extremes.

Influence. We can interpret the weights in an L-estimator as measures of the *influence* of each point (each order statistic), a useful characterization of a statistical task (Juni et al.,

2010; Maloney & Landy, 1989; Ota & Maloney, 2024) that is important for research in robust statistical estimation (Hampel et al., 1986; Huber, 1981).

In our experiment, we are presenting an observer with a sample whose order statistics — left to right – are $X_{(1)}, \ldots X_{(i)}, \ldots, X_{(n)}$ and recording the observer's response $T^O(X_{(1)}, \ldots X_{(i)}, \ldots, X_{(n)})$, the location estimate. The influence of the point at position *i* on the location estimate is the partial derivative

$$I_i = \frac{dT^O}{dX_i} \quad . \tag{8}$$

which allows us to measure the contribution of each point to the overall estimate. If the location estimator is an L-estimator, then the influence estimates are just the weights in Eq. (5). We compare the normative influence measures for each task (Figure 3) with the observed influence measures computed by linear regression.

To summarize, in this article we first examine how observers combine locations of individual points to estimate the location of the generating PDF, an exemplary parametric estimation problem in visual form. We then propose a model of human performance, the Visual Cluster Model (VCM) of location estimation. We test the VCM model against alternatives.

Results

The observers were first familiarized with the three distributional families (Gaussian, Laplacian, and Uniform) by viewing the typical samples generated for different sample sizes (Supplementary Figure1A; see Methods and Materials). In the main experiment observers were presented with samples of 9 points drawn from one of the three distributional families and asked to estimate the position of the center of the generating PDF given only 9 sample points and the knowledge of which distribution family had generated the sample. They were given feedback concerning their visual estimate (Supplementary Figure1B). We confirmed that behavioral performance did not systematically change across trials (Supplementary Figure 2AB).

We know the UMVUE weights (influences) for each task (Figure 3) and we can estimate the observers' weights by linear regression (see Methods and Materials). Figure 4 (and Supplementary Figure 3A) illustrates the estimated weights for each participant and the average across participants. Observers were sensitive to the distributional constraints. Observers' influence measure markedly changed with changes in distributional family (interaction between the order of points and the distributional families by ANOVA on the estimated weights: F [16, 304] = 44.53, $p < 0.0001$, partial $\eta^2 = 0.701$).

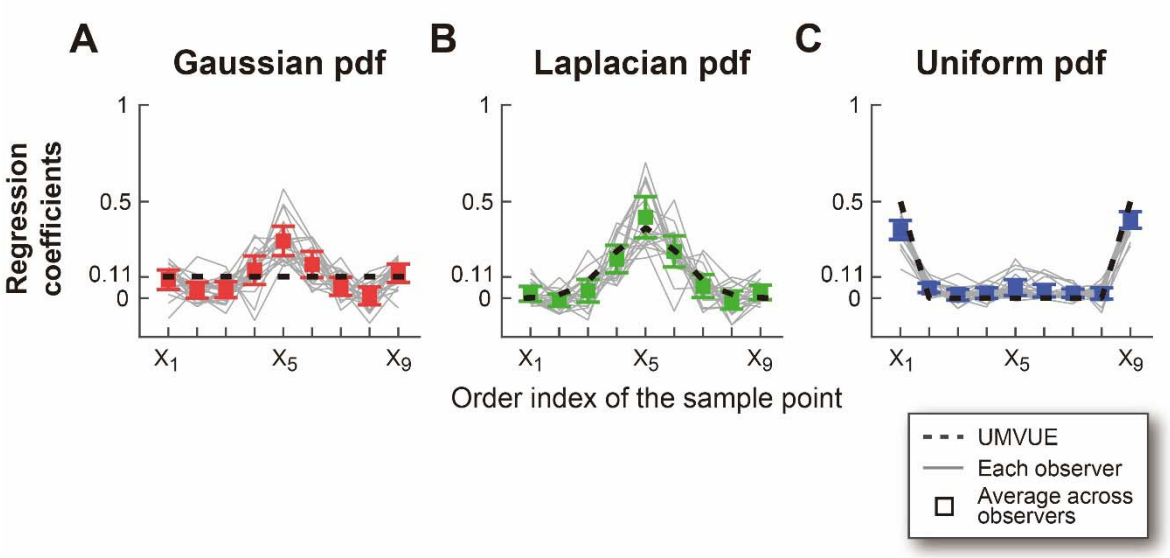


Figure 4: Observers' Estimated Influence Measures Compared to UMVUE Influence Measures. A. Gaussian Location Task. B. Laplacian Location Task. C. Uniform Location Task. Influence (i.e. regression coefficients) of each point in samples (ordered from left to right) on observer's visual estimates. Results for each observer are plotted as faint grey lines. Mean results across observers are plotted as color-coded squares**.** Error bar denotes 99.7% confidence interval (± 3 SEM). Weights employed by UMVUE are taken from Figure 3 and replotted as dashed lines. For the Laplacian distribution, observers' estimated influence measures are not significantly different from the $UMVUE_L$. The observers' influence measures for the other two distributions deviate from those of their respective UMVUEs.

Do observers' choice of estimators approximate the UMVUEs for each location task? We compared measured influence with normative (UMVUE or $UMVUE_L$) influence for each location task separately. For the Gaussian task, the normative estimator is just the arithmetic mean of the sample. In contrast, the observer's estimated influences look w-shaped and are significantly different from the normative influences (F [8, 152] = 19.56, p < 0.0001, partial η^2 = 0.507) by particularly placing larger weights on the median sample point than UMVUE (Supplementary Figure 3B), meaning that observers do something other than averaging points.

For the Uniform task, the measured influences are also different from normative influences (F [8, 152] = 35.27, p < 0.0001, partial η^2 = 0.650). Even though observers gave

most weight to the extreme sample points, they placed significantly less weight on them than the UMVUE (Supplementary Figure 3B). For the Laplacian task, observers approximated the UMVUE (F [8, 152] = 1.83, p = 0.076, partial η^2 = 0.088). We note that there were no significant changes in the measured influences from the beginning to the end of the experiment (Supplementary Figure 2C). The efficiency of the observers' estimators (Eq. 4) was overall 69.6% (Supplementary Figure 4). In other words, the observers were 1.44 times more variable than the UMVUE. There was no significant difference across the three location tasks (the main effect of the distributional families by ANOVA on the efficiency: F [2, 38] = 2.53, p = 0.093, partial η^2 = 0.118).

In summary, the observers are sensitive to the distributional constraints in the three location tasks – choosing significantly different estimators for the three location tasks. The observers' estimators did not differ significantly from the UMVUE_L for the Laplacian location task and roughly approximated the UMVUE for the Uniform location task.

The Visual Cluster Model (VCM)

The observers used different estimators for the three location families. *Here, we considered the possibility that the visual system's choice of estimator is based entirely on the cluster structure of the samples (Sun et al., 2019).* Intuitively, clusters convey information about the PDFs from which they are drawn. A cluster confined to an interval [a,b] will tend to have more points when the probability density across the interval is larger. We first illustrate the computation with an example.

In Figure 5A we plot a 9-point sample and in Figure 5B we partition the sample into K = 5 non-overlapping clusters denoted left to right as $C_1, C_2, \ldots, C_5$ using a standard clustering algorithm described below. Clusters in our experiment have a minimum size of 1 element, and a maximum size of 9 (in which case, there is only one cluster). We assign a location c_j to the jth cluster C_j, the midpoint of the endpoints of the cluster.

In our example, the largest cluster, C_4 contains three points. If we had to base our estimate of location on just one of the clusters in the cluster structure of the sample, it would plausibly be the 4th cluster. But there are competing clusters with either 1 or 2 points, and we can attempt to improve our estimator by weighting together the locations of the clusters with weights that depend on the clusters.

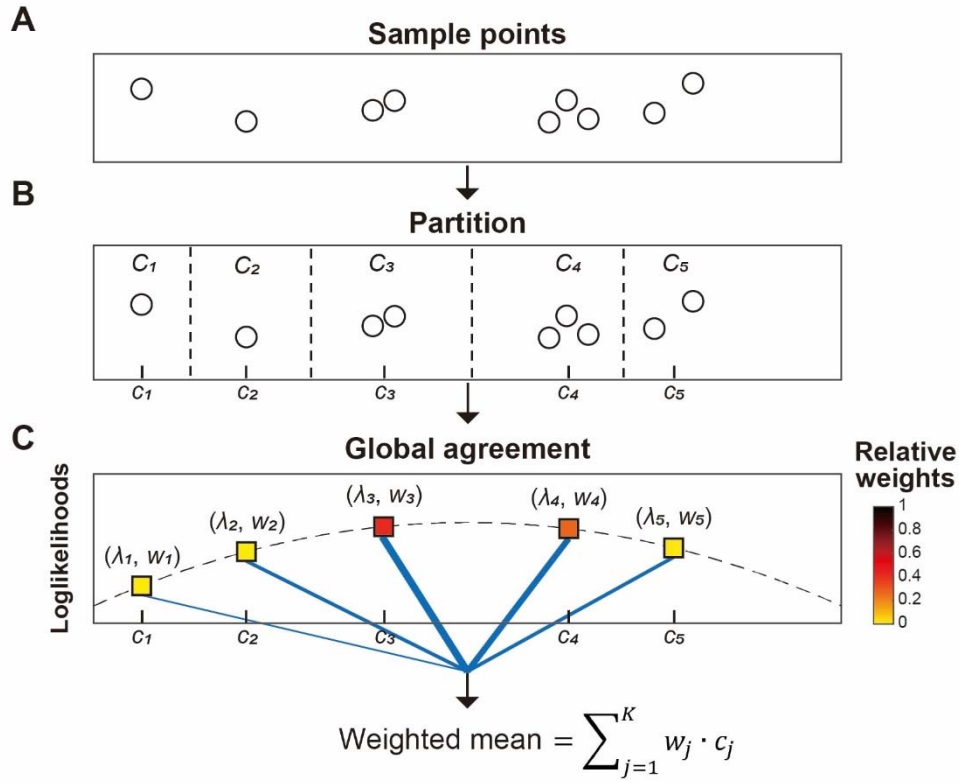


Figure 5: Visual Cluster Model. A. Gaussian Location task. A sample of 9 points is drawn from a Gaussian distribution. Observers were asked to estimate the location of the invisible center of the underlying distribution given only sample points. **B. Partition.** Visual cluster model partition points (partitions shown as dashed vertical lines) into *K* clusters of the sample, $\boldsymbol{C_1,\ldots,C_K}$. Each cluster is assigned a location $\boldsymbol{c_1,\ldots,c_K}$ defined as the average of its largest and smallest points if the cluster contains more than one point. Otherwise, the location of the single point is the location of the cluster. **C. Global Agreement.** Visual cluster model assigns the log likelihood λ_j of the center $\boldsymbol{\mu}$ of the PDF to the location $\boldsymbol{c}_j$ of the jth cluster. A dashed curve represents the log likelihood values across the entire space of the task. The log likelihood values $\lambda_1,\ldots,\lambda_K$ are converted to the relative weights $w_1,\ldots,w_K$ via a softmax transformation. The visual cluster model estimates the position of the center of the PDF as a weighted average of cluster locations $c_1,\ldots,c_K$ with weights $w_1,\ldots,w_K$.

We considered variations on the model at each step but – to simplify the presentation of the model – we first describe one version of the model and discuss

possible variations only after presentation of this base version of the model. One goal of the model is to reduce the complexity of the computation (Lieder & Griffiths, 2019) by relying on the cluster structure of the model (five clusters in our example) (Sun et al., 2019) rather than the nine points.

Step 1: Partition. The visual system partitions points in the sample into a small set of non-overlapping clusters. In Figures 5A and 5B, we apply a standard clustering algorithm, K-means clustering (Hartigan & Wong, 1979), to a sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ from the Gaussian distribution (Eq. 1). The K-means clustering algorithm has a free parameter K (the number of clusters). In Figure 5B, we obtained $K = 5$. The result is a partition into non-overlapping clusters $\boldsymbol{C}_1,\ldots,\boldsymbol{C}_K$ numbered from left to right, possibly varying in number of points. We assign a location to each cluster $\boldsymbol{c}_1,\ldots,\boldsymbol{c}_K$ as the average of its largest and smallest points (average of extremes) if the cluster contains more than one point, otherwise the location of the single point if the cluster contains only one. The first step is analogous to the cluster-of-samples representation developed by Sun et al. (2019). We extend their model by introducing *global agreement*.

Step 2: Global agreement. How well does each cluster C_j at location $\boldsymbol{c}_j$ capture not just the points in its own cluster but rather the entire sample of points $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$? We hypothesize that observers make use of the distributional constraint to assess the likelihood of the location of the distribution generating the points. In Figure 5C, for each cluster, we compute the log likelihood λ_j of the location parameter $\boldsymbol{\mu}$ of the population PDF with respect to the location $\boldsymbol{c}_j$ of the jth cluster:

$$\lambda_j = \sum_{i=1}^{9} \log f_\mu(X_i \mid \mu = c_j) \quad . \tag{8}$$

where f_μ is the likelihood function of the distribution chosen to generate the sample points, with unit variance. This step results in a measure of the likelihood that the location c_j of cluster *j* reflects the center of the distribution, taking into account all the points of the entire sample. Therefore, we refer to it as *global agreement*. A dashed curve in Figure 5C shows the log likelihood computed across the entire space of the task. In the Gaussian location task, finding the location of the maximum log likelihood value will correspond to averaging sample points. Instead of the performing the log likelihood computation over the entire space, we reduced the complexity of computation (Lieder & Griffiths, 2019) by computing the log likelihood value with respect to the locations c_j of the jth clusters.

We use a softmax transformation with one parameter β, i.e., known as inverse temperature (Reverdy & Leonard, 2016) to convert the log likelihood λ_j to weights

$$w_j = \frac{\exp(\beta\lambda_j)}{\sum_{j=1}^{K} \exp(\beta\lambda_j)} \tag{9}$$

The parameter β is a free parameter used to scale measures of global agreement. When β is fixed at 0, the relative weights w_j are all 1/K: all clusters get equal weight. When β increases, clusters having a large log-likelihood value gets relatively larger weights relative to those with smaller log-likelihood values. In the limit as β grows, the weight of the cluster with the largest log-likelihood converges to 1 while the other weights drop to 0.

On any given trial, the visual system's estimate of the position of the center of the distribution is modelled as the weighted mean of the locations of the clusters with softmax weights that depend on the log-likelihoods assigned to clusters (Figure 5C)

$$T^{VCM} = \sum_{j=1}^{K} w_j c_j + b \quad . \tag{10}$$

A bias term b is added to account for the small but constant bias that we found in the participants' visual estimates (on average, 2.07 mm, SD = 1.41 mm), possibly a visuo-motor error.

The weight on each cluster is determined by the value of the measure of global agreement (i.e., how accurately the center of the distribution is approximated) relative to the measures of global agreement in other clusters.

Factorial model comparison. We test two levels of the grouping parameter K. At the first level, we will fix K at 9. At the second level, the grouping parameter K will be treated as a free parameter and estimated separately from the data in each location task (maximum likelihood). Given that samples from the three possible distributions are color-coded, the observer can select different values K_G, K_L, K_U of the grouping parameter K for the three different distributional families.

For global agreement (β) we test three levels in the factorial design. At the first level, we will fix β at 0 so all clusters get equal weights. At the second level, the parameter β will be treated as a free parameter constrained to be the same value across the three distributional families. Last, at the third level, we will fix it at 1. By this constraint, global agreement is applied to the model fit but the β value is not informed from each observer's data.

We built six models by combining two levels of partition and three levels of global agreement and compare the models by a factorial model comparison (van den Berg et al., 2014; Zhang et al., 2020). Figure 6A shows the summed AICc score across participants for the six models. The two-stage visual cluster model (partition and global agreement)

outperformed the other five variants of the model. A group-level Bayesian model selection (Rigoux et al., 2014; Stephan et al., 2009) showed that the protected exceedance probability of the two-stage VCM, i.e. the probability for the model to outperform all the other models was 99.9%. This model was also better than the normative model where UMVUE was fitted to the observer's data. Furthermore, the three models based on partitioning points (K_G, K_L, K_U estimated from the data) outperformed the other three models without partitioning (K fixed at 9; the protected exceedance probability = 99.8%). A model where partitioning was applied while global agreement was partially applied ($\boldsymbol{\beta}$ fixed at 1) outperformed a model where partitioning was applied but global agreement was not ($\boldsymbol{\beta}$ fixed at 0), though the $\boldsymbol{\beta}$ value was not fitted to the data in these two models (the protected exceedance probability = 87.1%). *These results support the claim that both partitioning and global agreement are essential computations that underpin the observers' visual judgments.*

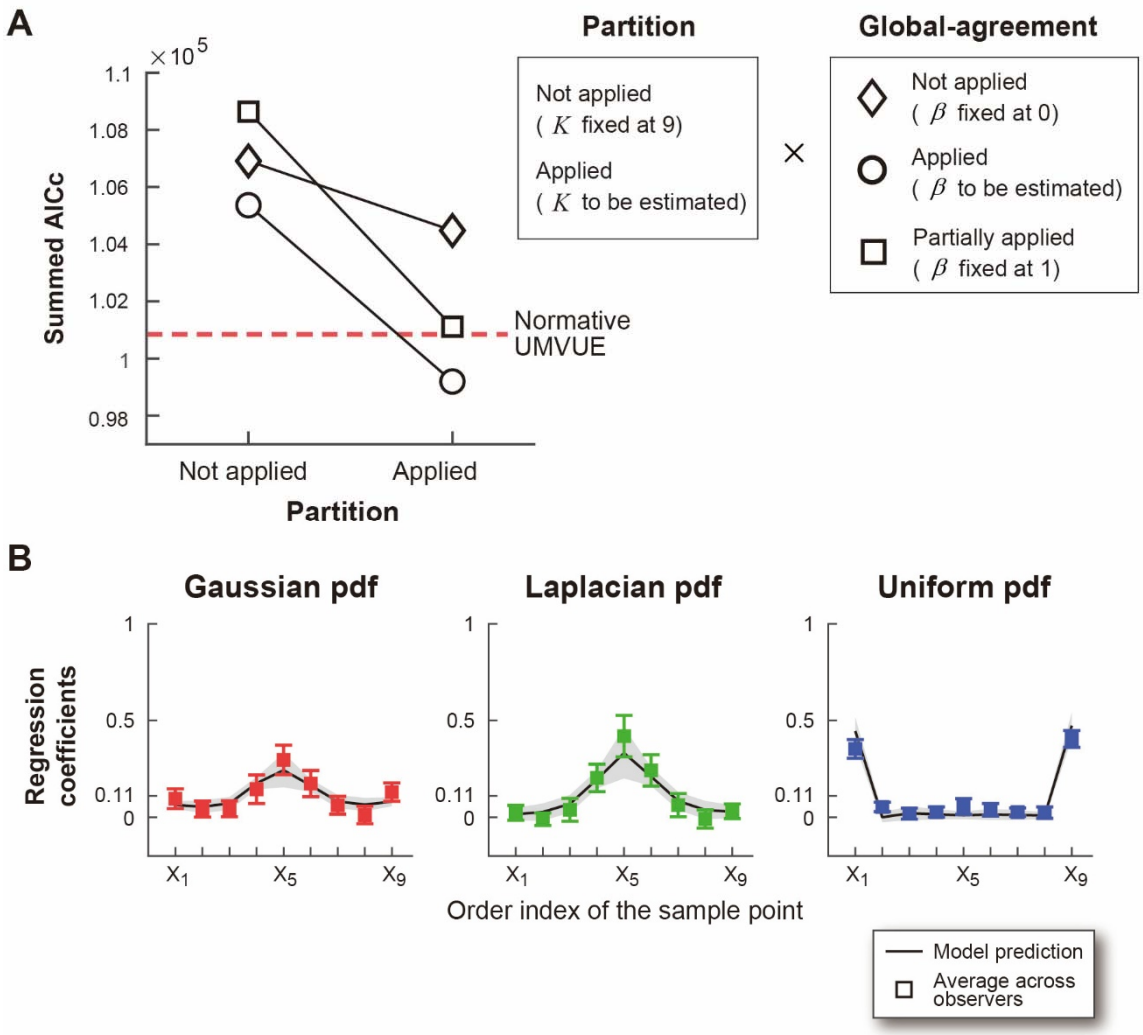


Figure 6: Visual Cluster Model Compared to Human Performance. A. Factorial model comparison. We built six models by combining two levels of partitioning (Step 1) and three levels of global agreement (Step 2). The y-axis denotes the AICc score summed across participants. The lower the AICc, the better the model. A dashed red line plots the AICc score for the normative UMVUE. A grouping parameter K is the cluster size in K-means clustering we used to fit to the data. A parameter $\boldsymbol{\beta}$ is inverse temperature in the softmax transformation which converts the log likelihood values to the relative weights (Figure 5). **B. Prediction of the Estimated Weights (Influence).** Regression weights estimated from the data are taken from Figure 4 and replotted as colored squares. The model prediction of the estimated weights for each location task was estimated using a linear regression on the visual estimates recovered from the best Visual Cluster Model (Partition applied × Global agreement applied). The prediction of the estimated weights was averaged across observers for each task and was plotted as a black solid line**.** Shaded gray area denotes 99.7% confidence interval (± 3 SEM).

We estimated the influence of each point in the sample on the visual estimates generated by the two-stage VCM. The prediction of the influence measure by the model captured human data well in all three distributional families (Figure 6B). Specifically, the model produced a w-shaped influence pattern in the Gaussian task, a median-biased influence in the Laplacian task, and the average of extremes in the Uniform task. It is remarkable that a single model can generate such different estimators each tailored to a distribution family.

Patterns of predicted influence by other variants of the model are summarized in Supplementary Figure 5. When we applied the global agreement partially (β fixed at 1) and had grouping parameters free, the model still produced distinct patterns of measured influence in the three location tasks (Supplementary Figure 5F), suggesting that we are not overfitting the data by including the global agreement parameter. However, without applying global agreement (β fixed at 0), we failed to properly fit the data (Supplementary Figure 5A&D). The models without partition (K fixed at 9) also failed to fit the data (Supplementary Figure 5A-C). These results again demonstrate that both partitioning and global agreement are essential computations.

Finally, we tested whether our results depend on the choice of the clustering algorithm. We repeated the model fit using a hierarchical clustering with single linkage algorithm (Nielsen, 2016; Shetty & Singh, 2021) and found that the results were almost identical (Supplementary Figure 6).

Discussion

A central task of visual perceptual organization is to model how the visual system combines ‘parts’ into ‘wholes’ (Groen et al., 2017; Peirce, 2015; Ullman, 1984; Ullman et al., 2002). When parts have spatial attributes such as location, how does the visual system combine the locations of the parts into an estimate of the location of the whole? We considered a statistical version of this problem modeled on parametric point estimation (Lehmann, 1998). We asked observers to estimate the center of the PDFs of three different location families of one-dimensional PDFs. Observers knew everything about the PDF except for its location parameter which they were asked to estimate. In instructing the observers, we used the neutral term ‘center’ to refer to what was in fact the mean, median, and axis of symmetry of the PDF. The task is a paradigmatic statistical task: given a sample, estimate the parameters of the PDF that generated it. It is also a part of perceptual organization.

We compare human performance to that of a Uniformly Minimum Variance Unbiased Estimator (UMVUE) for that distribution. The UMVUE for each distribution is unbiased and has the lowest possible variance among location estimators. The normative minimum variance estimator is markedly different for the different distributional families we considered, namely the Gaussian, the Laplacian, and the Uniform.

How does the observer carry out the task? For the Gaussian location family, for example, the statistician needs only compute the mean of the sample. The resulting estimate is the UMVUE but, as we have seen this is not what human observers do. We sought an alternative computation that would reproduce what human observers do, and that could be applied to estimation problems for the non-Gaussian distributions as well. *We propose that the visual system bases its estimates on the clusters in the sample and*

integrates the locations of the clusters, weighting them by how well each cluster alone accounts for the observed sample.

Using influence measures, we tested whether observers adopt the same estimator for the three different distributions. We found that observers used different location estimators for the three distributions. We also tested whether the estimator for each distribution was the UMVUE for that distribution. The observers' rules were different from UMVUE for Gaussian and Uniform location tasks but we could not reject the hypothesis that observers used the UMVUE in the Laplacian location task.

In the Gaussian location task, observers exhibit a systematic bias that deviate from equal weighting of each sample point, demonstrating a w-shaped influence pattern – placing large weights on the median and the two extreme points. This result is counterintuitive because it suggests that observers are not simply averaging visual elements, despite averaging being an elementary statistical operation that they are presumably familiar with. Deviations from equal weighting in the Gaussian location task have also been confirmed in previous studies, which show that observer's estimates are biased by outliers in a sample (Juni et al., 2010; Moreland & Boynton, 2017; Rashid & Chubb, 2021). However, the influence patterns were flexibly adapted to the distributional constraints. In the Uniform location task, the two extreme sample points exerted strong influence on the observer's estimates, while their influences were minimal in the Laplacian. Intuitively, the extreme points, if these are drawn from the Laplacian distribution, are poor indicators of the location of Laplacian distribution. If observers were estimating the likelihood of the location of the distribution, the influence of these extreme points would be expected to be small in the Laplacian location task.

It is implausible that the human visual system has different rules of location estimation for any choice of location family. We wondered whether a common estimation algorithm, ideally based on Gestalt structure, could account for how the visual system assigns locations to samples from arbitrary location families.

We developed a two-stage Visual Cluster Model (VCM). Our model comparison analysis validated the two stages on the VCM. The first partition stage edits individual visual points so that points can be grouped into a cluster which itself defines a location. At the second stage (global agreement), the VCM evaluates how likely each cluster approximates the location of the distribution generating the points. A cluster with a high likelihood value is weighted more in the visual judgment than a cluster with a low likelihood value.

The VCM then combines the cluster locations using weights that are proportional to the likelihood values assigned to clusters. The distributional constraints provided information that was necessary to estimate the likelihood of the location parameter of the distribution (Eq. 8). We demonstrated that a two-stage VCM simultaneously accounts for distinct patterns of the observer's visual judgments across all the three location tasks. We note that the VCM is not an L-estimator as the location estimator by the VCM is not the weighted sum of the order statistics of a sample. However, the VCM produces good approximations of observers' estimators and therefore provides a descriptive account for how their location estimations were produced.

Sun et al (2019) also used a clustering algorithm to model the representation of human visual experience. In their task, observers were asked to estimate the location of the empirical mean or mode of the sample of 70 visual points, and points were drawn from a mixture of distributions that had three different peaks. The nature of their stimuli likely

enhanced the visual segmentation of points beneath each peak, making clusters more salient. In contrast, the cluster structure in our task may be visually less salient, as we generated just 9 points from a unimodal (Gaussian, Laplacian, or Uniform) distribution. Nevertheless, models without the partition stage did not explain the data well.

Gestalt psychologists (Koffka, 1935/2013; Wagemans, Elder, et al., 2012; Wagemans, Feldman, et al., 2012) identified Gestalt rules underlying perceptual organization (for review see (Kubovy & Gepshtein, 2003)). Visual objects are grouped together based on a similarity of spatial attributes and these new *Gestalten* are assigned attributes including locations in space. The theory is descriptive and provides no clear justification for why Gestalten such as clusters are part of our visual experience (Jakel et al., 2016). The VCM offers two possible reasons for this.

First, the visual system can reduce the amount of information that must be processed by aggregating parts (Stigler, 2016). This becomes progressively advantageous as the number of visual elements increases. It is not simply the clusters that are useful but the fact that there can be many fewer clusters than sample points. Second, a cluster of points triggers useful information for visual judgments beyond just spatial attributes. In the VCM, each cluster is assigned its own location, which serves as an anchor where the likelihood of the center of the generating PDF is evaluated. By combining these two advantages, the VCM reduces the number of computations required to carry out the task, analogous to the resource rational principle in perception, cognition and decision-making (Lieder & Griffiths, 2019). For instance, in Figure 5C, the likelihood of the location of the PDF underlying the sample points is evaluated only at five cluster locations. Nevertheless, it approximates the maximum likelihood estimation as if we applied a low-pass filter at the limited sampling points. The resulting estimated likelihood

function is not perfectly accurate. To compensate for this, the VCM computes a weighted average of these sampling points based on their estimated likelihoods. In this sense, Gestalten are not just fascinating visual illusions but may also correspond to intermediate steps in optimal computations.

In Marr's framework of three levels of analysis (computational, algorithmic, or implementational) (Marr, 1982), the VCM exists on the computational level, concerning how the visual system assigns a location to a 'whole' composed of 'parts' and what objective is optimized in perceptual organization. The goal of the VCM can be interpreted as balancing task accuracy with minimizing computational costs such as time (Griffiths et al., 2015). The resource rational principle is considered a variant of Bayesian inference, which posits that the visual system aims to identify the most plausible interpretation of the structure of the image given its constituent visual elements (Froyen et al., 2015; Gershman et al., 2016; Shivkumar et al., 2025).

How the VCM is implemented algorithmically or biologically is a separate question. In particular, we implemented the first partition stage using two standard clustering algorithms, a centroid-based K-means clustering (Hartigan & Wong, 1979) and a connectivity-based hierarchical clustering (Nielsen, 2016; Shetty & Singh, 2021). We found that both algorithms reproduced behavioral patterns equally well. It should be noted that recent work has successfully modeled human perceptual experience from a clustering perspective in tasks such as auditory streaming (Larigaldie et al., 2025), motion perception (Gershman et al., 2016; Shivkumar et al., 2025), and the representation of uncertainty (Sun et al., 2019). These studies employed a stochastic clustering approach called the Chinese restaurant process, assuming that perceptual noise leads to variability in cluster

size. Such an approach may produce similar behavioral consequences as the clustering algorithms we used.

In conclusion, we addressed a central question in perceptual organization concerning how the ‘whole’ and its attributes are estimated from the attributes of individual ‘parts’. We formalized a statistical version of this problem by presenting observers with samples of points (‘parts’) drawn from three probability density functions (Gaussian, Laplacian, and Uniform) and asking them to estimate the location of the center of the PDF (‘whole’). We demonstrated that the observer’s visual judgments can be explained by a two-stage computational process: partition and global agreement. These results suggest that the observer’s estimate of the location of the ‘whole’ is based not directly on the local locations of the ‘parts’ but rather on the locations of ‘clusters’ they form. As such our findings imply that to further understand how the human visual system estimates the location of the ‘whole’, it is necessary to consider both the perceptual process which structures the groups based on ‘parts’ and the probabilistic structure (i.e., probability density function) that generates local locations of the ‘parts’.

Methods and Materials

Participants

Twenty participants (mean age 22.9, range 18–29, 13 females, 7 males, 0 other) participated in the experiment. This study was approved by the University Committee on Activities Involving Human Subjects of New York University (IRB-FY2019-3450) and carried out in accordance with its approved guidelines. Informed consent was given by each participant before the experiment. None of the participants was aware of the hypotheses under test in the experiment. Participants knew that they would receive $12 per hour plus a performance-related bonus (average bonus: $16.90, range: $11–$24). The experiment took approximately 2 hours.

Apparatus

Stimuli were displayed on a horizontal monitor (VPIXX, VIEWPIXX, 514 mm × 288 mm). The monitor resolution was 1920 × 1080 pixels with a 60-Hz refresh rate and we refer to location on the monitor by horizontal coordinates ranging from -257 mm to 257 mm and vertical coordinates -144 mm to 144 mm. The participants were seated at a viewing distance of approximately 43 cm and manipulated a computer mouse to carry out tasks. All stimuli were presented and controlled using the Matlab Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Training Task

Participants were first familiarized with the three distributional families: Gaussian, Laplacian, and Uniform (see Supplementary Figure 1A). In this training stage, participants were not yet asked to make any estimates but were asked to observe the population PDFs, the population means, and the samples drawn from these population distributions. On each trial, we plotted a PDF drawn from one of the three distributional families and a

sample from the same PDF. The origin (0,0) was set at the vertical and horizontal center of the screen. The location μ of the PDF was drawn uniformly from the interval -100 mm and 100 mm. Following this, a sample of *N* points was randomly drawn. The sample points were displayed underneath the shape of the PDF. The points were discs with a radius of 1.5 mm and colored by the same color that was used to draw the underlying PDF (for instance, red Gaussian; green Laplacian; blue Uniform). The color codes were counter-balanced across participants.

The population variances σ^2 of all PDFs was fixed at 2500 mm^2 (σ = 50 mm). We added vertical jitter to the vertical positions of the sample points to reduce the chance that sample points would occlude one another. The variance of vertical jitter was fixed at 6.25 mm^2. To aid the observer in interpreting the PDF and sample during training, we marked the location μ of the PDF by a white solid line segment and dashed lines (width: 2 mm, height: 35 mm).

During training, the sample size *N* was initially 300 and was gradually reduced to 50, 20 and finally 9. With 300 points the displayed sample could be readily matched to the PDF from which it was drawn.

Presentation of the sample when N > 9 was extended across time: 18 additional points were drawn every 100 ms from the underlying PDF until the required number of points (300, 50, 20, or 9) points appeared on the screen. The sample points were displayed for 2s. After an inter-trial interval of 1s, a new PDF with randomly chosen location μ was displayed together with a sample drawn from the PDF.

A distribution was chosen from one of the three distributional families (Gaussian, Laplacian, and Uniform) and was repeated for five trials. In each trial, a new sample was generated from the distribution chosen. After the presentation of 300 sample points in

groups of 15 = 5 x 3, the second block of 50 sample points started. The third block and the fourth block used 20 and 9 points, respectively, giving 60 trials in total. The final sample size (*N* = 9) was the one used in the main experimental task, the location task. The training task provided substantial opportunity to observe typical Gaussian, Laplacian, and Uniform samples while learning the color codes.

In the instructions, we used the following analogy:

> On every trial, you will first see a white apple tree located somewhere on the screen. You will then see a curve which tells you where the apples are most likely to fall. The higher the curve is, the more likely an apple will fall below that point on the curve. Where the curve is high, there probably will be more apples. Where the curve is low, fewer. The density curve is always symmetric around the apple tree. The horizontal location of the apple tree is not fixed but it will vary from trial to trial along with the curve and the apples. The tree, the apples and the curve will all move together from left to right on each trial. There are three types of density curves from where apples will fall, each coded by color. One is bell-shaped (red), the second is “pointy” (green), and the last is like a box (blue).

Visual Location Task [Main Experiment]

The stimuli in the visual location task were the same as in the visual training task with the following differences: On each trial in the Visual Location Task, participants saw only a colored-coded sample of N = 9 points randomly drawn from the Gaussian, Laplacian or Uniform distribution. The PDF and the location of the PDF were not shown on the screen. To help participants remember the three density curves, small icons representing the shape of the three distributions were drawn near the top of the screen (see Supplementary Figure 1B and Supplementary Videos 2&3).

Participants were asked to estimate the unknown position of the center (i.e., location μ) of the underlying PDF using only the positions of sample points. They could move a white upward pointing arrow (10 mm) to any location on the screen using a mouse. Participants set the arrow at a position where they believed the horizontal center of the

underlying PDF was located and clicked the mouse button to record their estimates. We ignored the vertical displacement of the arrow in analyzing the observer's settings.

Following a response, we presented a white vertical line (35 mm) marking the location of the underlying PDF as feedback on the error between the participant's estimates and the true position of the center of the PDF. Participants received 500 – D^2 points where D^2 is the squared magnitude of their error in mm. If the position of the estimates perfectly matched the true center ($D^2 = 0$), the participant received 500 points. Points received could be negative if the error in distance was above 22.4 mm or below -22.4 mm. At the end of each trial, the estimation error (mm), the score (points) and the cumulative score were presented as a graphical display for 1.5 s. Participants received a bonus payment proportional to the total score they earned across all trials ($1 per 10,000 points). The estimator that maximized the participants' expected bonus for each distribution family was the UMVUE for that distribution family.

We provided warm-up trials before the main experiment. During warm-up, the sample size was initially 300 and gradually reduced to 50, 20, and 9. The warm-up task became more difficult as the sample size was reduced. As in the training task, a distribution was chosen from one of the three distributional families (Gaussian, Laplacian, and Uniform) and was repeated for five trials. In each trial, a new sample was generated from the distribution chosen, giving 15 = 5 x 3 trials for the first block of 300 sample points. Then the second block of 50 sample points started. The third block and the fourth block used 20 and 9 points, respectively, giving 60 warm-up trials in total. Following a short break, the main experiment started. One block consisted of 10 trials. Samples from the three distributions were interleaved across blocks. Each participant ran 90 blocks (30 blocks for each distribution) with a short break after every 6 blocks.

Uniformly Minimum Variance Unbiased Estimator among L-estimators

The Laplacian location family has no closed-form solution for its UMVUE. Therefore we computed the unbiased L-estimator that minimized variance numerically as follows. We denote the UMVUE among L-estimators as $UMVUE_L$.

Given the n × n covariance matrix Σ of the order statistics $X_1, \ldots, X_N$ we can compute the choice of weights that leads to a minimum variance estimator as

$$w \propto \Sigma^{-1} e \qquad (11)$$

where $e = [1,1,\ldots,1]'$, assuming that the correlations between the random variables are non-zero (for proof, see Appendix A in Oruc et al., 2003), which is our case because the random variables are ordered. We ran the numerical simulation by resampling the order statistics 1,000,000 times and then computing the inverse of the covariance matrix and the weights in Eq. 11. We normalized the weights so they sum to 1.

Measuring influence

Order statistics. We performed a regression analysis to estimate the influence of each point in the sample on the participant's estimate of the position of the horizontal location of the distribution (Gaussian, Laplacian, or Uniform) that had generated the sample. We ordered 9 sample points in each sample in ascending order from the leftmost point to the rightmost point. For each of the three distributions, we regressed the vector of the participant's estimators $\mathbf{T^O}$ which contained 300 data points by the 300 by 9 matrix $\mathbf{X}$. Columns were vectors of order statistics for each trial $\mathbf{X_{(1)},\ldots,X_{(9)}}$. Our linear regression model is thus $\mathbf{T^o = X\widehat{W}}$ where $\widehat{\mathbf{W}}$ is the estimated weight vector. We used an ordinary

least squares regression with the constraint that weights add to 1: $\sum_{i=1}^{9} \widehat{w}_i = 1$. We estimated influence for each participant and each distribution family of the population distribution. In addition, we estimated the weight vector of the visual cluster model by regressing the vector of the visual estimates predicted by the VCM by the matrix $\mathbf{X}$. For each variant of the VCM, we retrieved their visual estimates from the best fit parameters in each model (see *Model fitting and evaluation*). See Supplementary Figure 5 for the estimated weights in each model. The influence measures are a sensitive way to uncover how participants behave in carrying out the task and discover discrepancies between human performance and optimal performance (Dal Martello et al., 2023; Ota et al., 2024; Ota et al., 2025b; Ota et al., 2025a; Ota & Maloney, 2024).

Model fitting and evaluation

For each participant, we fitted the model's visual estimates [Eq. 10] to the participant's estimates by minimizing the negative log-likelihood of the observed estimates using Bayesian adaptive direct search (BADS) (Acerbi & Ma, 2017). We assumed that the observer's location estimate $T^{O} \sim N(T^{VCM}, \sigma^2)$ was Gaussian distributed.

In the first stage of the VCM, we tested two levels of partitioning sample points. At the first level, we did not apply partition and fixed the grouping parameter K (i.e., the number of clusters) in K-means clustering at 9 across all three location tasks. At the second level of applying partition, we estimated K separately from data in each of the three location tasks as a free parameter. Because K must be an integer, we estimated K by a grid search in steps of 1. In each iteration of the grid search, the same integer value of K was used across trials in each task. The estimated value of K could differ across the

three location tasks as the tasks for the three conditions were interleaved. We optimized other parameters (described below) for each combination of the values of the three grouping parameters K_G, K_L, K_U and chose the best combination that minimized negative log-likelihood on each individual's data.

In the second stage of the VCM, we tested three levels of global agreement. At the first level, we did not apply global agreement and fixed the parameter β (the inverse temperature in the softmax decision rule) at 0. At the second level of applying global agreement, we treated β as a free parameter with the boundary $\beta \in [0,4]$. Here we fitted the value of β jointly across the three location tasks. Because the probability density did not convey any information about how the relative weights should be assigned to each cluster, we reasoned that participants did not adjust the global agreement parameter according to the distribution family. Finally, at the third level, we partially applied global agreement by fixing β at 1. Therefore, this factorial design allowed us to compare six variants of the model. In all six models, we included a Gaussian response variance σ^2 and a constant bias parameter b for individual's location estimates with the boundary $b \in [-50,50]$. Again, we fitted the values of σ^2 and b jointly across the three location tasks.

We also used hierarchical clustering with single linkage to test whether the results of the VCM depend on a clustering algorithm we chose. Instead of K, we used $\theta \in [0,150]$ as a grouping parameter. We fitted $\theta_G, \theta_L, \theta_U$ to the data while other parameters remained the same (Supplementary Figure 6). This parameter is a cutoff threshold in millimeter. If the distance between two sample points or clusters is below a threshold, these two are grouped into one cluster.

In the normative model, we fitted UMVUEs matched to the distribution family to the data in the corresponding location task with two free parameters σ^2 and b. In all models, we repeated the search process with different starting points to verify that we had found the global minimum. For model comparison, we applied AICc—Akaike information criterion with a correction for finite sample size—to each participant and each model as the information criterion for goodness-of-fit (Burnham & Anderson, 2002; Hurvich & Tsai, 1989). The Bayesian information criterion (BIC) is motivated as a dimension-consistent model selection criterion designed to select the "true model" when it is contained in the candidate models being considered. Burnham & Anderson (2002) argues that expecting any candidate model to represent the "full truth" is often unrealistic in biological sciences. They therefore recommend AIC-type criteria for the analysis of empirical data, because AIC is not designed to identify the true model but rather to select the model expected to be closest to the data-generating process by minimizing the relative information loss among the candidate models. The AIC's justification is asymptotic and it can be biased toward more complex models when sample size is small relative to the number of model parameters. AICc introduces an additional penalty term that reduces this small-sample bias, helping to avoid potential overfitting and favour more parsimonious models. Based on AICc values across participants and across models, we computed the protected exceedance probabilities (Rigoux et al., 2014; Stephan et al., 2009) using a Matlab function 'spm_BMS' in SPM 12 (https://www.fil.ion.ucl.ac.uk/spm/). The sign of AICc values was reversed by multiplying with $-\frac{1}{2}$ so that an entry to the function 'spm_BMS' can be converted to log model evidence (https://mbb-team.github.io/VBA-toolbox/wiki/BMS-for-group-studies/).

802

Competing Interests

The authors declare no competing interests.

Author's Contributions

KO, QW, PM and LTM designed the experiment and analysis, QW and KO collected data. KO carried out the analyses. All authors contributed to developing and evaluating models and writing the article.

Data and code availability

The data frames and the codes used to generate the figures are shared at GitHub repository: https://github.com/keijiota/Given-the-birds-where-is-the-flock-

Acknowledgments

This research was supported by Grant-Aid for JSPS Fellows No. 17J07822 awarded to KO, by a Fellowship from the Institute for Advanced Studies of Paris and a Guggenheim Fellowship awarded to LTM, and an ANR/NIH grant "Probabilistic models of perceptual grouping and segmentation in natural vision" (ANR-19-NEUC-0003-01) awarded to PM.

References

Acerbi, L., & Ma, W. J. (2017). Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Advances in neural information processing systems*, *30*, 1834-1844.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/https://doi.org/10.1163/156856897X00357

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference* (2nd ed.). Springer.

Chen, Y. J., Sun, Z., & Nishida, S. (2025). Human visual grouping based on within- and cross-area temporal correlations. *PLoS Comput Biol*, *21*(9), e1013001. https://doi.org/10.1371/journal.pcbi.1013001

D'Angelo, G., Voto, S., Iacono, M., Glover, A., Niebur, E., & Bartolozzi, C. (2025). Event-driven figure-ground organisation model for the humanoid robot iCub. *Nat Commun*, *16*(1), 1874. https://doi.org/10.1038/s41467-025-56904-9

Dal Martello, M. F., Ota, K., Pietralla, D. E., & Maloney, L. T. (2023). Detecting visual texture patterns in binary sequences through pattern features. *J Vis*, *23*(13), 1. https://doi.org/10.1167/jov.23.13.1

Froyen, V., Feldman, J., & Singh, M. (2015). Bayesian hierarchical grouping: Perceptual grouping as mixture estimation. *Psychol Rev*, *122*(4), 575-597. https://doi.org/10.1037/a0039540

Gan, L., Sun, P., & Sperling, G. (2023). Deriving the number of salience maps an observer has from the number and quality of concurrent centroid judgments. *Proc Natl Acad Sci U S A*, *120*(21), e2301707120. https://doi.org/10.1073/pnas.2301707120

Gershman, S. J., Tenenbaum, J. B., & Jakel, F. (2016). Discovering hierarchical motion structure. *Vision Res*, *126*, 232-241. https://doi.org/10.1016/j.visres.2015.03.004

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Top Cogn Sci*, *7*(2), 217-229. https://doi.org/10.1111/tops.12142

Groen, II, Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philos Trans R Soc Lond B Biol Sci*, *372*(1714). https://doi.org/10.1098/rstb.2016.0102

Hampel, F. R., Ronchetti, E. M., & Rousseeuw, P. J. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, *28*(1). https://doi.org/10.2307/2346830

Huber, P. J. (1981). *Robust_statistics*. Wiley.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297-307. https://doi.org/10.1093/biomet/76.2.297

Jakel, F., Singh, M., Wichmann, F. A., & Herzog, M. H. (2016). An overview of quantitative approaches in Gestalt perception. *Vision Res*, *126*, 3-8. https://doi.org/10.1016/j.visres.2016.06.004

Jeurissen, D., van Ham, A. F., Gilhuis, A., Papale, P., Roelfsema, P. R., & Self, M. W. (2024). Border-ownership tuning determines the connectivity between V4 and V1 in the macaque visual system. *Nat Commun*, *15*(1), 9115. https://doi.org/10.1038/s41467-024-53256-8

Juni, M. Z., Singh, M., & Maloney, L. T. (2010). Robust visual estimation as source separation. *J Vis*, *10*(14), 2. https://doi.org/10.1167/10.14.2

Koffka, K. (1935/2013). *Principles of Gestalt psychology*. Routledge.

Kubovy, M., & Gepshtein, S. (2003). Perceptual grouping in space and in space-time: An exercise in phenomenological psychophysics. In R. Kimchi, M. Behrmann, & C. R. Olson (Eds.),

Perceptual organization in vision: Behavioral and neural perspectives (pp. 57-98). Psychology Press.

Landy, M. S., Goutcher, R., Trommershauser, J., & Mamassian, P. (2007). Visual estimation under risk. *J Vis*, *7*(6), 4. https://doi.org/10.1167/7.6.4

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Res*, *35*(3), 389-412. https://doi.org/10.1016/0042-6989(94)00176-m

Larigaldie, N., Yates, T., & Beierholm, U. R. (2025). Perceptual clustering in auditory streaming. *PLoS Comput Biol*, *21*(7), e1013189. https://doi.org/10.1371/journal.pcbi.1013189

Lehmann, E. L. (1998). *Theory of point estimation*. Springer.

Lieder, F., & Griffiths, T. L. (2019). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behav Brain Sci*, *43*, e1. https://doi.org/10.1017/S0140525X1900061X

Maloney, L. T., & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. *Proceedings of the SPIE: Visual Communications and Image Processing IV*(1199), 1154-1163.

Maloney, L. T., & Thomas, E. A. C. (1991). Distributional assumptions and observed conservatism in the theory of signal detectability. *Journal of Mathematical Psychology*, *35*(4), 443-470. https://doi.org/10.1016/0022-2496(91)90043-s

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

McGowan, J. W., Kowler, E., Sharma, A., & Chubb, C. (1998). Saccadic localization of random dot targets. *Vision Res*, *38*(6), 895-909. https://doi.org/10.1016/s0042-6989(97)00232-0

Melcher, D., & Kowler, E. (1999). Shapes, surfaces and saccades. *Vision Res*, *39*(17), 2929-2946. https://doi.org/10.1016/s0042-6989(99)00029-2

Mood, A. M., Graybill, F. A., & C., B. D. (1974). *Introduction to the theory of statistics* (3rd ed.). McGraw-Hill.

Moreland, J. C., & Boynton, G. M. (2017). A neurophysiological explanation for biases in visual localization. *Atten Percept Psychophys*, *79*(2), 553-562. https://doi.org/10.3758/s13414-016-1251-z

Nielsen, F. (2016). Hierarchical Clustering. In *Introduction to HPC with MPI for Data Science* (pp. 195-211). https://doi.org/10.1007/978-3-319-21903-5_8

Nielsen, K. J., & Connor, C. E. (2024). How Shape Perception Works, in Two Dimensions and Three Dimensions. *Annu Rev Vis Sci*, *10*(1), 47-68. https://doi.org/10.1146/annurev-vision-112823-031607

Oruç, İ., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*(23), 2451-2468. https://doi.org/10.1016/s0042-6989(03)00435-8

Ota, K., Charles, L., & Haggard, P. (2024). Autonomous behaviour and the limits of human volition. *Cognition*, *244*, 105684. https://doi.org/10.1016/j.cognition.2023.105684

Ota, K., Christofilea, E., Charles, L., Daunizeau, J., & Haggard, P. (2025b). Freedom through understanding: instructed knowledge shapes voluntary action choices. *Preprint at PsyArXiv*. https://doi.org/https://doi.org/10.31234/osf.io/mgksv_v2

Ota, K., Ciston, A., Haggard, P., Gajdos, T., & Charles, L. (2025a). Integrating explicit reliability for optimal choices: effect of trustworthiness on decisions and metadecisions. *Preprint at bioRxiv*. https://doi.org/https://doi.org/10.1101/2025.01.12.632598

Ota, K., & Maloney, L. T. (2024). Dissecting Bayes: Using influence measures to test normative use of probability density information derived from a sample. *PLoS Comput Biol*, *20*(5), e1011999. https://doi.org/10.1371/journal.pcbi.1011999

Ota, K., Shinya, M., Maloney, L. T., & Kudo, K. (2019). Sub-optimality in motor planning is not improved by explicit observation of motor uncertainty. *Sci Rep*, *9*(1), 14850. https://doi.org/10.1038/s41598-019-50901-x

Papale, P., Zuiderbaan, W., Teeuwen, R. R. M., Gilhuis, A., Self, M. W., Roelfsema, P. R., & Dumoulin, S. O. (2024). V1 neurons are tuned to perceptual borders in natural scenes. *Proc Natl Acad Sci U S A*, *121*(46), e2221623121. https://doi.org/10.1073/pnas.2221623121

Peirce, J. W. (2015). Understanding mid-level representations in visual processing. *J Vis*, *15*(7), 5. https://doi.org/10.1167/15.7.5

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437-442. https://doi.org/10.1163/156856897x00366

Rashid, J. A., & Chubb, C. (2021). The density effect in centroid estimation is blind to contrast polarity. *Vision Res*, *186*, 41-51. https://doi.org/10.1016/j.visres.2021.04.005

Rashid, J. A., & Chubb, C. (2022). On the flexibility of strategies for center estimation. *Vision Res*, *201*, 108125. https://doi.org/10.1016/j.visres.2022.108125

Reverdy, P., & Leonard, N. E. (2016). Parameter Estimation in Softmax Decision-Making Models With Linear Objective Functions. *IEEE Transactions on Automation Science and Engineering*, *13*(1), 54-67. https://doi.org/10.1109/tase.2015.2499244

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies - revisited. *Neuroimage*, *84*, 971-985. https://doi.org/10.1016/j.neuroimage.2013.08.065

Shetty, P., & Singh, S. (2021). Hierarchical Clustering: A Survey. *International Journal of Applied Research*, *7*(4), 178-181. https://doi.org/10.22271/allresearch.2021.v7.i4c.8484

Shivkumar, S., DeAngelis, G. C., & Haefner, R. M. (2025). Hierarchical motion perception as causal inference. *Nat Commun*, *16*(1), 3868. https://doi.org/10.1038/s41467-025-58797-0

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage*, *46*(4), 1004-1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Sun, J., Li, J., & Zhang, H. (2019). Human representation of multimodal distributions as clusters of samples. *PLoS Comput Biol*, *15*(5), e1007047. https://doi.org/10.1371/journal.pcbi.1007047

Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016). The centroid paradigm: Quantifying feature-based attention in terms of attention filters. *Atten Percept Psychophys*, *78*(2), 474-515. https://doi.org/10.3758/s13414-015-0978-2

Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2018). High-capacity preconscious processing in concurrent groupings of colored dots. *Proc Natl Acad Sci U S A*, *115*(52), E12153-E12162. https://doi.org/10.1073/pnas.1814657115

Tanae, M., Ota, K., & Takiyama, K. (2021). Competition Rather Than Observation and Cooperation Facilitates Optimal Motor Planning. *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.637225

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining priors and noisy visual cues in a rapid pointing task. *J Neurosci*, *26*(40), 10154-10163. https://doi.org/10.1523/JNEUROSCI.2779-06.2006

Trommershauser, J., Kording, K., & Landy, M. S. (2011). *Sensory cue integration*. Oxford University Press.

Trommershauser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends Cogn Sci*, *12*(8), 291-297. https://doi.org/10.1016/j.tics.2008.04.010

Ullman, S. (1984). Visual routines. *Cognition*, *18*(1-3), 97-159. https://doi.org/10.1016/0010-0277(84)90023-4

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat Neurosci*, *5*(7), 682-687. https://doi.org/10.1038/nn870

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychol Rev*, *121*(1), 124-149. https://doi.org/10.1037/a0035234

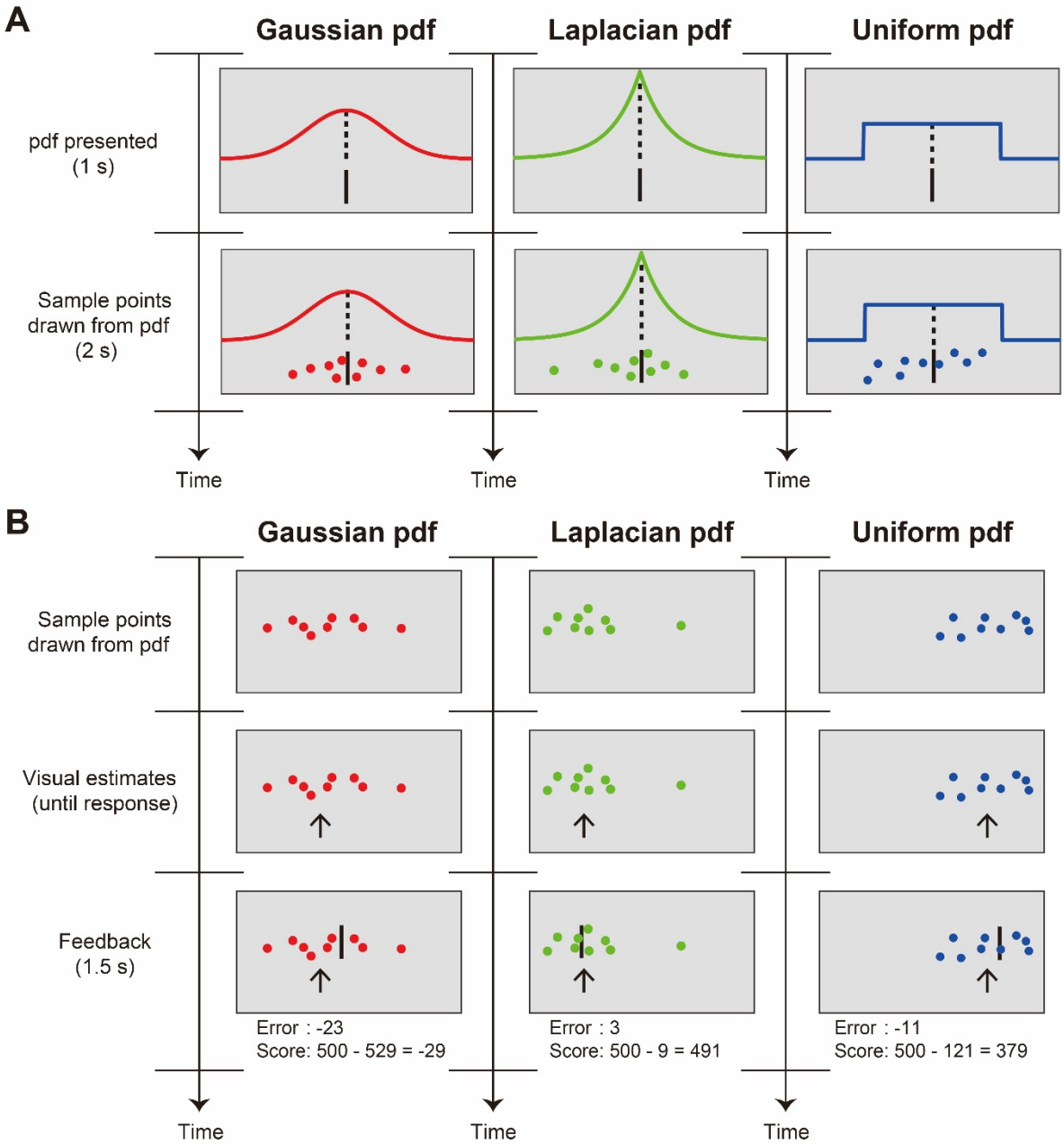
Vilares, I., Howard, J. D., Fernandes, H. L., Gottfried, J. A., & Kording, K. P. (2012). Differential representations of prior and likelihood uncertainty in the human brain. *Curr Biol*, *22*(18), 1641-1648. https://doi.org/10.1016/j.cub.2012.07.010

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychol Bull*, *138*(6), 1172-1217. https://doi.org/10.1037/a0029333

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., van der Helm, P. A., & van Leeuwen, C. (2012). A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychol Bull*, *138*(6), 1218-1252. https://doi.org/10.1037/a0029334

Zhang, H., Ren, X., & Maloney, L. T. (2020). The bounded rationality of probability distortion. *Proc Natl Acad Sci U S A*, *117*(36), 22024-22034. https://doi.org/10.1073/pnas.1922401117

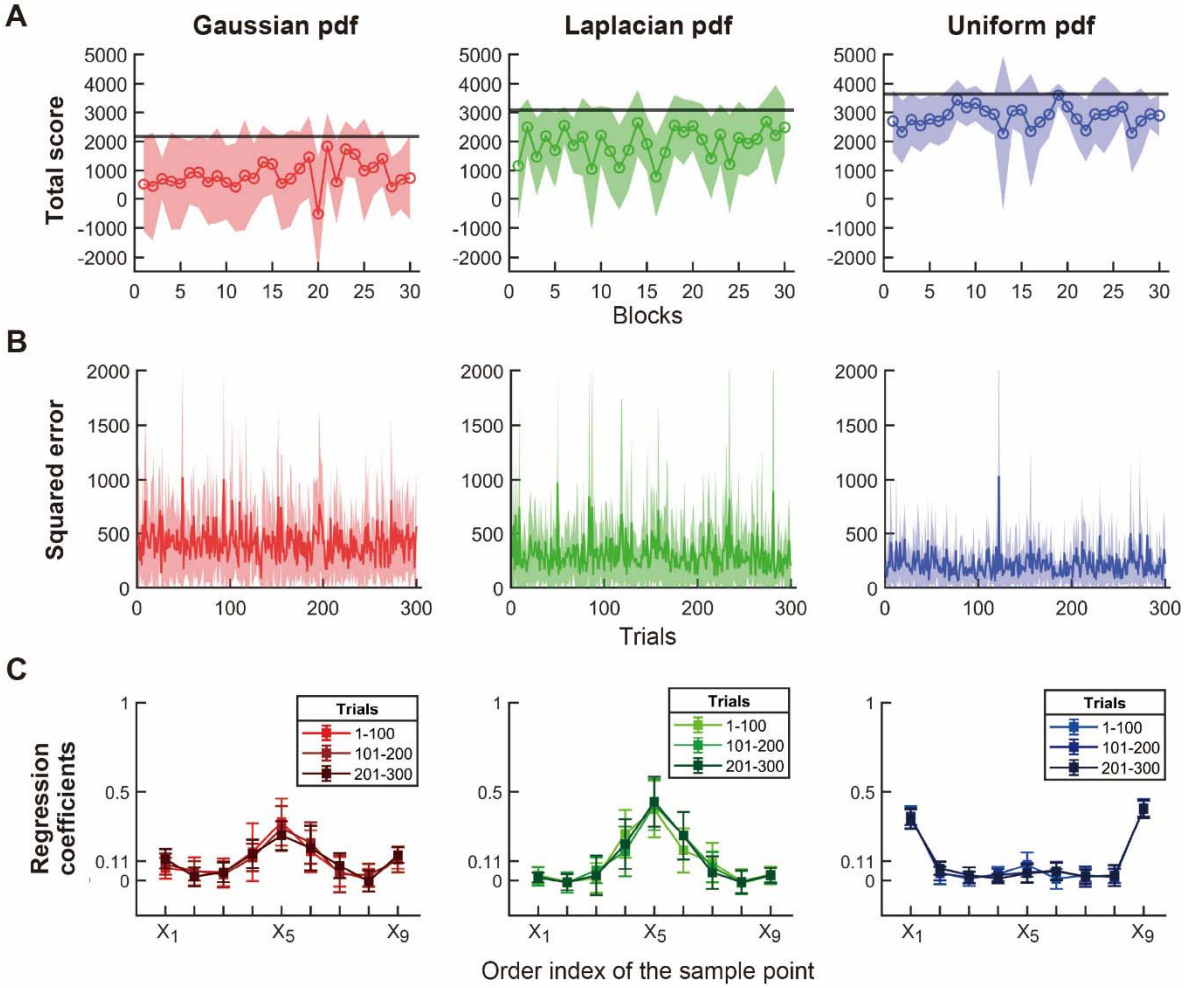
Supplementary Information



Supplementary Figure 1. Training Task and Visual Location Task.

A. Training Task. Three distributional families were color-coded. On each trial, we presented one of the three PDFs with varying its location μ along the horizontal axis and a sample drawn from the same PDF. The sample size reduced from N = 300 to N = 9 between blocks. Participants were asked to observe a color-coded sample drawn from the PDF with the same color code. **B. Visual Location Task**. On each trial, a sample of 9

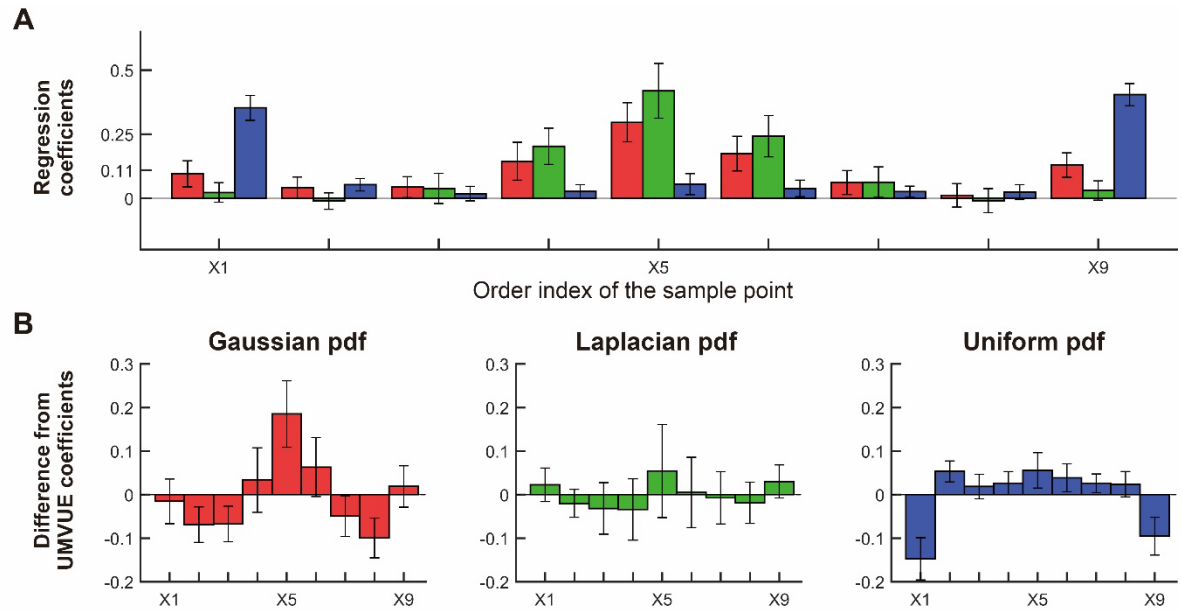
points is drawn from one of the three distributional families with its location μ varied along the horizontal axis. Participants were asked to estimate the location of the invisible center of the PDF that had generated the sample by moving an arrow horizontally. Following a response, the true location of the center of the PDF was presented as a vertical line. A squared error and a score for a trial were also provided as feedback.



Supplementary Figure 2. Behavioral Performance did not Change Systematically Across Trials.

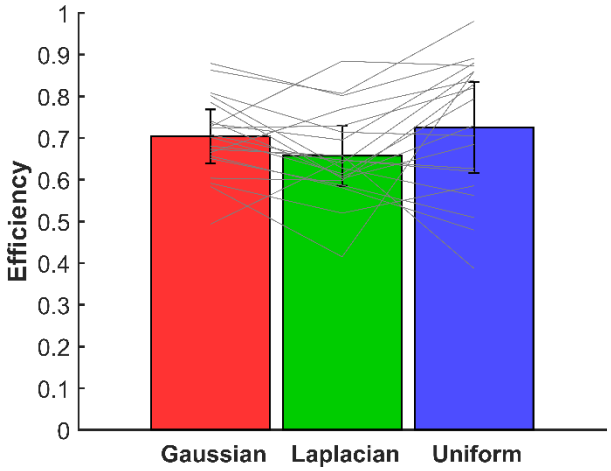
A. Total Score. The participant's total score in each location task is plotted across blocks. Data are averaged across all participants. The shaded areas indicate 99.7% confidence interval (± 3 SEM). Black horizontal lines are the theoretically maximum total score obtained by UMVUE that corresponds to each distribution family. Two-way within participant ANOVA, using the number of blocks (30) and the distribution family (3) as independent variables, shows neither main effect of the block (F [299, 5681] = 1.11, p = 0.095, partial η^2 = 0.055) nor a significant interaction between the block and the distribution (F [598, 11362] = 1.08, p = 0.091, partial η^2 = 0.054). **B. Squared Error.** The squared error between the participant's visual estimate and the true location of the center of the distribution is plotted versus trial. Two-way within-participant ANOVA, using the number of trials (300) and the distribution family (3) as independent variables, shows neither main effect of the block (F [29, 551] = 1.43, p = 0.069, partial η^2 = 0.070) nor a significant interaction between the block and the distribution (F [58, 1102] = 1.19, p =

0.155, partial η^2 = 0.059). **C. Estimated Weights (Influence).** To additionally confirm a lack of learning effect over trials, we computed influence measures separately for the first, the second, and the last 100 trials using a linear regression. The estimated weights are averaged across all participants. The error bars indicate 99.7% confidence interval (± 3 SEM). A two-way within participant ANOVA, using the order of points (9) and the bin of trials (3) as independent variables, shows no significant interaction between the order index of the sample point and the trial bin in the Gaussian location task (F [16, 304] = 0.70, p = 0.796, partial η^2 = 0.035), in the Laplacian location task (F [16, 304] = 0.74, p = 0.751, partial η^2 = 0.038), nor in the Uniform location task (F [16, 304] = 1.22, p = 0.250, partial η^2 = 0.060).



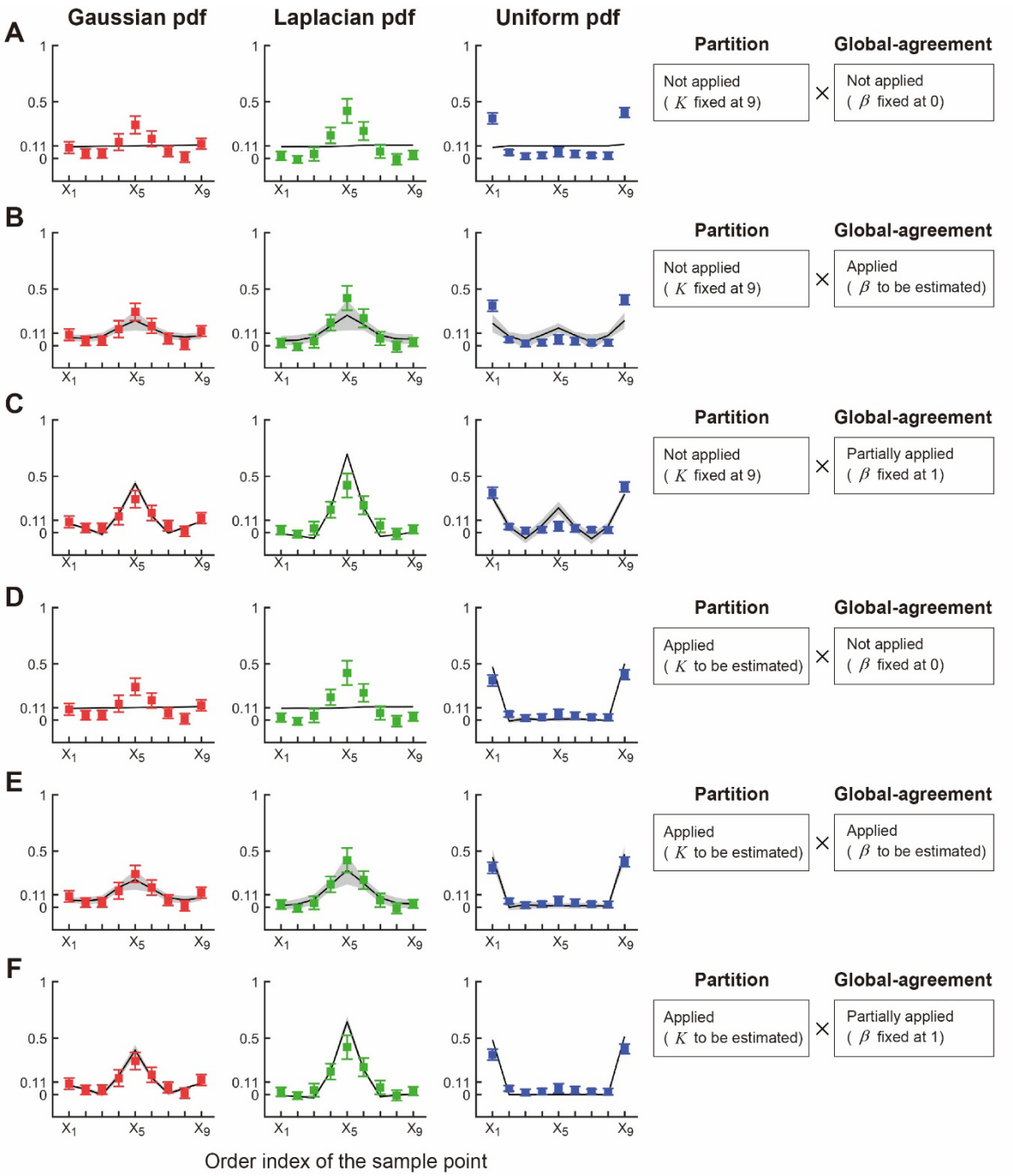
Supplementary Figure 3. Pairwise Comparisons of the Estimated Weights.

A. The values of the estimated weights (influence measure) from the data are taken from Figure 4 and replotted as a bar graph as a function of the order statistics. **B.** Difference between the estimated weights and the weights employed by the UMVUE (for the Laplacian the $UMVUE_L$). In all panels, bars represent the average across participants. The error bars indicate 99.7% confidence interval (± 3 SEM).



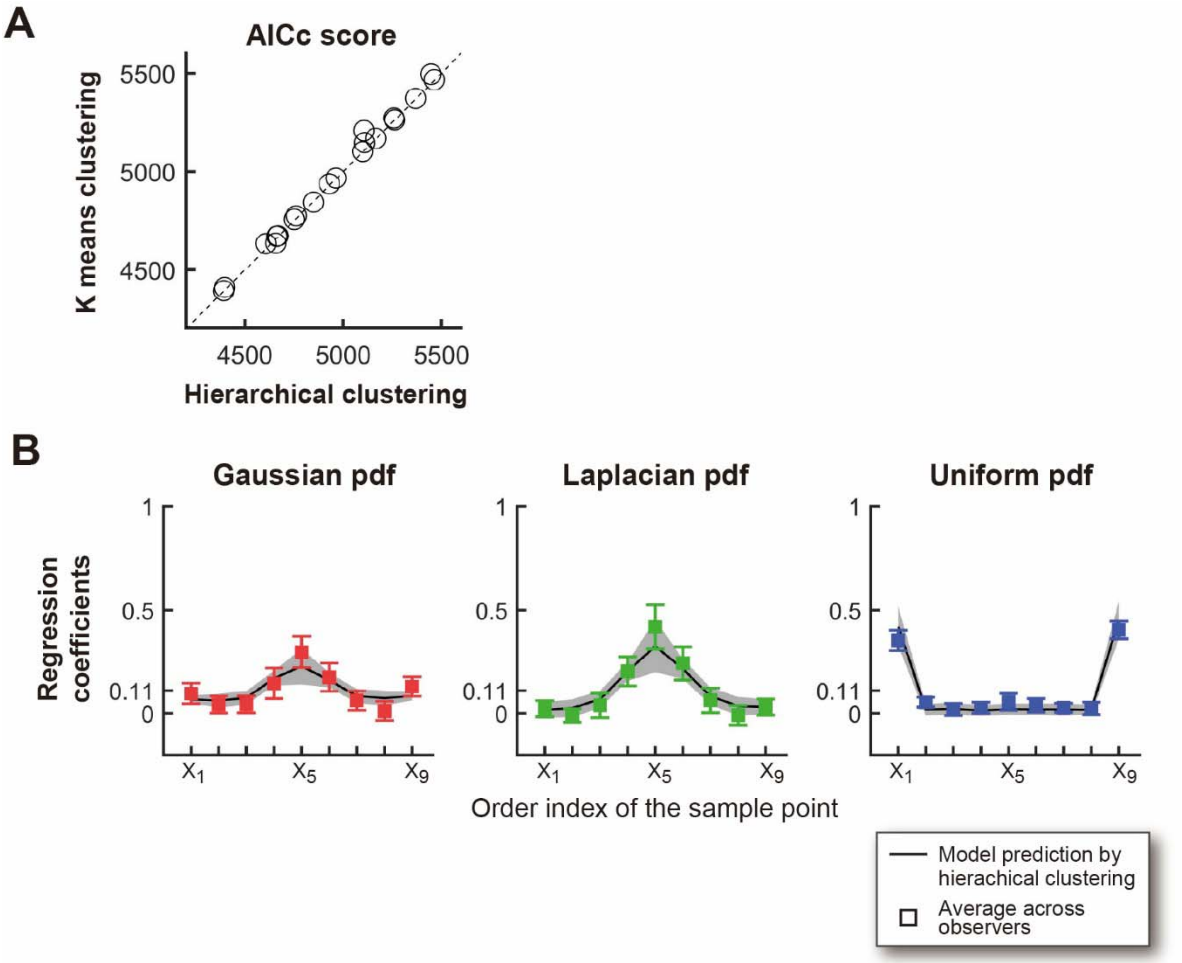
Supplementary Figure 4. Efficiency of the obersers' estimators.

For each of the three location tasks and for each observer, we computed the variance of observer's estimator across trials and the variance of the UMVUE (which matched each location family) across trials. We then calculated the efficiency based on Equation 4. Bars represent the average across participants, whereas gray lines represent individual participants. The error bars indicate 99.7% confidence interval (± 3 SEM).



Supplementary Figure 5. All Variants of the Visual Cluster Model.

The predictions of the estimated weights for each model in the factorial model comparison. Regression weights estimated from the data are taken from Figure 4 and replotted as colored squares. Mean model fits across observers are plotted as black solid lines while shaded gray areas denote 99.7% confidence interval (± 3 SEM).



Supplementary Figure 6. Comparison of the Model Fit across Two Clustering Algorithms.

A. A Comparison of AICc Scores. We adopted K-means clustering in the main text to fit the visual cluster model to the data. We repeated the clustering algorithm analysis using hierarchical clustering with single linkage. **B.** Prediction of the Estimated Weights (Influence). Regression weights estimated from the predictions of the VCM using hierarchical clustering. AICc scores and the estimated weights are almost identical between the two clustering algorithms (see Figure 6).