Empirical test of ideal observer models of individual and ensemble spatial perception

Yanina E. Tena Garcia, Bianca R. Baltaretu, Katja Fiehler
{yanina.tena-garcia,bianca.baltaretu,katja.fiehler}@psychol.uni-giessen.de
Justus-Liebig-Universität Gießen
Dominik M. Endres, dominik.endres@uni-marburg.de
Marburg University

Abstract

Individual and ensemble perception are crucial for interacting with objects in our environment. Individual perception processes single objects, while ensemble perception extracts summary information from object groups (Melcher et al., 2021; Neumann et al., 2018). To investigate how these two modes of perception work with different set sizes (3, 6, 10) in naturalistic settings, we compare two bayesian models on our data. The first model, a variant of the summation model, is the 'Individual Encoding Model'. The second model is the 'Ensemble Encoding Model', which is related to the automatic averaging model. We conducted an experiment in which participants encoded the position of an individual object or an ensemble position that summarized multiple objects in a 3D rendered scene and indicated its remembered position by mouse click on the screen. The 'Individual Encoding Model' assumes that each object's position is encoded in memory, the ensemble position is only evaluated on demand. In the 'Ensemble Encoding Model', the ensemble position is part of the process that generates the scene and is inferred from the observable object locations. We found that the accuracy of reproducing individual object positions increased as set size increased, while the estimation of the ensemble position (arithmetic mean) only differed between the 6- and 10-object set size conditions, with smaller deviations observed for scenes with 6 objects. The 'Ensemble Encoding Model' specifically explains the variability in human behavioral data better. The subject-specific bayes factors in its favor increase with set size. We conclude that in naturalistic scenes the choice between individual versus ensemble encoding is likely driven by the more compact scene representation of the ensemble model.

Keywords: ensemble perception; spatial perception; scene perception; cognitive model; bayesian model

All scripts used in this study and more detailed descriptions of the behavioral experiment, the modeling and their results are available here: `https://doi.org/10.60834/tam-datahub-10.2`.

Behavioral Experiment

To assess participants ability to locate individual and ensemble information, a computer-based experiment was conducted with 29 participants. Participants who viewed scenes with 3, 6 or 10 objects were instructed to recall either the location of one object (individual task) or the average position around which all objects were arranged (ensemble task). They responded by clicking the respective target position as accurately as possible. Tasks were blocked, with each participant completing 216 trials. Locating accuracy was measured as the distance the clicked and the actual (veridical) 2D object position. Individual positions were defined relative to the center of mass of an object, and ensemble positions by the centroid of all objects.

Models

We created and explored two bayesian perception models for the behavioral data, which differ in their assumptions about how individual and ensemble percepts are computed. We incorporated zero-mean visual uncertainty with a standard deviation of 2° in both models, (Pertzov et al., 2015).

Individual Encoding Model (IEM)

In the IEM the ensemble location $\vec{E}$ is computed by averaging individual object representations $\vec{X}_i$ (Harrison et al., 2021; Robinson & Brady, 2023). For K objects in a scene which are presented at locations $\vec{O}_i$, that differ from the internal representations $\vec{X}_i$ by independently drawn visual noise/uncertainty $\vec{V}_i$, the model assumptions are

$$\vec{X}_i \sim \mathcal{N}(\vec{0},\Sigma_X)\,;\vec{V}_i \sim \mathcal{N}(\vec{0},4\cdot\mathbb{I}_2)\,;\Sigma_X \sim \mathcal{W}(4,12/\sqrt{4}\cdot\mathbb{I}_2)$$

$$\vec{O}_i = \vec{X}_i + \vec{V}_i\,;\vec{E} = \frac{1}{K}\sum_{i=1}^{K}\vec{X}_i \qquad (1)$$

where $\mathcal{N}$ refers to a multivariate normal distribution and $\mathbb{I}_2$ is the 2×2 identity matrix. $\mathcal{W}(4,12/\sqrt{4}\cdot\mathbb{I}_2)$ is a wide Wishart prior on the covariance matrix Σ_X with an expectation of 144 for the diagonal elements, which reflects the experimental design: a standard deviation of 12° was used for object placement around the screen center, across all test scenes used in the behavioral experiment.

Ensemble Encoding Model (EEM)

The EEM describes a generative process for scenes that independently draws the ensemble position $\vec{E}$ and the object positions $\vec{X}_i$ relative to $\vec{E}$ (Lew & Vul, 2015). Visual uncertainty is the same as for the IEM (Pertzov et al., 2015). The model is therefore specified as

$$\vec{E} \sim \mathcal{N}(\vec{0},\Sigma_E)\,;\vec{X}_i \sim \mathcal{N}(\vec{0},\Sigma_X)\,;\vec{V}_i \sim \mathcal{N}(\vec{0},4\cdot\mathbb{I}_2)$$

$$\vec{O}_i = \vec{E} + \vec{X}_i + \vec{V}_i\,; \Sigma_X \sim \mathcal{W}(4, 9/\sqrt{4}\cdot\mathbb{I}_2)\,; \Sigma_E \sim \mathcal{W}(4, 9/\sqrt{4}\cdot\mathbb{I}_2)$$

The prior diagonal expected covariances of 81 reflect the experimental design here, too. Both the standard deviation of the average object position from the screen center, as well as the average deviation of the individual object from the average position was 9°.

In both models, the ensemble percept is given by the posterior distribution of (the latent) $\vec{E}$ after encoding, i.e. we evaluate $P(\vec{E}|\vec{O}_{1,\ldots,K}, \Sigma_X)$ for the IEM, and $P(\vec{E}|\vec{O}_{1,\ldots,K}, \Sigma_X, \Sigma_E)$ for the EEM. Both posteriors can be computed analytically, details can be found in the model scripts, see DOI above.

Model Fitting

We fit the models to simulated data (for recovery tests) and real data (for model evaluation) by maximizing the posterior probability of the model's predictions and parameters with respect to Σ_X (and Σ_E). We used a Laplace approximation to the model evidence (Bishop, 2006) for model comparison.

Recovery of individual object positions and ensemble positions were within the covariance limits prediced by the models. Covariance parameters were recovered with a regression coefficient of $r \in [0.74, 0.93]$. Model type recovery (EEM vs. IEM) yielded log bayes factors > 10 in favor of the generating model.

Results

Behavioral Experiment

For both the individual and ensemble reproduction task, there was a significant main effect of set size (individual: $F_{2,56} = 55.50, p < .001, \eta_p^2 = 0.67$; ensemble: $F_{2,56} = 5.04, p = .010, \eta_p^2 = 0.15$), as shown in Figure 1.

Models

We computed the average angular distances between model posterior means and participants reports (see 'Individual' data points in Fig.1). The results indicate that both model predicts participants' reports for individual object locations better than the veridical position. For ensemble perception, we compared predicted posterior standard deviations with participants' estimates, see 'Ensemble' data points in Fig.1. Here, the EEM predictions are clearly better than IEM, which tends to be 'overconfident' as a consequence of Eqn. 1. We also calculated the log-bayes factors (bf) for each participant across and within each set size condition, as well as the mean log-bayes factor per set size condition across all participants. On average, the EEM is the better explanation for all participants, monotonically increasing from 3 objects (Mdn=3.48, IQR=-4.3-13.5), up to 10 objects (Mdn=94.67, IQR=-63.7-140.1).

Finally, to assess the variability associated with the latent variables X and E in the two models, we calculated the average estimated Σ across and within each set size condition (see preprint in the repo for details). In the EEM, the total variance (in both directions) of the object locations is divided between Σ_X and Σ_E, hence these (co)variances are smaller than Σ_X in the IEM. Furthermore, for three objects $\Sigma_X < \Sigma_E$ in the EEM. This results in an ensemble location that is 'pulled' towards the geometric center of the individual objects and away from the screen center, reflecting participants' behavior.

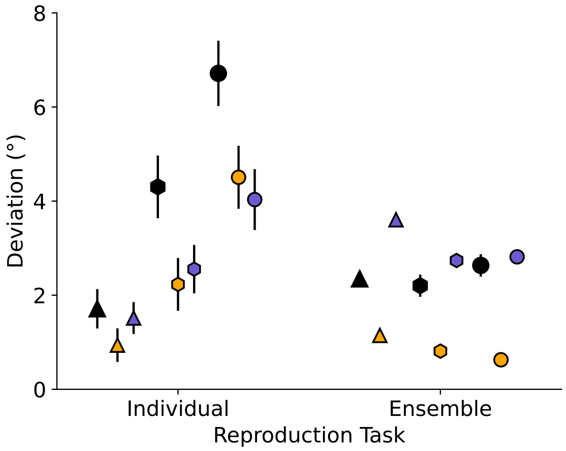

Figure 1: Mean reproduction errors (black) for individual and ensemble average reproduction task for all three set sizes with confidence intervals as error bars. Individual errors are relative to veridical positions, ensemble errors are relative to average veridical positions. Also shown are the median angular distances between model posterior means and participants reports in the indivudal reproduction taks for the IEM (orange) and EEM (purple). For the ensemble perception task, mean posterior standard deviations are shown.

Discussion

Set size influences both individual and ensemble perception within naturalistic scenes.Furthermore, the EEM demonstrated a better fit for the observed locating behavior. Notably, the advantage of the EEM became more pronounced at larger set sizes. The worse fit of the IEM can possibly be explained by its 'overconfidence' of locating the ensemble position. This likely reflects that the current models account only for encoding variance, ignoring factors like memory-based uncertainty (Robinson & Brady, 2023). Other sources of locating errors (memory, motor) will be investigated in the future.

Our study contributes to a more dynamic view of individual and ensemble perception in real-world contexts, by flexibly adapting the perceptual strategies to the task demands. Building models that fuse both modes of perception and can flexibly weight them in a continuous fashion, rather than contrasting them as we did here, will be interesting for future research.

Acknowledgements: this work was supported by the Research Cluster "The Adaptive Mind", funded by the Excellence Program of the HMWK and the Collaborative Research Center SFB/TRR 135, project A4 and C6, under grant agreement no. 62202647.

References

Bishop, C. M. (2006). *Pattern recognition and machine* *learning*. Springer. Retrieved from /bib/bishop/Bishop2006/Pattern-Recognition-and-Machine-Learning-Christophe-M-Bishop.pdf,/bib/bishop/Bishop2006/978-0-387-31073-2_sm.pdf,https://www.microsoft.com/en-us/research/people/cmbishop/#!prml-book

Harrison, W. J., McMaster, J. M., & Bays, P. M. (2021). Limited memory for ensemble statistics in visual change detection. *Cognition*, *214*, 104763. doi: 10.1016/j.cognition.2021.104763

Lew, T. F., & Vul, E. (2015). Ensemble clustering in visual working memory biases location memories and reduces the weber noise of relative positions. *Journal of Vision*, *15*(4), 10. doi: 10.1167/15.4.10

Melcher, D., Huber-Huber, C., & Wutz, A. (2021). Enumerating the forest before the trees: The time courses of estimation-based and individuation-based numerical processing. *Attention, Perception, & Psychophysics*, *83*(3), 1215–1229. doi: 10.3758/s13414-020-02188-0

Neumann, M. F., Ng, R., Rhodes, G., & Palermo, R. (2018). Ensemble coding of face identity is not independent of the coding of individual identity. *Quarterly Journal of Experimental Psychology*, *71*(6), 1357–1366. doi: 10.1080/17470218.2017.1327988

Pertzov, Y., Heider, M., Liang, Y., & Husain, M. (2015). Effects of healthy ageing on precision and binding of object location in visual short term memory. *Psychology and Aging*, *30*(1), 26–35. doi: 10.1037/pag0000029

Robinson, M. M., & Brady, T. F. (2023). A quantitative model of ensemble perception as summed activation in feature space. *Nature Human Behaviour*, *7*, 1638–1651. doi: 10.1038/s41562-023-01634-0