

Models of allocentric coding for reaching in naturalistic visual scenes

Parisa Abedi Khoozani (Okpa@queensu.ca)
Queen's Univ., Kingston, ON, Canada

Dominik Endres
(dominik.endres@uni-marburg.de)
Philipps-University Marburg, Marburg, Germany

Paul R. Schrater (schrater@umn.edu)
Univ. Minnesota, Minneapolis, MN

Katja Fiehler
(Katja.Fiehler@psychol.uni-giessen.de)
Justus-Liebig Univ. Giessen, Giessen, Germany

Gunnar Blohm (blohmq@queensu.ca)
Queen's Univ., Kingston, ON, Canada

Abstract:

To reach to objects, humans rely on relative positions of target objects to surrounding objects (allocentric) as well as to their own bodies (egocentric). Previous studies demonstrated that scene configuration and object relevancy to the task modulates the combination weights of allocentric and egocentric information. Egocentric coding for reaching is studied extensively; however, how allocentric information is coupled and used in reaching is unknown. Using a computational approach, we show that clustering mechanisms for allocentric coding combined with causal Bayesian integration of allocentric and egocentric information can account for the observed reaching behavior. To further understand allocentric coding, we propose two strategies, global vs. distributed landmark clustering (GLC vs. DLC). Both models can replicate the current data but each has distinct implications. GLC efficiently encodes the scene relative to a single virtual reference but loses all the local structure information. In contrary, DLC stores more redundant inter-object relationship information. Consequently, DLC is more sensitive to the changes of the scene. Further experiments must differentiate between the two proposed strategies.

Keywords: reference frames; scene configuration; clustering; statistical ensemble; causal Bayesian inference

Introduction

Previous studies demonstrated that humans combine allocentric and egocentric information when reaching toward visual targets (Byrne & Crawford, 2010; Klinghammer, Blohm, & Fiehler, 2015, 2017). While a vast number of studies investigated how egocentric reference frames are used in movement planning, little is known about how humans incorporate allocentric information in movement planning.

To this aim, a series of experiments investigated the role of allocentric information in reaching (Klinghammer et al., 2015, 2017). Participants were asked to memorize a scene configuration (encoding). After a short delay, a new scene appeared and participants were instructed to reach to missing object

position (decoding). Figure 1 illustrates the task procedure.

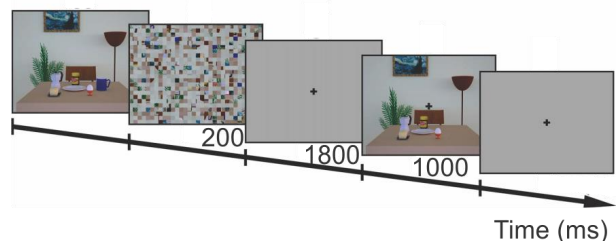


Figure 1. Task procedure, adapted from (Klinghammer et al., 2015).

To change the integrity of allocentric information, sometimes in the decoding scene objects were shifted (horizontally), either in the same or different directions. Figure 2 shows examples of the encoding and decoding scenes. In addition, the number of shifted objects varies; either all the objects in the scene were shifted or only a subset (e.g. 1 out of 5 or 3 out of 5) were shifted. We observed that shifting the objects in opposite direction or shifting only a subset of objects as opposed to the whole group of objects resulted in lower reliance on the allocentric information. That is the shift in reaching end points were smaller for conditions that violated the scene consistency. While shifting objects in the same direction or shifting all the objects together resulted in higher reaching end point biases. In addition, participants only considered the shifts of the objects that were potential future targets (relevant objects, RO) and almost ignored the changes in the configuration of the rest of the objects (irrelevant objects, IO). Finally, we observed that violating scene consistency, also, increased movement variability. Whereas this data provides clear evidence that contextual factors of the scene modulate the combination weights of egocentric and allocentric information, the underlying mechanisms of allocentric coding is unknown.

To this aim, we propose two chunking/clustering approaches for allocentric landmark coding. Clustering is a suggested method to compress the information of

complex scenes (Lew & Vul, 2015). These models are used to reconstruct the memorized scene and casual Bayesian inference was used to identify the consistency between the memorized scene and the current scene. Thus, egocentric and allocentric information are combined using causal Bayesian integration. Through simulations, we demonstrate that both models can replicate the current experimental data, while each model has distinct implications.

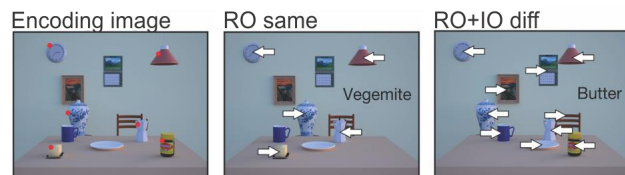


Figure 2. One example of encoding scene and two different decoding scenes. RO same: ROs are shifted in the same direction, and RO+IO diff: ROs and IOs are shifted in different directions. adapted from (Klinghammer et al., 2017).

Materials and Methods

Modeling the allocentric coding

Similar to the experimental setup, we included two phases in our model: encoding and decoding. In the encoding phase the goal is to memorize the scene configuration. Similar to suggested mechanisms in the visual working memory literature, we use chunking or clustering to compress the amount of information of the complex scenes (Lew & Vul, 2015) and propose two modeling paradigms: global versus distributed allocentric landmark coding. The former encodes the scene by creating a global landmark (cluster) and calculates the distance of objects from this point. The latter encodes the position of the target using Barycentric coordinates relative to a distributed set of landmark points (object clusters). At the decoding phase, the goal is to infer the position of the target from a new scene (allocentric) and remembered information from encoding (egocentric). To make this inference, the global model combines the egocentric and the new scene's global cluster points to reconstruct target position, while the distributed model combines the reconstructed target position from the new scene's clusters and encoding Barycentric coordinates with its remembered egocentric position. In both paradigms, causal Bayesian integration would be used to infer the reliability of allocentric information and its contribution in final estimation of the missing target position. Following demonstrates the mathematical implementation of our proposal.

Encoding

The main challenge in the encoding is to memorize the scene configuration given the limited memory

resources. We are proposing two compression mechanisms: global landmark clustering (GLC) and distributed landmark clustering (DLC). We assumed that the internal representation of the objects in the brain can be modeled as a Gaussian distribution.

Global landmark clustering (GLC)

GLC aims to represent the scene configuration with the least amount of resources. This can be achieved by coding the distance of all the objects with regard to the center of the mass of the collection of all the objects. Based on the assumption that the objects have Gaussian distributions, the center of mass (CP) can be calculated as a weighted summation of the object positions. With higher weights assigned to ROs compared to IOs. To summarize the scene configuration, one should calculate the distance vector of each object from the central point. Since all the computations are linear combination of Gaussian distributions, both CP and distance vectors have Gaussian distributions. Storing CP and distance vectors supplies enough information to reconstruct the scene configuration.

Distributed landmark clustering (DLC)

While GLC provides a framework to store scene configuration with the least amount of resources, it ignores all the local statistical structure. In addition, GLC is not sensitive to the relative distances of objects with regard to each other. It has been shown that humans rely less on the far visual landmark. Therefore, DLC aims at restoring the local statistical structure and encodes the positions of the possible target with regards to local clusters of objects. One approach to create local clusters is to use familiar shapes such as triangles or rectangles. In this study we chose triangles and created Barycentric coordinate systems to memorize the scene configuration. Using generalized Barycentric coordinates we can generalize our model to other shapes as well.

Let us assume $T = [T_1 \dots T_{nt}]$ is the set containing all the triangles (nt triangles) created based on the scene configuration. Therefore, each $T_j \in Y_j = [y_{ij}]_{i=1,2,3}$ where i is the vertex's ID and j is the triangle's ID. We can, then, calculate the barycentric coordinates of the target object (X) for each triangle: $X = \Lambda_j * Y_j = \sum \lambda_{ij} * y_{ij}$, where $\Lambda_j = [\lambda_{ij}]_{i=1,2,3}$. Thus, using this method, X is coded nt times.

The goal in the encoding phase is to calculate and store the Barycentric coordinates of the targeted object; $P(\Lambda_j | X, Y_j)$. Using Bays theorem:

$$P(\Lambda_j | X, Y_j) = \frac{P(X, Y_j | \Lambda_j) * P(\Lambda_j)}{P(X, Y_j)} \quad (1)$$

Since we assumed that objects are having Gaussian distribution, we can write $P(X, Y_j | \Lambda_j)$ in the form of:

$$P(X, Y_j | \Lambda_j) = \mathcal{N}\left(\begin{bmatrix} a \\ \mu_{Y_j} \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & \Sigma_{Y_j} \end{bmatrix}\right) \quad (2)$$

Employing conditional probability of Gaussian distributions, we have:

$$\mu_{X|\Lambda_j, Y_j=y} = a + C * \Sigma_{Y_j}^{-1} * (y - \mu_{Y_j}) \quad (3)$$

$$\Sigma_{X|\Lambda_j, Y_j=y} = A - C * \Sigma_{Y_j}^{-1} * C^T \quad (4)$$

Thus, a, A, C can be derived as:

$$a = \Lambda_j * \mu_{Y_j} \quad (5)$$

$$C = \Lambda_j * \Sigma_{Y_j} \quad (6)$$

$$A = \Lambda_j * (\Sigma_{Y_j} + \Sigma_{Y_j}^T) * \Lambda_j^T. \quad (7)$$

Therefore, by storing T and Λ_j 's all the information regarding scene configuration and position of the targeted object can be retrieved later.

Decoding

In the decoding phase the goal is to estimate the position of the missing object based on the current scene configuration ($P(\check{x}_N | \check{x}_1 \dots \check{x}_{N-1})$, where \check{x}_N is the estimated position of the missing object and \check{x}_i is the positions in the current scene). Here, one first needs to use the memorized information from the encoding to reconstruct the scene configuration and then identify if the current scene in the decoding represents the same scene ($c = 1$) as before or if they are different scenes ($c = 2$). Finally, the missing object position should be estimated by combining allocentric and egocentric information. It is worth noting that, since egocentric information is fixed the whole time, we assumed that the CP in GLC and the possible target position in DLC is stored in egocentric coordinates.

Therefore, similar to (Körding et al., 2007), the problem can be formulated as:

$$\check{x}_N = p(c = 1 | \check{x}_1 \dots \check{x}_{N-1}) \check{x}_{N,c=1} + (1 - p(c = 1 | \check{x}_1 \dots \check{x}_{N-1})) \check{x}_{N,c=2} \quad (8)$$

, where $\check{x}_{N,c=1}$ represents the position of the missing object (target) when one is certain that encoding and decoding phase represent the same scene. Similarly, $\check{x}_{N,c=2}$ is the estimation for different scenes in the encoding and decoding phase.

Reconstructing scene configuration

In GLC, if the CP is recovered the whole scene can be reconstructed, therefore the problem can be written as: $P(\check{C}\hat{P} | \check{x}_1 \dots \check{x}_{N-1}, r_1 \dots r_{N-1}) = P(\check{C}\hat{P} | \check{C}\hat{P})$, Where $\check{C}\hat{P}$ is a vector of all the central position estimations ($\check{C}\hat{P} = [\check{C}\hat{P}_1 \dots \check{C}\hat{P}_{N-1}]$) and $\check{C}\hat{P}_i = \check{x}_i - r_i$. Since both \check{x}_i and r_i have Gaussian distribution, $\check{C}\hat{P}_i$ has Gaussian distribution. The task is to estimate the distribution of the CP based on the predicted CPs provided by each individual object in the scene. This problem is similar causal inference in multisensory integration. If the two scenes are related the memorized information and the

current information should be integrated and segregated otherwise. We deployed a similar procedure as (Körding et al., 2007) to build our causal Bayesian integration. Therefore, here we only provide our final analytical solutions to the relevant probabilities and not the overall procedure. To solve equation (8), three components should be calculated and in the following, we briefly explain our solution for each component.

Estimating the probability of the similar scenes

The probability of common cause can be written as:

$$p(c = 1 | \check{C}\hat{P}) = \frac{p(\check{C}\hat{P} | c = 1)p(c = 1)}{p(\check{C}\hat{P} | c = 1)p(c = 1) + p(\check{C}\hat{P} | c = 2)p(c = 2)} \quad (9)$$

, where $p(c = 1) = p_{common}$. For $p(\check{C}\hat{P} | c = 1)$ we obtain:

$$p(\check{C}\hat{P} | c = 1) = \iiint p(\check{C}\hat{P} | CP_{ego}) p(CP_{ego}) dCP_{ego} \quad (10)$$

It is worth noting that, when there is a common cause, $\check{C}\hat{P}_i$ s are correlated and the correlation can be calculated based on the encoding assumption. All the factors in the integral are Gaussian and therefore there is an analytical solution:

$$p(\check{C}\hat{P} | c = 1) = \frac{1}{\sqrt{(2\pi)^{k-1} * \det(\Sigma) * \sum_{i,j} \Lambda_{ij}}} * \exp(-0.5(\sum_{ij} \Lambda_{ij} \check{C}\hat{P}_i \check{C}\hat{P}_j - \frac{(\sum_{i,j} \Lambda_{ij} (\check{C}\hat{P}_i + \check{C}\hat{P}_j))^2}{2\sqrt{\sum_{i,j} \Lambda_{ij}}})) \quad (11)$$

, where $\Lambda = \Sigma^{-1}$ and Σ is the covariance matrix of $\check{C}\hat{P}_i$ s.

For $p(\check{C}\hat{P} | c = 2)$, we note that $\check{C}\hat{P}_i$'s are independent and thus we obtain:

$$p(\check{C}\hat{P} | c = 1) = \prod \int p(\check{C}\hat{P}_i | CP_{ego}) p(CP_{ego}) dCP_{ego} \quad (12)$$

Since all these distributions are Gaussian, we can find an analytical solution:

$$p(\check{C}\hat{P} | c = 1) = \sum \frac{1}{\sqrt{2\pi(\delta_{\check{C}\hat{P}_i}^2 + \delta_{CP_{ego}}^2)}} \exp(-0.5(\frac{(\check{C}\hat{P}_i - \mu_{CP_{ego}})^2}{\delta_{\check{C}\hat{P}_i}^2 + \delta_{CP_{ego}}^2})) \quad (13)$$

, where $\mu_{CP_{ego}}$ and $\delta_{CP_{ego}}^2$ are the mean and standard deviation of CP estimation from encoding. Similarly, $\check{C}\hat{P}_i$ and $\delta_{\check{C}\hat{P}_i}^2$ are the estimated CP position and the variability of this estimation from decoding.

Estimating the position of the missing object

To finally estimate the position of the missing object, one needs to calculate the position estimations for when one is certain that the two scenes are related ($C=1$) or not related ($C=2$).

When one is certain that the two scenes are related, the position can be estimated using Bayesian integration. Since all the position estimations from all objects are correlated and therefore based on (Winkler, 1981) the solution is as following:

$$\bar{C}P_{N,c=1} = e'\Sigma^{-1}CP/e'\Sigma^{-1}e \quad (14)$$

, where e is a vector of ones, Σ is the covariance matrix of CP_i 's. Similarly, when one is certain that the two scenes are not related, the segregation is selected and therefore $\bar{C}P_{N,c=2} = CP_{ego}$.

Similar procedures can be used for the DLC.

Model predications and implications

As expected both proposed model could replicate the human behavior. Similar to (Klinghammer et al., 2017), allocentric weight is defined as the regression of the amount of shift in the final reach error to the expected shift in end point reach error for shifted allocentric information. For instance, if all the objects are shifted 5 cm to the right, the predicted end point reach error for allocentric information should be also 5 cm. As Figure 3 illustrates, violating the scene coherency modulated the allocentric weight. Specifically, when all the objects shifted coherently, participants relied on allocentric information the most and decreased their reliance as the number of shifted object decrease. But this is only valid for ROs. In other words, only the changes in the scene imposed by ROs were considered in the reaching movements.

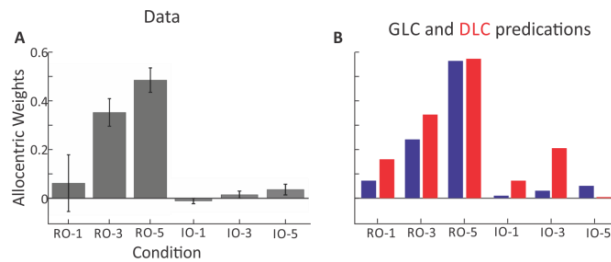


Figure 3. Both GLC and DLC can replicate the measure allocentric weight. The numbers beside the RO/IO represent the number of shifted objects; e.g. RO-1: one relevant object was shifted.

While both models predict human behavior, each has distinct implications. GLC encodes the statistical structure in a very compact format. However, it causes the loss of the local structure information. This means that when the overall structure of the scene remains intact, such as magnification of the scene, GLC predicts the same position for the missing object even though the surrounding objects are shifted. On the other hand, DLC requires more resources to store some of the local structures. This added complexity can provide higher sensitivity to the changes of the scene. This tradeoff can be taken into account to choose different clustering strategies for task relevant and task irrelevant objects. For instance, one can use the GLC to memorize the overall structure of the irrelevant objects. While a more detailed approach such as DLC would be more beneficiary to memorize

the task relevant objects configuration. Future experiments can shed light on how humans select between these strategies.

Acknowledgments

This project was funded by NSERC (Canada) and DAAD (Germany) and DFG (Germany).

References

- Byrne, P. A., & Crawford, J. D. (2010). Cue Reliability and a Landmark Stability Heuristic Determine Relative Weighting Between Egocentric and Allocentric Visual Information in Memory-Guided Reach. *Journal of Neurophysiology*. <https://doi.org/10.1152/jn.01008.2009>
- Klinghammer, M., Blohm, G., & Fiehler, K. (2015). Contextual factors determine the use of allocentric information for reaching in a naturalistic scene. *Journal of Vision*, *15*(13), 1–13. <https://doi.org/10.1167/15.13.24>
- Klinghammer, M., Blohm, G., & Fiehler, K. (2017). Scene configuration and object reliability affect the use of allocentric information for memory-guided reaching. *Frontiers in Neuroscience*, *11*(APR), 1–13. <https://doi.org/10.3389/fnins.2017.00204>
- Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *2*(9). <https://doi.org/10.1371/journal.pone.0000943>
- 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. <https://doi.org/10.1167/15.4.10>
- Winkler, R. L. (1981). Combining Probability Distributions from Dependent Information Sources. *Management Science*, *27*(4), 479–488.