# The Role of the Center-of-Mass in Evaluating Spatial Language

Thomas Kluth[1], Michele Burigo[1], Holger Schultheis[2], and Pia Knoeferle[3]

[1] CITEC (Cognitive Interaction Technology Excellence Cluster), Bielefeld University, Inspiration 1, 33619 Bielefeld, Germany, {tkluth, mburigo}@cit-ec.uni-bielefeld.de
[2] Cognitive Systems Group, Department of Computer Science, University of Bremen, Enrique-Schmidt-Str. 5, 28359 Bremen, Germany, schulth@informatik.uni-bremen.de
[3] Department of German Language and Linguistics, Humboldt University, Unter den Linden 6, 10099 Berlin, Germany, pia.knoeferle@hu-berlin.de

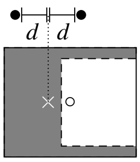
**Abstract.** Consider a display with a circle and a rectangle and the sentence "The circle is above rectangle." How well does the sentence describe the display? For such an acceptability rating of a spatial preposition, the location of the center-of-mass of the rectangle (reference object, RO) is assumed to play an important role. However, there is only little empirical evidence that favors the use of the center-of-mass over other possible reference points of the RO (e.g., the center-of-object). We present an empirical rating study that contrasts the center-of-mass with the center-of-object of an RO by using asymmetrical ROs. The results of the study suggest that people base their acceptability ratings on the center-of-object instead of on the center-of-mass of an RO. Computer simulations of cognitive models implementing this strategy support this view.

**Keywords:** spatial language; center-of-mass, center-of-object, cognitive modeling.

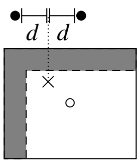
**Introduction** Spatial language is an important part of spatial cognition. People use spatial terms to express their mental representations of space. In this paper, we focus on the acceptability of projective spatial prepositions such as "above" or "to the left of" for describing a scene. Imagine you look at a picture that contains geometrical shapes and hear a sentence like "The circle is above the rectangle". This sentence locates the circle (*located object*, LO) relative to the rectangle (*reference object*, RO). Whether this sentence is an acceptable description of the scene depends on the relative locations of these two objects.

According to [4,5], people use two points of the RO as anchor for their acceptability ratings: The proximal point (the point of the RO that is closest to the LO) and the *center-of-mass* (CoM) of the RO. The orientations of the two imaginary lines that connect each of these two points with the LO (simplified as a single point) are called *proximal orientation* or *center-of-mass orientation*, respectively. [5] provide evidence that both of these orientations affect acceptability ratings of spatial prepositions. In this paper, we focus on the role of the CoM orientation and contrast the role of the CoM with another central point in the RO, the *center-of-object* (CoO). We define the CoO as the following point: $CoO(x,y) = \left(RO_{x0} + \frac{RO_{width}}{2}, RO_{y0} + \frac{RO_{height}}{2}\right)$, where $RO_{x0}$ is the leftmost point of the RO and $RO_{y0}$ is the point of the RO with the lowest y-coordinate (y-axis grows from bottom to top). The CoO coincides with the center of the bounding box of the RO (the smallest rectangle that includes all points of RO). In Figs. 1a and 1b, the bounding box is depicted as solid line, the CoO is depicted as ○, and the CoM is depicted as ×.

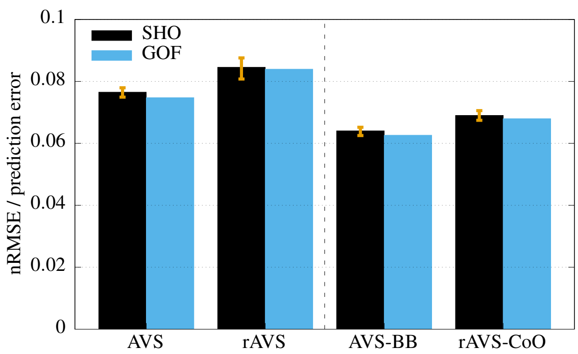
Although research on saccadic and perceptual localization has revealed that the CoM may not be the only critical point for object localization (e.g., [2,7]), the possibility of reference points other than the CoM have so far not been studied in spatial language. In most spatial language acceptability rating tasks, symmetrical ROs were used for which the location of the CoM coincides with the location of the CoO. To our knowledge, there exists only one experiment explicitly designed to dissociate the CoM from the CoO (exp. 4 from [5]). In this experiment, however, only four LOs above two ROs were tested (in total eight LOs). For ROs with a cavity on their top, the results suggest that the CoM is more important than the CoO. For ROs that have a cavity at their bottom (i.e., ROs with a flat top), the results are less clear. We conducted a study to more closely contrast the importance of the CoM with the importance of the CoO using a larger set of items (28 LOs above 4 ROs with flat tops, in total 112 LOs).




(a) Asym. "C"




(b) Asym. "L"




(c) Goodness-Of-Fit (GOF) and Simple-Hold-Out (SHO) results.

Fig. 1: (a) and (b): Stimuli used for the computational and empirical studies (dashed line = borders of the RO, solid line = bounding box of the RO, × = center-of-mass, ○ = center-of-object, ●= LO), (c): Results of model simulations.

**Empirical Study** We designed asymmetrical ROs for which the CoM is dissociated from the CoO (see Figs. 1a and 1b where the CoM is depicted as × and the CoO is depicted as ○). To control for left-right biases we also included vertically mirrored versions of these ROs. We placed 28 LOs at different positions above each RO, resulting in 112 tested LOs in total (28 LOs × 4 ROs). (We also tested 4 more ROs as well as LOs placed below the ROs with the preposition "unter" (below/under). The results for these other 4 ROs and "unter" are not discussed here.)

For the predictions of the study, consider Fig. 1a. Since the two LOs (black dots) are placed at the same elevation with equal horizontal distance $d$ to the CoM of the RO, both LOs have the same CoM orientation. Since they also have equal proximal orientations, people should rate these two LOs identically (following the reasoning by [4,5]). However, if instead the CoO is more important for the acceptability of spatial language, the right LO should be rated higher than the left LO, as it is closer to being directly above the CoO.

Each of our 34 participants saw the German sentence “Der Punkt ist über dem Objekt” (“The point is above/over the object”). After they read the sentence they had to press the space bar. Then, a picture appeared on the screen showing one RO and one LO. Participants had to rate how well the sentence described the depicted scene on a scale from 1 (sentence does not describe the picture at all) to 9 (sentence describes the picture perfectly) using the number keys above the letter keys on a keyboard. Each participant rated all LOs in a pseudo-randomized order (the same RO never appeared twice in a row.)

*Results* Interestingly, the CoM orientation did not have the expected effect as LOs with the same CoM orientation were rated differently: LOs that were placed above the mass of the RO were rated lower than LOs that were placed above the cavity of the RO (mean difference: 0.518, 95% confidence intervals: 0.619, 0.428). In contrast, our results suggest that people use the CoO of an RO in rating its acceptability. LOs with the same CoO orientation were rated equally: LOs on the left side of the RO received equal ratings compared to LOs on the right side of the RO (mean difference: 0.034, 95% confidence intervals: -0.101, 0.165).

We dissociated the CoM from the CoO in our stimuli, permitting us to dissociate the effects of the corresponding orientations on ratings. The results of this comparison suggested that what was previously thought to be an effect of the CoM is in fact an effect of the CoO, at least for ROs with flat tops. [5] also used asymmetrical ROs in their experiment 4 to contrast the CoO with the CoM. While they found the CoM to be more important for the RO without a flat top, they could not find such effect for the RO with a flat top.

**Model Simulations** Two cognitive models that compute spatial language acceptability ratings rely on the CoM for their computations: the *Attentional Vector Sum* model (AVS, [5]) and the *reversed AVS* (rAVS) model, a recently proposed modification of the AVS Model (see [1] for a motivation of the rAVS model and details of both models). According to our empirical findings, however, people seem to use the CoO instead. This is why we next present refined versions of both models.

The AVS model relies on the CoM because it computes a vector sum using all points of the RO. We modified the AVS model so that it uses all points that are in the bounding box of the RO and call this modification AVS-BB model. The bounding box is the smallest rectangle that includes all points of the RO (see solid line in Figs. 1a and 1b). In particular, the bounding box also includes the points inside the cavity of asymmetrical ROs. Computing a vector sum with all points inside the bounding box then means that the AVS-BB model relies on the CoO instead of on the CoM.[4]

The rAVS model explicitly uses the location of the CoM of the RO and can be easily modified to use the CoO instead by replacing every occurence of the CoM with the CoO. This yields the rAVS-CoO model. The AVS-BB model and the rAVS-CoO model both show the same output pattern as the empirical data: LOs above the mass are rated lower than LOs above the cavity of the RO because ratings peak above the CoO.

[4] One could also add a parameter $\gamma$ that gives different attentional weights to points that are in the bounding box but outside the RO (i.e., in the cavity of the RO) compared to points that are in the RO. This would create a model that could behave like the AVS model ($\gamma = 0$) or like the AVS-BB model ($\gamma = 1$). Since this adds possibly unneeded flexibility to the model, we decided against this implementation.

We fitted all four models to the 112 empirical mean ratings by searching for values of the free model parameters that provide minimal normalized Root Mean Square Errors (nRMSE). The resulting Goodness-Of-Fit (GOF) values are plotted in Fig. 1c. An nRMSE of 0.0 means that the model is able to reproduce the empirical data exactly, while an nRMSE of 1.0 means that model output and empirical data are maximally different. As can be seen in Fig. 1c, all models can closely fit the data (GOF $< 0.084$). The versions of the models that use the CoO (AVS-BB and rAVS-CoO) fit the data better than the original versions (GOF $< 0.068$).

Since a good fit to data is a necessary but not sufficient property of a model (e.g., the good fit might be the result of overfitting due to an overly flexible model, see [3]), we also assessed the model with the simple hold-out (SHO) method proposed by [6]. [6] showed that this model selection method provides results comparable to other model selection methods. The SHO method is a cross-validation method: The data is randomly split into a training and a test set and the parameters of the model are estimated on the training set. Using the best parameters for the training set, a prediction error on the test set is computed (again an nRMSE). This is done several times with different random splits of the data. Fig. 1c shows the median prediction error of 101 SHO iterations together with their bootstrapped 95% confidence intervals. These results now clearly favor the modified versions of the models (lower SHO without overlapping confidence intervals for AVS-BB and rAVS-CoO compared to AVS and rAVS). Also, both versions of the AVS model outperform the corresponding versions of the rAVS model: AVS performs better than rAVS and AVS-BB performs better than rAVS-CoO.

**Conclusion** We presented an acceptability rating study of spatial prepositions using asymmetrical ROs that allowed us to explicitly contrast the importance of the CoM with the CoO. In contrast to previous literature claiming that people use the CoM of the RO as reference point ([4,5]), our results suggest that people rather select the CoO of the RO as reference point. Furthermore, we modified two cognitive models in order to implement this strategy. These modified models performed considerably better than the original models that rely on the CoM of the RO. This corroborates the importance of the CoO over the CoM.

## References

1. Kluth, T., Burigo, M., Knoeferle, P.: Shifts of attention during spatial language comprehension: A computational investigation. In: Proc. of ICAART 2016, Vol. 2. pp. 213–222 (2016)
2. Nuthmann, A., Henderson, J.M.: Object-based attentional selection in scene viewing. Journal of vision 10(8), 20–20 (2010)
3. Pitt, M.A., Myung, I.J.: When a good fit can be bad. Trends Cogn Sci 6(10), 421–425 (2002)
4. Regier, T.: The human semantic potential: Spatial language and constrained connectionism. MIT Press, Cambridge, Mass. (1996)
5. Regier, T., Carlson, L.A.: Grounding spatial language in perception: An empirical and computational investigation. Journal of Experimental Psychology: General 130(2), 273–298 (2001)
6. Schultheis, H., Singhaniya, A., Chaplot, D.S.: Comparing model comparison methods. In: Proc. of the Ann Conf of the Cog Sci Soc. pp. 1294 – 1299. Cog Sci Soc, Austin, TX (2013)
7. Vishwanath, D., Kowler, E.: Localization of shapes: Eye movements and perception compared. Vision Research 43(15), 1637–1653 (2003)