Universität Bielefeld

Language and Cognition
Center of Excellence
Cognitive Interaction Technology

CITEC

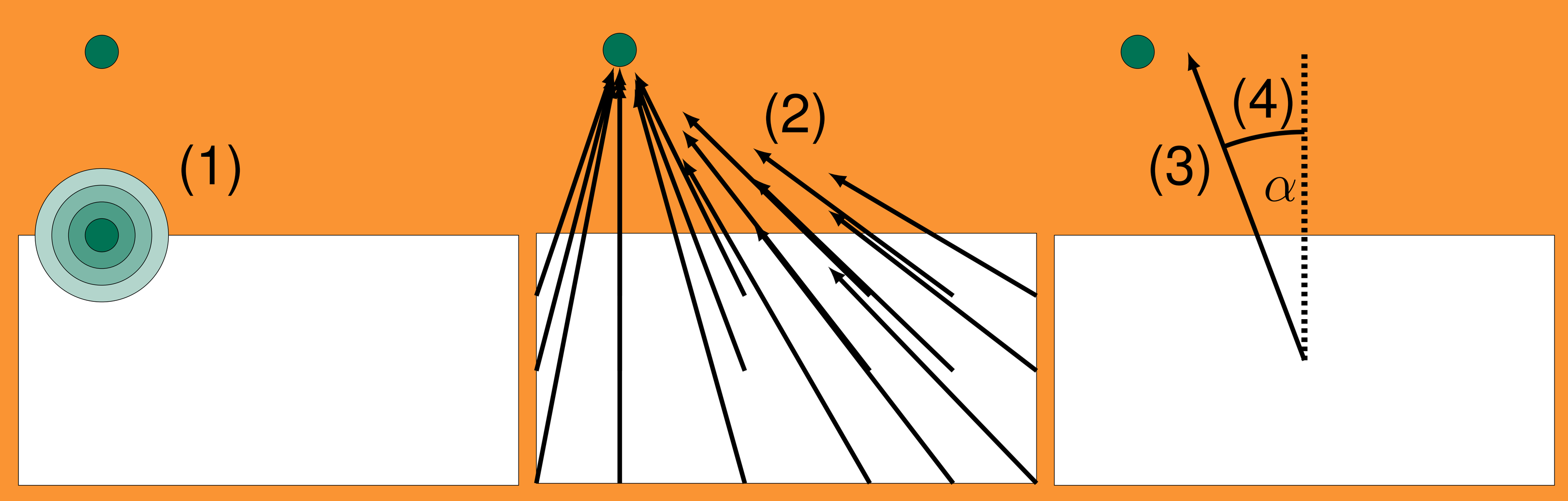
Spatial Language Comprehension: A Computational Investigation of the Directionality of Attention

Thomas Kluth, Michele Burigo, Pia Knoeferle (contact: tkluth@cit-ec.uni-bielefeld.de)

Introduction

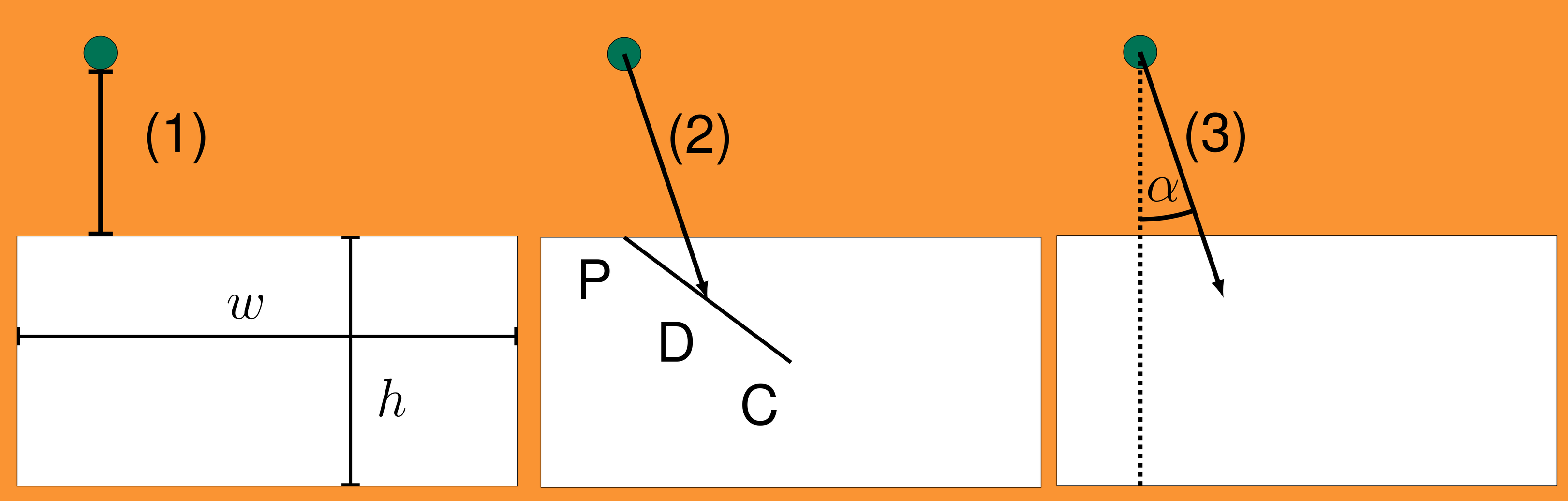
- How do humans understand spatial language? Example:
  - The circle (located object, LO) is above the rectangle (reference object, RO).
- Two conflicting views:
  - Attentional Vector Sum model (AVS, Regier & Carlson, 2001, see below): attention shifts from the RO to the LO (inspired from Logan & Sadler, 1996)
  - visual world paradigm findings (Burigo & Knoeferle, 2015; Chambers et al., 2002): attention shifts from the LO to the RO (in the order the objects are mentioned; further evidence: Roth & Franconeri, 2012)
- In what order are the two objects attended?
- To integrate the visual world findings (Burigo & Knoeferle, 2015) in the AVS model, this work proposes a modified version of the AVS model with a **reversed attentional shift** from the LO to the RO (the **rAVS** model, see below).

AVS model (Regier & Carlson, 2001)



(1) Attention is focused on top of the RO and decays exponentially.
(2) At every point i of the RO a vector is rooted, pointing to the LO and weighted by the amount of attention at that point i.
(3) All vectors are summed.
(4) The vector sum is compared with upright vertical (in the case of above): the higher the deviation α, the lower the final rating.

rAVS model (proposed model)



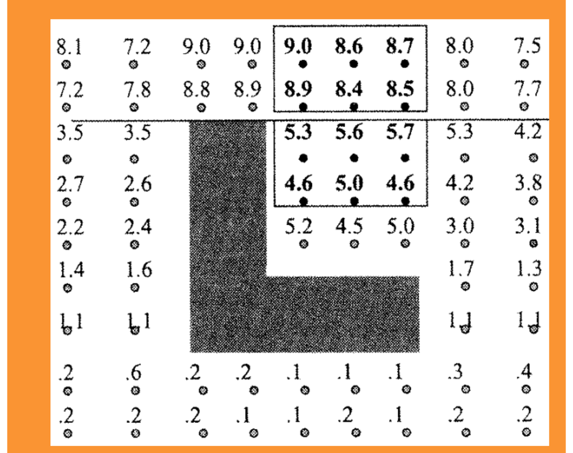
(1) Compute relative distance between LO and proximal point on the RO (*relative*: considering width and height of the RO).
(2) One vector points from the LO to the RO. The end point of the vector always lies on the drawn line that connects the proximal point P with the center-of-mass of the RO (C). Great relative distance: vector points more toward C; low relative distance: vector points more toward P.
(3) Deviation α from canonical downwards is computed: The higher the deviation, the lower the rating (as in the AVS model).

- no vector sum (because LO consists of only one point)
- attentional distribution is computationally irrelevant

Method

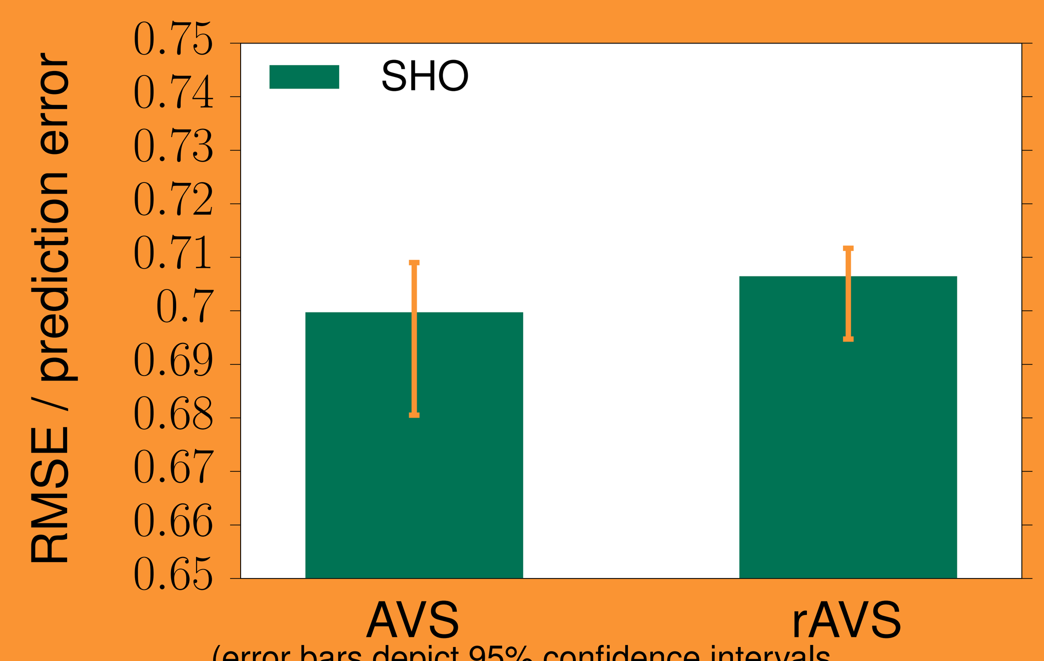
- Both models compute acceptability ratings given a RO, a LO and a spatial preposition.
- We fitted both models to empirical ratings (all seven experiments from Regier & Carlson, 2001; see box Data below) with the simple hold-out method (SHO, Schultheis et al., 2013). SHO controls for the complexity of the models.
- Root Mean Square Error (RMSE) as error measure:
  $RMSE = \sqrt{\frac{1}{n}\sum_i^n (data_i - modelOut_i)^2}$
- lower RMSE → better model:
  - best possible RMSE = 0 (model and data are exactly the same)
  - worst possible RMSE = 9 (model and data are maximally different)

Data (example)



example; complete data set contains 10 ROs and 337 LOs (see Regier & Carlson, 2001)
(Image Source: Regier & Carlson, 2001, p. 287)

Results



(error bars depict 95% confidence intervals computed with 100,000 bootstrap samples)

Discussion

- Both models account equally well for the data.

→ Simulations support both directionalities of the attentional shift.

- rAVS is computationally less complex:
  - single vector instead of vector sum (due to simplification of the LO as in the AVS model)
  - attentional distribution is computationally irrelevant
- Models predict differences for particular configurations:
  - AVS' predictions are hard to grasp (due to its flexibility).
  - PSP analysis (Pitt et al., 2006) confirms that rAVS is less flexible than AVS and leads to precise predictions.

→ Next steps: conduct an experiment to contrast (distinct predictions of) the two models.

References

Burigo, M., & Knoeferle, P. (2015). Visual attention during spatial language comprehension. *PloS ONE*, *10*(1), e0115758. doi: 10.1371/journal.pone.0115758
Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49.
Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and Space* (pp. 493–530). The MIT Press.
Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, *113*(1), 57.
Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, *130*(2), 273–298.
Roth, J. C., & Franconeri, S. L. (2012). Asymmetric coding of categorical spatial relations in both language and vision. *Frontiers in Psychology*, *3*(464). doi: 10.3389/fpsyg.2012.00464
Schultheis, H., Singhaniya, A., & Chaplot, D. S. (2013). Comparing model comparison methods. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1294 – 1299). Austin, TX: Cognitive Science Society.

Architectures and Mechanisms for Language Processing (AMLaP) 2015, Valetta, Malta