Investigating the Parameter Space of Cognitive Models of Spatial Language Comprehension

Thomas Kluth * **Michele Burigo** * **Pia Knoeferle** **

* *CITEC (Cognitive Interaction Technology Excellence Cluster), Bielefeld University, Inspiration 1, 33619 Bielefeld, Germany (e-mail: tkluth@cit-ec.uni-bielefeld.de, mburigo@cit-ec.uni-bielefeld.de)*
** *Department of German Language and Linguistics, Humboldt University, Unter den Linden 6, 10099 Berlin, Germany (e-mail: pia.knoeferle@hu-berlin.de)*

Abstract: Cognitive models are – due to their computational nature – useful for the development and improvement of artificial cognitive systems. However, if two models perform equally well on the existent data, comparing them directly can permit us to select the more appropriate one. One way of comparing to models is to perform an in-depth analysis of their predictions. In this study, we compared the predictions of two similar cognitive models of spatial language comprehension using the Parameter Space Partitioning (PSP) algorithm proposed by Pitt, Kim, Navarro, and Myung (2006).

Keywords: Spatial Language, Cognitive Modeling, Parameter Space Partitioning

1. INTRODUCTION

Humans should be able to interact with technical systems as intuitively as possible. An intuitive way of interaction between humans and technical systems is by means of natural language. If technical systems were able to understand and generate natural language, humans could easily interact with them. In order to enable technical systems to use natural language it is crucial to understand how humans use language. Insights into linguistic processes and representations could then be used for technical replication of natural language processing. Here, it is helpful if the linguistic theories that should be implemented are precisely formulated, for instance as a computational model.

Computational models of human language processing are one specific example of the general endeavor to investigate human cognition with the help of computational modeling, also called cognitive modeling. A cognitive model aims to precisely formulate the assumed processes and mechanisms of human cognition. Once a model has been formulated, it can be implemented and assessed with empirical data to test whether the model reflects human cognition (both qualitatively and quantitatively). Cognitive modeling started in the 1960s and shows many facets nowadays (for recent overviews see McClelland, 2009; Shiffrin, 2010; Sun, 2008). Cognitive models can accommodate a wide variety of cognitive and behavioral processes in humans (e.g., visual attention: Bundesen & Habekost, 2008; working memory: Miyake & Shah, 1999; emotion: Marsella, Gratch, & Petta, 2010; analogy-making: French, 2002). Apart from different approaches to model cognition (e.g., symbolic models, neural networks, dynamical systems), cognitive architectures also exist (e.g., ACT-R: Anderson et al., 2004, SOAR: Laird, Newell, & Rosenbloom, 1987). Instead of focusing on an isolated part of human cognition, cognitive architectures strive to accommodate the processes (from stimulus perception to a motor response) implicated in completing a specific task.

Linguists have formalized their theories as computational models since the 1980s (e.g., Elman, 1990; McClelland & Elman, 1986) and computational modeling is still an active research area in linguistics (e.g., Crocker, Knoeferle, & Mayberry, 2010; Kukona & Tabor, 2011; Smith, Monaghan, & Huettig, 2013; see Crocker, 2010, for an overview). Recent computational models that were developed to investigate human language processing reflect the state-of-the-art in linguistic research. Due to their computational nature they can be implemented into technical systems and ideally enable these technical systems to use natural language in a human-like manner.

In the domain of spatial language, we have proposed a cognitive model that integrates recent empirical findings into an existing model (Kluth, Burigo, & Knoeferle, 2016). This model had been inspired by the AVS model (Regier & Carlson, 2001) , a model which had been developed to accommodate results from acceptability rating studies (how good is the fit between a spatial term such as “above” and an object depiction). The AVS model accommodates these acceptability ratings as a shift in attention between two objects. In the AVS model, the shift of attention is assumed to occur from a reference object to a to-be-located object (as theorized by Logan, 1995 and Logan & Sadler, 1996). This directionality, however, conflicts with recent findings of a shift of (visual) attention in the opposite direction (from the LO to the RO, see Burigo & Knoeferle, 2015; Franconeri, Scimeca, Roth, Helseth, & Kahn, 2012; Roth & Franconeri, 2012). Accordingly, we have proposed the reversed AVS (rAVS) model that reverses the direction of the attentional shift in the AVS model based on these empirical findings. As most cognitive models, both the AVS model and the rAVS model use free parameters to closely fit empirical data (e.g., to adapt their output to individual differences or noise). The two models provide a comparable fit on the data from Regier and Carlson (2001), and this despite the fact that they implement the attentional shift in different ways (see Kluth et al., 2016).

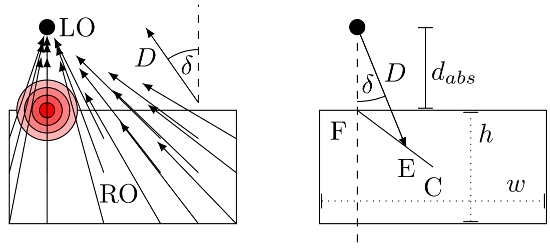
Having two different models that perform equally well on the existing data, the question arises whether one of the models better reflects the human use of spatial language. Selecting one among two or more competing cognitive models is a question that can be addressed by several different methods (for reviews see, e.g., Myung, Tang, & Pitt, 2010; Shiffrin, Lee, Kim, & Wagenmakers, 2008). Providing a good fit to empirical data is a necessary feature of a good model but not the only one. Apart from that a good cognitive model is also as simple as possible (Vandekerckhove, Matzke, & Wagenmakers, 2015), generalizes well to unseen data (Pitt & Myung, 2002), and makes distinct predictions (Roberts & Pashler, 2000). Roberts and Pashler (2000) discuss the use of goodness-of-fit measures for the assessment of computational models and suggest: "A better way to test a theory with free parameters is to determine how the theory constrains possible outcomes (i.e., what it predicts) [...]" (Roberts & Pashler, 2000, p. 358). A model with few predictions can then be falsified by providing empirical data that conflict with the model's predictions. However, if a model generates numerous predictions, its fit to any empirical data is less impressive (Roberts & Pashler, 2000).

In this contribution, we analyze whether the AVS model and the rAVS model make different predictions for stimuli that have not yet been tested. In doing so, we assess the predictive ability of the models as suggested by Roberts and Pashler (2000). Moreover, if the two models predict different outcomes, we could run an empirical study that tests these distinct predictions. Such a test would permit us to see which of the model predictions are supported by the empirical results. The models supported by the test data would then be selected as a more appropriate approximation of how humans process spatial language.

Based on the mechanisms of the models, we designed stimuli for which we hypothesized that the models predict different outcomes. However, our hypothesized predictions may not be the only output that the models are able to generate. Given the range and the interaction of the models' parameters, the models might produce output that conflict with our hypothesized predictions. To analyze the range of possible predictions, we applied the Parameter Space Partitioning (PSP) algorithm proposed by Pitt et al. (2006). The results of the PSP analysis can be used to see what qualitative data patterns the models are able to generate. These results in turn can inform an empirical study that could provide data on which the models perform differently and thus could be distinguished.

2. THE MODELS

To discuss the different predictions of the AVS model and the rAVS model, we will first briefly introduce the two



(a) In the AVS model, vectors point from the RO to the LO and are weighted with attention (red). The sum D of these vectors is compared to a reference direction (dashed) and yields the deviation δ.

(b) In the rAVS model, one vector points from the LO to one point E on the line that connects point F with point C (depending on the relative distance of the LO). This vector D is compared to a reference direction (dashed) and yields the deviation δ.

Fig. 1. Schematized steps of (a) the AVS model and (b) the rAVS model.

models. A more detailed description of the two models can be found in Regier and Carlson (2001) and Kluth et al. (2016). Both models compute an acceptability rating of a spatial preposition – given the location and shape of a reference object (RO), the location of a located object (LO), and the spatial preposition. In one example context of house-hold robotics, a possible utterance that contains spatial language is the following: "Robot, bring me my key. It is to the left of the phone." Here, the spatial preposition "to the left of" describes the location of the key (the located object, LO) relative to the phone (the reference object, RO). Where should the robot look first for the key? Probably, a good start would be the location "to the left of the phone". The AVS model and the rAVS model are able to compute acceptability ratings for all possible locations around the RO. To this end, both models rely on the angular deviation δ of a direction D from a reference direction (e.g., canonical upright: if the LO is directly above the RO, the angular deviation δ would be zero). If the angular deviation δ is high, a low acceptability rating is returned (and vice versa). The direction D that results from the relative placement of the two objects is computed differently in the two models.

2.1 The AVS model

In the AVS model, the direction D is the weighted sum of a population of vectors that point from every point of the RO to the LO (see Figure 1a). This sum is weighted with attention, where attention is defined as an exponential decay function which is highest at the focal point F (point on top of the RO that is vertically aligned with the LO). The distance of the LO from the RO affects the distribution of attention: A nearby LO results in a narrow distribution of attention (i.e., a large decline of attention from point F) whereas a distant LO gives a broad distribution of a attention (i.e., a small decline of attention from point F).

2.2 The rAVS model

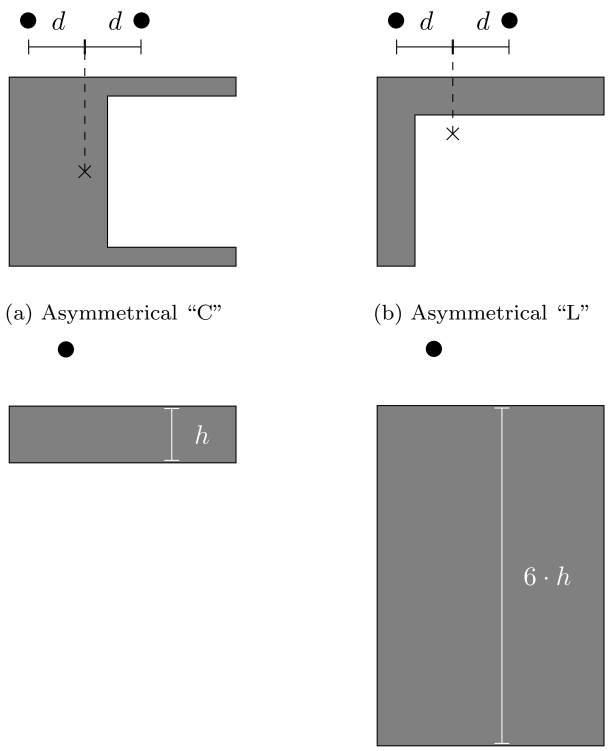
In the reversed AVS (rAVS) model, the relative placement of the LO and the RO is used differently to obtain the direction D that is compared to a reference direction. To integrate recent findings about shifts of attention during spatial language comprehension (see Kluth et al., 2016, for details), the direction of the vectors is reversed in the rAVS model: Instead of pointing from the RO to the LO, the vectors are pointing from the LO to the RO. Since the LO is simplified as a single point in the AVS model and the rAVS model tries to stay as similar as possible to the AVS model, this change results in a vector sum that consists of a single vector only (see Figure 1b). The end point E of this vector lies always on the line that connects the point F (the same point as the focal point in the AVS model) with the point C (the center-of-mass of the RO). The relative distance of the LO to the RO determines the end point E: If the LO is close to the RO, E is close to F, if the LO is far from the LO, E is close to C. Here, *relative* distance is defined as the distance of the LO to the RO divided by both the width and the height of the RO.

3. PREDICTIONS

We have shown that the two models (the AVS model and the rAVS model) accommodate the existing results equally well (see Kluth et al., 2016). In order to assess whether any of the models reflects human use of spatial language more than the other, we follow the reasoning of Roberts and Pashler (2000) and "[d]etermine the predictions [of the models] [u]sing intuition, experience, and trial and error[...]"(Roberts & Pashler, 2000, p 363-364). Based on the mechanisms of the models, we found two types of configurations of an RO and an LO for which the models seem to predict qualitatively different ratings. We will first discuss these stimuli and why we hypothesize the models predict different ratings for them. Because both models comprise four free parameters, the models might produce predictions different from our hypotheses. This is why we subsequently apply the parameter space partitioning (PSP) algorithm (Pitt et al., 2006) that helps us to see whether our intuitive predictions reflect the output that the models are actually able to compute.

Asymmetrical Objects The first prediction is based on the representation of asymmetrical ROs (see Figures 2a and 2b). Consider, for example, the two LOs shown in Figure 2a (represented by black dots and placed with equal horizontal distance from the center-of-mass $\times$ of the RO): the AVS model would produce a higher rating for the LO placed on the left of the center-of-mass of the RO than for the LO placed on the right of the center-of-mass. This is because the AVS model represents every single point of the RO with one vector and the asymmetrical RO has more points below the left than the right LO. Thus, the vector sum consists of more vectors with lower deviation from the reference direction for the left LO compared to the right LO. This should yield an overall direction D that has a lower deviation δ (leading to a higher rating) for the left LO compared to the right LO.

Instead of representing all points of the RO, by contrast, the computation of the rAVS model is mainly based on the center-of-mass of the RO. Due to the same horizontal distance to the center-of-mass, the rAVS model computes basically the same deviation for both LOs in Figure 2a. Therefore, the rAVS model predicts the same rating for both LOs in Figure 2a. For both models, the same reasoning applies for the two LOs in Figure 2b.



(a) Asymmetrical "C" (b) Asymmetrical "L"

(c) "Thin" rectangle (d) "Tall" rectangle

Fig. 2. Displays for which the AVS model and the rAVS model seem to predict different ratings. LOs are displayed as black dots, center-of-mass of the ROs in (a) and (b) are displayed as $\times$. The LOs in (a) and (b) have the same horizontal distance d from the center-of-mass of the RO. The height of the RO in (d) is six times the height h of the RO in (c).

Relative Distance The second prediction concerns the implementation of the distance between the LO and the RO in the models. The rAVS model explicitly uses *relative* distance for its computation, where relative distance is computed as the absolute distance divided by both the width and the height of the RO. Given two LOs with the same absolute distance but different relative distances from the RO, the rAVS model predicts higher ratings for the LO with the lower relative distance than for the LO with the higher relative distance. In Figures 2c and 2d, the rAVS model would rate the LO above the "thin" rectangle (Figure 2c) lower in acceptability than the LO above the "tall" rectangle (Figure 2d).

In the AVS model, by contrast, the *absolute* distance influences the attentional distribution. The width and height of the RO are incorporated in the vector sum. The vector sum depends on the attentional distribution that is additionally controlled by a free parameter. This parameter provides flexibility for the vector sum: On the one hand, a narrow distribution of attention might result in an equal representation of the "thin" and the "tall" rectangle (the lower part of the "tall" rectangle only receives a negligible amount of attention and does not change the final direction D). On the other hand, a broader distribution of attention might change the final direction

D through the additional vectors for the "tall" rectangle. Therefore, the prediction of the AVS model in this case is unclear.

4. METHOD: PARAMETER SPACE PARTITIONING

Both models have four parameters that interact in a complex way, making it hard to understand in a principle-based manner the model predictions for different RO-LO configurations. The parameter space partitioning algorithm (PSP) proposed by Pitt et al. (2006) helped us to investigate whether the two models actually predict different rating patterns for the specific RO-LO configurations shown in Figures 2a-2d and whether the models also compute ratings that do not follow our hypothesized predictions.

The PSP algorithm is a Markov Chain Monte Carlo (MCMC) based method that searches in the parameter space of the models for regions of patterns that are qualitatively different and also estimates the volumes of these pattern regions. With these volumes it is possible to see what qualitative patterns the model is able to generate (i.e., what the model predicts) and which volume in the parameter space is occupied by these patterns. The volume can be interpreted as the importance of the predictions for the model: The larger the volume of a pattern, the more important this prediction is for the model. A pattern with a large volume means that the model generates the same qualitative pattern throughout a substantial range of parameter settings.

In our study, we coded qualitative patterns in the following way: We code the comparison of two LOs in one digit. If the ratings for two compared LOs do not differ by more than an equality threshold t, we code this as a '`0`'. If the first LO has a higher rating than the second one, we code this as a '`1`'. If the second LO has a higher rating than the first one, we code this as a '`-1`'. We compared ratings for the two LOs above the asymmetrical "C" (Figure 2a), resulting in a first code. We next compared ratings for the two LOs above the asymmetrical "L" (Figure 2b), resulting in a second code. Finally, we compared ratings for the LO above the "thin" rectangle (Figure 2c) to ratings for the LO above the "tall" rectangle (Figure 2d), resulting in a third code. Accordingly, we obtain a three-digit code. Applying this coding to our hypothesized predictions described above, the rAVS model should generate the pattern '`00-1`' (no differences for the LOs above the asymmetrical objects and a lower rating for the relatively far LO above the "thin" rectangle compared to the relatively close LO above the "tall" rectangle). The AVS model should generate the pattern '`11?`' (higher ratings for the LOs with more mass below them and an unclear prediction for the condition in which the relative distance is manipulated).

But when should two ratings be considered equal? In the studies reported by Regier and Carlson (2001), a difference in mean ratings of 0.17 (exp. 1), 0.2 (exp. 2), 0.3 (exp. 4), or 0.7 (exp. 6) was needed to reach significance. In the study conducted by Carlson-Radvansky, Covey, and Lattanzi (1999) this difference was 0.3. In the study by Hörberg (2008) a difference of 0.57 was required. Experiment 2 by Burigo, Coventry, Cangelosi, and Lynott (in press) needed differences of 0.27, 0.34, 0.36, 0.43 to reach significance. Although all these studies investigated different effects on the acceptability of spatial prepositions and also used slightly different rating scales, these values provide hints at the magnitude for the equality threshold t. Based on these values, we used the following three equality thresholds t for our PSP analysis: 0.1, 0.5 and 1.0.

To conduct the PSP analysis, we used the `MATLAB` implementation of the PSP algorithm made available by Pitt et al. (2006) [1] together with the `C++` implementation of the rAVS and the AVS model available at Kluth (2016). We constrained the boundaries of the parameter space as reported in Kluth et al. (2016).

5. RESULTS

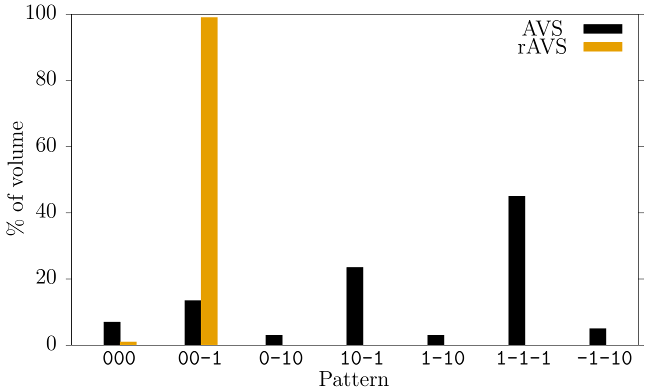
The results of the PSP analysis separately for each value of t are shown in Figure 3. Plotted are mean relative volume estimates of three PSP runs.

rAVS Throughout all ts, the rAVS model generates only 2 out of 27 theoretically possible patterns: '`000`' and '`00-1`'. With a larger threshold t the volume of pattern '`000`' increases while the volume of pattern '`00-1`' decreases. This is reasonable: If the difference d of two ratings lies between two equality thresholds (say, $d = -0.7$), this difference is either coded as a '`-1`' for $t = 0.5$ or as a '`0`' for $t = 1.0$. The PSP analysis confirms the hypothesized predictions described earlier: For all parameter settings, the LOs above the two asymmetrical objects were rated equally by the rAVS model (the first two digits in the patterns are always '`0`'). For most parameter settings, the LO above the "thin" rectangle is rated lower than the LO above the "tall" rectangle (the last digit of the pattern is '`-1`'). However, there exist parameter settings for which the rAVS model predicts no differences in ratings for these two LOs.

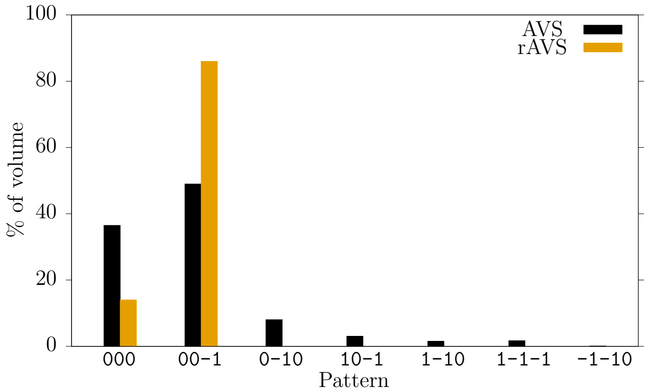
AVS The AVS model generates a greater range of qualitative patterns. For $t = 0.1$ (Figure 3a) and $t = 0.5$ (Figure 3b), it generates 7 out of 27 theoretically possible patterns. Interestingly, the pattern we predicted earlier ('`11?`') is not generated by the model (for none of the values of t). Moreover, for $t = 0.1$, large proportions of the parameter space yield patterns that conflict with our hypothesized predictions ('`10-1`', '`1-10`', '`1-1-1`'). There is no intuitive explanation why the AVS model should predict different directions for the differences in ratings for the LOs above the asymmetrical objects ('`1`' vs '`-1`'). Due to its flexibility, the vector sum incorporated in the AVS model seems to generate a wide range of ratings.

If the equality threshold t is set to 0.5, the AVS model still generates the same seven patterns (Figure 3b). The proportions of the parameter space in which the model generates these patterns, however, have changed. Mostly, the AVS model now generates the same patterns as the rAVS model, but in a different proportion. Roughly half of the parameter space generates '`00-1`' and almost 40 % generates '`000`'.

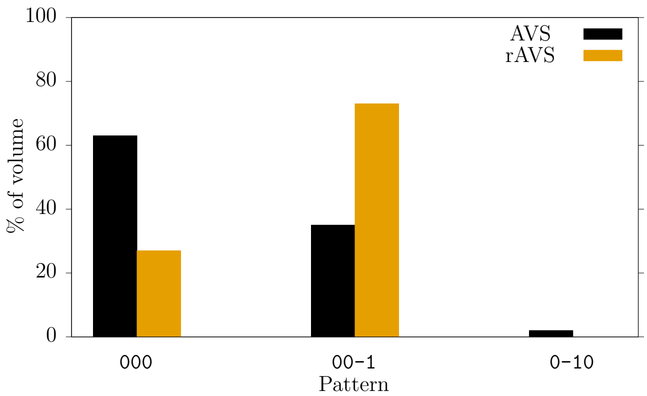
[1] This implementation is available at `http://faculty.psy.ohio-state.edu/myung/personal/psp.html`. We slightly changed the implementation to be able to use it with `GNU Octave` (Eaton, Bateman, Hauberg, & Wehbring, 2014).



(a) Equality of Ratings $t = 0.1$



(b) Equality of Ratings $t = 0.5$



(c) Equality of Ratings $t = 1.0$

Fig. 3. Results of the PSP analysis: Proportions of volume that each pattern occupies in the parameter space. Patterns that were not generated are not shown. Two generated ratings were considered equal if they differed by less than (a) 0.1 (b) 0.5 (c) 1.0.

For $t = 1.0$, the proportion of these two patterns switches. Now, a greater set of parameter settings generates '000' (more than 60 %) whereas the pattern '00-1' is generated in less than 40 % of the parameter space. The AVS model still generates a third pattern ('0-10'), but this pattern only occupies a small volume in the parameter space.

5.1 Discussion

The PSP analysis confirms the hypothesized predictions for the rAVS model: the LOs above the asymmetrical objects receive the same ratings and – for most parameter settings – the LO above the "thin" rectangle is rated lower than the LO above the "tall" rectangle. For the AVS model, the PSP results uncover a greater range of output for the same input compared with the output range of the rAVS model. However, the AVS model does not generate our hypothesized prediction '11?'. Arguably, the predictions of the AVS model are harder to grasp intuitively.

The rAVS model generates more distinct predictions, as it only produces 2 out of 27 possible data patterns (in contrast to the AVS model that computes 7 out of 27 patterns). If humans produce the patterns predicted by the rAVS model, this would provide more support for the rAVS model than for the AVS model (Roberts & Pashler, 2000). This is because the AVS model produces a greater range of predictions and is thus more difficult to falsify.

For $t = 1.0$, both models are able to generate the two patterns '000' and '00-1', i.e., both models predict the same qualitative patterns. If humans also generate either of these patterns, both models should fit these empirical data quite well. Accordingly, conducting a study with the stimuli used for the PSP analyis might not provide data that can distinguish the two models. The rAVS model, however, more often predicts '00-1' than '000', whereas the AVS model more often predicts '000' than '00-1'. So, there is a trend in these predictions that might show up in empirical data and thus would slightly support one of the models over the other.

6. CONCLUSION

The interaction of humans with technical systems would be facilitated if the technical systems were able to interpret and generate natural language. To implement natural language into technical systems, cognitive models of language could prove useful. Recent cognitive models reflect the state-of-the-art in linguistics while being thoroughly assessed with empirical data. In this contribution, we investigated the parameter space of two similar cognitive models of spatial language understanding (the AVS model, Regier & Carlson, 2001, and the rAVS model, Kluth et al., 2016). Since the two models cannot be distinguished on the existing data, we analyzed the predictions of the models to see whether any of the models better accounts for human comprehension of spatial language. Following Roberts and Pashler (2000), a good model of human cognition should constrain the range of predictions. Based on the mechanisms of the models we identified stimuli on which we hypothesized they predict different outcomes. We then applied the PSP algorithm (Pitt et al., 2006) to see whether the models follow our hypotheses. The PSP analysis confirmed the hypotheses for the rAVS model. For the AVS model, however, our hypotheses were not confirmed. Arguably then, it is more difficult to translate the mechanisms of the AVS model into testable predictions. Moreover, the rAVS model constrains the range of possible outcomes to a greater extent than the AVS model.

ACKNOWLEDGEMENTS

This research was supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG) and by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement nº 316748 awarded to PK.

REFERENCES

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036. doi: 10.1037/0033-295X.111.4.1036

Bundesen, C., & Habekost, T. (2008). *Principles of Visual Attention: Linking Mind and Brain.* Oxford, New York: Oxford University Press. doi: 10.1093/acprof:oso/9780198570707.001.0001

Burigo, M., Coventry, K. R., Cangelosi, A., & Lynott, D. (in press). Spatial Language and Converseness. *Quarterly Journal of Experimental Psychology*.

Burigo, M., & Knoeferle, P. (2015). Visual attention during spatial language comprehension. *PloS ONE*, *10*(1), e0115758. doi: 10.1371/journal.pone.0115758

Carlson-Radvansky, L. A., Covey, E. S., & Lattanzi, K. M. (1999). What effects on where: Functional influences on spatial relations. *Psychological Science*, *10*(6), 516–521.

Crocker, M. W. (2010). Computational psycholinguistics. In A. Clark, C. Fox, & S. Lappin (Eds.), *The Handbook of Computational Linguistics and Natural Language Processing* (chap. 17). Wiley-Blackwell.

Crocker, M. W., Knoeferle, P., & Mayberry, M. R. (2010). Situated sentence processing: The coordinated interplay account and a neurobehavioral model. *Brain and Language*, *112*(3), 189–201. doi: 10.1016/j.bandl.2009.03.004

Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. (2014). *GNU Octave version 3.8.1 manual: a high-level interactive language for numerical computations.* CreateSpace Independent Publishing Platform. Retrieved from `http://www.gnu.org/software/octave/doc/interpreter` (ISBN 1441413006)

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Franconeri, S. L., Scimeca, J. M., Roth, J. C., Helseth, S. A., & Kahn, L. E. (2012). Flexible visual processing of spatial relationships. *Cognition*, *122*(2), 210–227.

French, R. M. (2002). The computational modeling of analogy-making. *Trends in Cognitive Sciences*, *6*(5), 200–205.

Hörberg, T. (2008). Influences of form and function on the acceptability of projective prepositions in swedish. *Spatial Cognition & Computation*, *8*(3), 193–218. doi: 10.1080/13875860801993652

Kluth, T. (2016). *A C++ Implementation of the reversed Attentional Vector Sum (rAVS) model.* Bielefeld University. doi: 10.4119/unibi/2900103

Kluth, T., Burigo, M., & Knoeferle, P. (2016). Shifts of attention during spatial language comprehension: A computational investigation. In *Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART 2016) – Volume 2* (pp. 213–222). SCITEPRESS – Science and Technology Publications, Lda.

Kukona, A., & Tabor, W. (2011). Impulse processing: A dynamical systems model of incremental eye movements in the visual world paradigm. *Cognitive Science*, *35*(6), 1009–1051. doi: 10.1111/j.1551-6709.2011.01180.x

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, *33*(1), 1–64. doi: 10.1016/0004-3702(87)90050-6

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, *28*(2), 103–174.

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.

Marsella, S., Gratch, J., & Petta, P. (2010). Computational models of emotion. In K. R. Scherer, T. Bänziger, & E. B. Roesch (Eds.), *A Blueprint for Affective Computing: A Sourcebook and Manual* (pp. 21–46). Oxford, New York: Oxford University Press.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*, 11–38. doi: 10.1111/j.1756-8765.2008.01003.x

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Pressm.

Myung, J. I., Tang, Y., & Pitt, M. A. (2010). Evaluation and comparison of computational models. In M. L. Johnson (Ed.), *Essential Numerical Computer Methods* (Vol. 454, p. 511-527). Elsevier.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, *113*(1), 57–83. doi: 10.1037/0033-295X.113.1.57

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*(10), 421–425.

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. doi: 10.1037//0096-3445.130.2.273

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358-367. doi: 10.1037//0033-295X.107.2.358

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

Shiffrin, R. M. (2010). Perspectives on modeling in cognitive science. *Topics in Cognitive Science*, *2*(4), 736–750. doi: 10.1111/j.1756-8765.2010.01092.x

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical bayesian methods. *Cognitive Science*, *32*(8), 1248–1284. doi: 10.1080/03640210802414826

Smith, A. C., Monaghan, P., & Huettig, F. (2013). An amodal shared resource model of language-mediated visual attention. *Frontiers in Psychology*, *4*(528). doi: 10.3389/fpsyg.2013.00528

Sun, R. (2008). Introduction to computational cognitive modeling. In R. Sun (Ed.), *Cambridge Handbook of Computational Psychology* (pp. 3–20). Cambridge University Press. doi: 10.1017/CBO9780511816772.003

Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2015). Model comparison and the principle of parsimony. In J. T. Townsend & J. R. Busemeyer (Eds.), *The Oxford Handbook of Computational and Mathematical Psychology.* Oxford, New York: Oxford University Press. doi: 10.1093/oxfordhb/9780199957996.013.14