Authors
Kriz, Sarah
Trafton, J. Gregory
McCurry, J. Malcolm

Peer reviewed

The Role of Spatial Information in Referential Communication: Speaker and Addressee Preferences for Disambiguating Objects

Sarah Kriz (kriz@itd.nrl.navy.mil)
J. Gregory Trafton (trafton@itd.nrl.navy.mil)
J. Malcolm McCurry (mccurry@itd.nrl.navy.mil)

Navy Center for Applied Research in Artificial Intelligence, Naval Research Laboratory
4555 Overlook Ave., SW, Washington, DC 20375 USA

Abstract

This study evaluated the extent to which speakers and addressees utilize information about an object's spatial location during referential communication. In Experiment 1 participants verbally requested target objects in an array. Results showed that speakers preferred to use color information to disambiguate the target object, and they described spatial location only when color could not serve to disambiguate. Experiment 2 tested the comprehension of requests like the ones generated in Experiment 1. For a simple triangular array, participants were faster at comprehending spatial requests than color requests. With a more complex array, the opposite pattern was found. Taken together, these results suggest that while there is some overlap between speaker and addressee preferences, what is easiest for the speaker is not necessarily what is easiest for the addressee.

Keywords: Spatial language; referential communication, psycholinguistics.

There are many ways in which a speaker can verbally refer to an object. A cup on a table may be requested by mentioning its object class, such as, "Get me the cup." If there is potential confusion as to *which* cup, a speaker may refer to the desired cup by describing its object-based properties like color, size, or shape. Likewise, a speaker may refer to the cup's spatial location in order to disambiguate it from other cups. Given the numerous choices available for object reference, what factors influence how a speaker requests an object?

Research on referential communication has shown speakers often refer to a target object by specifying more than one of its properties (i.e., color *and* size), even though only one feature is necessary to disambiguate the object from others (Belke & Meyer, 2002; Deutsch & Pechmann, 1982; Hermann & Deutsch, 1976; Whitehurst, 1976). Most often, color is the feature that is overspecified (i.e., it is described but it does not aid in disambiguating), and properties such as size are less likely to be overspecified.

Previous experiments on referential communication have controlled tasks in such a way that spatial location information could not be used as a potential disambiguating feature. Likewise, studies of spatial language production have generally limited tasks so that speakers are forced to refer to objects based solely on their spatial locations (e.g., Emmorey & Casey, 2002; Schober, 1995; Tenbrink, 2005). While these studies have been instrumental in assessing how either object-based or spatial location features are specified when referring to an object, they have not addressed the situations in which speakers actually use spatial descriptions to disambiguate objects, given the availability of other perceptual features. To our knowledge, no one has yet evaluated when speakers choose to use spatial location information versus non-spatial, object-based features to refer to a target object in an array.

Whether speakers choose to use spatial location information in referential communication can inform us on the interaction between perception, conceptualization, and communication. In order to produce a referential utterance, a speaker must first visually evaluate the features of a target object. These features include its color, shape, size, pattern, and presumably, spatial location. Dimensions on which the target object differs from the other objects must be identified, and the speaker then has to formulate a linguistic message that provides enough contrastive information to disambiguate the target object from the others.

It is crucial to note that the spatial location is a dimension in which an object will always differ from other objects, as every object occupies a unique area of space. Thus, the use of spatial location information to disambiguate a target object would ensure an unambiguous reference to the target object. For this reason, computing and specifying spatial location may be a good strategy to adopt. However, conceptualizing spatial relations may require more cognitive resources than simply comparing visual features like the color or size of objects. The question we seek to answer is whether spatial location, which is always reliable, is a more viable option than visually comparing object-based features of the target object and the other objects. We attempt to answer this question by evaluating the frequency with which speakers choose to produce spatial location specifications over color specifications.

Most previous studies assessing multidimensional features in referential communication have focused on either language production or comprehension. This presents a problem for most referential communication theories because the theories cannot adequately address both sides of communication. For instance, researchers studying object-based feature descriptions have come to a general consensus that the overspecification of color in referential descriptions is due to the ease and speed in which color is processed, and that suppressing the inclusion of color in a description takes

more cognitive effort than overspecifying the information (Belke & Meyer, 2002; Whitehurst, 1976). However, the overspecification of color, especially when it does not provide disambiguating information, creates non-contrastive information that the addressee must process. Obviously, the inclusion of such redundant information is not ideal from the addressee's perspective. In this sense, there is a trade-off between what is easier for the speaker and what is easier for the addressee. By attempting to evaluate both referential language production and comprehension, we hope to build a better theory of how referential communication is optimized for both the speaker and the addressee.

In this study, we have isolated the processes of language production and comprehension in order to gain as much experimental control as possible. Experiment 1 evaluates factors that influence speakers to refer to an object either by its spatial location, its color, or both. Experiment 2 explores how listeners comprehend the types of descriptions that were generated in Experiment 1.

Figure 1: Schematics of simple and complex arrays.

Experiment 1: Language Production

Experiment 1 evaluated the factors that influence how speakers refer to a target object in order to disambiguate it from other objects in an array. Participants were shown objects placed on a tabletop in either a simple or complex configuration (see Figure 1), and were asked to request a target object.

Due to previous findings suggesting that children had more difficulty in producing good contrastive descriptions for larger arrays (Whitehurst, 1976), we varied whether participants viewed simple or complex arrays in order to determine whether this effect would be replicated in adult language. In Whitehurst's (1976) experiment, the addition of more objects created more comparisons for the children to compute. Along these lines, more objects also create spatial relations that are more complex and, in principle, take more effort to compute. For instance, note that positions 2, 3, 6, and 7 in the complex array of Figure 1 are much harder to conceptualize than the relations shown in the simple array. Thus, we hypothesized that speakers would use fewer spatial requests in the complex array. We predicted that speakers in the simple array condition would be more likely to use spatial descriptions because of the ease

in which the spatial relations between the objects could be conceptualized and described linguistically.

This experiment was designed so that one-third of the trials showed objects of the same color, thereby forcing speakers to disambiguate the target object on the basis of spatial location. The remaining two-thirds of the trials allowed a speaker to choose whether to disambiguate the target object by its color or spatial location, or both. Based on the ease in which color information is visually processed (Belke & Meyer, 2002; Treisman & Gelade, 1980), we predicted that, in general, when color could disambiguate the target object from the others, speakers would give color requests. Taking this prediction together with our prediction of the effect of array complexity on spatial descriptions, we hypothesized an interaction between array complexity and target object color, such that speakers viewing the complex array would specify the color of the target object (as opposed to giving a spatial description) whenever possible. However, for the simple array, we predicted speakers would be more likely to give spatial descriptions even when color could disambiguate the target object from the others.

Based on previous findings that suggest speakers often overspecify object-based feature information (Belke & Meyer, 2002; Deutsch & Pechmann, 1982; Hermann & Deutsch, 1976), we also hypothesized that when speakers produced requests containing spatial location information, they would give overspecifications of color as well, even when color information would not be useful in disambiguating the target object from the other objects in the array.

Method

Participants Thirty-eight undergraduate students from George Mason University were randomly assigned to the simple array or complex array condition so that each condition had 19 participants.

Materials Bitmap image files were created by taking digital photographs of objects arranged on a tabletop. Objects were arranged in either a simple triangular formation or a complex formation (see Figure 1). For the simple array stimuli, six object classes were used: pen caps, jellybeans, thumbtacks, and paper cutouts of triangles, squares, and circles.[1] Each bitmap image showed three objects from the same class in a triangular formation. The objects were either all the same color (i.e., all blue jellybeans), all different colors (i.e., one red, one blue, and one green jellybean), or one object of a single color and the other two objects of another color (i.e., one blue jellybean and two red jellybeans). In the latter case the object of a different color was always the target object. The bitmaps were loaded into

---

[1] The purpose of using different object classes was to vary the stimuli so that participants would not lose interest in the experiment.
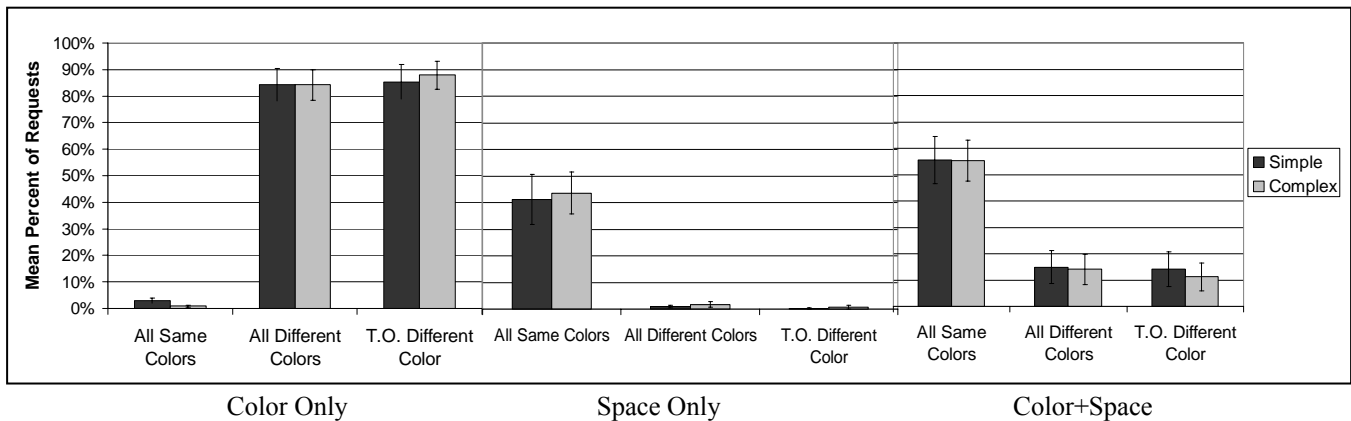
Figure 2: Mean percent of A) color only requests, B) space only requests, and C) color+space requests, by color of objects and array complexity.

E-Prime™ and individual trials were created. For each trial a box briefly appeared around one object in the array in order to signal that object as the target object. Trials were created so that each object in each array was signaled once. The complex trials were constructed similarly, but three object classes were used to create the stimuli, and each array contained eight objects.

To summarize, the simple array trials were constructed from the following factors: 6 (object class: pen caps, jellybeans, triangles, circles, squares, thumbtacks) x 3 (array type: all same color, all different colors, target object different color) x 3 (target object: position 1, 2, or 3). The complex array trials consisted of the following factors: 3 (object class: jellybeans, circles, pencaps) x 3 (array type: all same color, all different colors, target object different color) x 8 (target object placement: positions 1-8). This yielded 72 complex array trials and 54 simple array trials.

Procedure After giving informed consent, participants were seated in front of a Tobii 1750 17" computer monitor and were shown pictures of the objects they would see during the experiment in order to get acquainted with the names of the object classes. At the beginning of each trial, a bitmap showing an array (either simple or complex, depending on the randomly assigned condition) appeared on the monitor. After 1000 msec a box appeared around one of the objects and remained on the screen for 500 msec. The box served to signal the target object. Participants were asked to imagine they had a friend in the room with them, and to request from their friend the object that was signaled. After producing the verbal request, participants pressed a button on the keyboard to advance to the next trial. The stimuli were presented in a random order. Participants were video recorded and eye tracked during the experiment, however, due to length restrictions eye movement data will not be reported in this paper.

Results and Discussion

Verbal requests were transcribed and coded in one of three ways: specifying color information only (e.g., "Get me the red square"), specifying spatial information only (e.g., "Get me the square on the left"), or specifying color+space information (e.g., "Get me the red square on the left"). Raw number counts were then converted into proportions in order to compare across the simple and complex array conditions.

Three separate 2 (array complexity) x 3 (color of objects) ANOVAs were run with the frequency of request type as the dependent variables. As Figure 2 illustrates, no main effect of array complexity was found for any of the request types (Color: $F_{(1,36)}=.01$, n.s.; Space: $F_{(1,36)}=.08$, n.s.; Color+Space: $F_{(1,36)}=.03$, n.s.). Contrary to our predictions, speakers tended to structure their requests similarly regardless of the complexity of the array.

One third of the trials in this experiment presented arrays in which the target object was the same color as the other objects, one third of the trials showed arrays in which all the objects differed in color, and one third of the trials contained target objects that were a different color than the other objects. We predicted that when the target object was not the same color as the other objects (i.e., trials in which all objects were different colors and trials in which the target object was a different color than the others), speakers would request the target object using a color specification. When the target object was the same color as the others, we expected speakers to give a spatial specification of the target object, but to also include color overspecifications (redundancies).

In fact, a main effect of object color was found for all three request types (Color: $F_{(2,72)}=365.02$, $p<.01$, $\eta^2=.91$; Space: $F_{(2,72)}=47.80$, $p<.01$, $\eta^2=.57$; Color+Space: $F_{(2,72)}=51.60$, $p<.01$, $\eta^2=.59$). As Figure 2 shows, when the target object was a different color than the other objects, and

when all of the objects were different colors, speakers gave color requests significantly more often than when the target object was the same color as the others (p<.01 for all comparisons). On the other hand, when the target object was the same color as the other objects in the array, requests specifying spatial location and color+space information were significantly more frequent than when the target object was a different color than the others (p<.01 for both comparisons) or when all the objects were different colors (p<.01 for both comparisons). No significant interactions of array complexity and target object color were found.

We hypothesized that speakers would prefer to redundantly provide color information along with their spatial requests. Color overspecifications accounted for roughly 30% of the requests obtained in this experiment while space-only requests accounted for 15% of all utterances. In other words, two of every three spatial requests contained color overspecifications. As Figure 2 shows, the majority of these overspecifications occurred when the target object was the same color as the others, thus the overspecifications did not provide information that could help in disambiguating the target object.

To summarize, the findings from this experiment suggest that, given the choice, speakers use color specifications rather than spatial information to disambiguate a target object. Furthermore, speakers tended to give redundant color information in their spatial requests.

Experiment 2

Experiment 2 was designed to address the issue of referential communication from the perspective of the addressee. To assess whether the preferences exhibited in Experiment 1 benefit addressees (in terms of time taken to process different request types), we created requests similar to those obtained in Experiment 1 and required participants to respond by choosing the correct target object.

If the preferences from Experiment 1 benefit addressees, we would expect reaction times to pattern similarly—participants should be fast to respond to color, slow to respond to space, and color+space requests should fall somewhere in the middle.

Although array complexity did not affect how speakers produced their requests in Experiment 1, we expected an effect of array complexity on language comprehension. Specifically, we expected slower response times for spatial requests in the complex array, as there are more objects that require attention. As Carlson and Logan (2001) have shown, distractor objects cause an increase in time to verify spatial descriptions. Moreover, the complexity of a larger array requires the use of more complex spatial descriptions (i.e., 'between'), which are conceptually difficult (Quinn, 2005), and presumably take longer to process than simple relational words such as left, right, top, and bottom. However, because color attributes are normally processed pre-attentively and in parallel (Treisman & Gelade, 1980), array complexity should not affect response times for color-only requests.

Method

Participants Fifteen undergraduate students from George Mason University participated in the experiment for course credit.

Materials One hundred and fifty three trials (72 complex, 81 simple) were created in E-Prime™. The trials were essentially the same as those from Experiment 1, except the target object was not signaled visually. Instead, wave files requesting the target object were recorded and were linked in E-Prime to their related bitmaps. For each target object, three requests were made: color-only, space-only, color+space. All requests took the same syntactic structure, namely, "Get me the *OBJECT* that's *MODIFIER*." Color specifications were structured as, "Get me the jellybean that's green." Space specifications were, "Get me the jellybean that's to the right." Color and space specifications were arranged with the color modifier first, followed by the spatial modifier as in, "Get me the jellybean that's green and that's to the right."

The simple trials consisted of the following factors: 3 (object class: pen caps, jellybeans, triangles) x 3 (array type: all same color, all different colors, target object different color) x 3 (target object placement: position 1, 2, or 3) x 3 (request type: color, spatial, color+space). The complex trials were constructed similarly: 1 (object class: circles) x 3 (array type: all same color, all different colors, target object different color) x 8 (target object placement: positions 1-8) x 3 (request type: color, spatial, color+space).

Procedure After giving informed consent, participants were seated in front of a Tobii 1750 17" computer monitor. At the beginning of each trial a bitmap picture appeared and a recorded request for the target object simultaneously played. Participants were told to drag the mouse cursor to the correct object and to click on it as soon as they knew which object was being requested. After clicking on an object the mouse cursor was reset to the center of the screen and a new image and request were simultaneously presented. Stimuli were presented in a random order. Participants were eye tracked during the experiment, but again, eye movement data will not be discussed here.

Results and Discussion

Participants were extremely accurate in selecting the requested object (*M*=98.97%), and only accurate trials were included in the reaction time analysis. Reaction time data was submitted to a 3 (request type) x 2 (array complexity) repeated measures ANOVA. There was a significant main effect of request type ($F(2,28)=80.36$, p<.01, $\eta^2=.85$), and Bonferroni post-hoc tests showed that participants were significantly faster at responding to color requests than space requests, and space requests were significantly faster than color+space requests (p<.01 for all comparisons). This pattern is different than the color<color+space<space

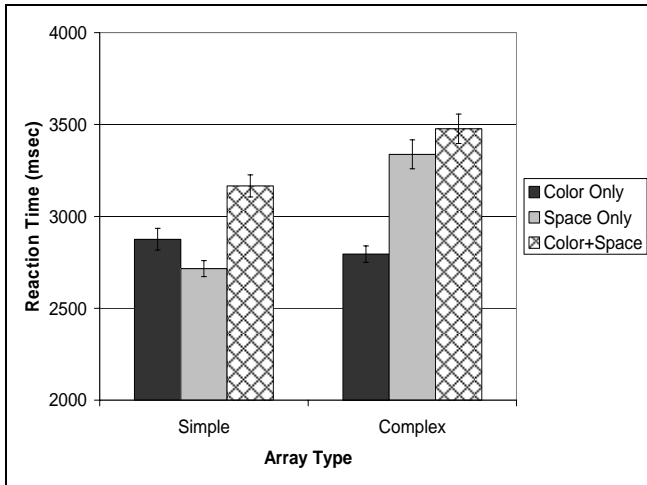pattern we predicted based on the production data from Experiment 1.

Figure 3: Mean reaction times for description type by array complexity.

A significant main effect of array complexity was also obtained (F(1,14)=138.63, p<.01, η^2=.91), such that participants were faster at responding to object requests in the simple array than the complex array. A significant interaction (F(2,28)=66.60, p<.01, η^2=.83) was also found. As Figure 3 suggests, in the simple array participants were significantly faster at responding to spatial descriptions than color and color+space descriptions (p≤.01 for all comparisons). However, the complex array showed a different pattern: participants were significantly faster at responding to color descriptions than spatial or color+space descriptions (p<.01 for all comparisons). Moreover, participants were equally fast at responding to color requests in both arrays, but were faster at responding to spatial descriptions and color+space descriptions in the simple array compared to the complex array (p<.01 for all comparisons). As predicted, array complexity affected the speed in which participants comprehended spatial requests, but not color requests.

One possible confound with these analyses concerns the length of time each spoken utterance took. It could be that the pattern of response times simply reflected the length of the different request types, rather than the time taken to process the verbal information. For instance, spatial requests for the simple array such as, "Get me the jellybean that's on the left," were much shorter than the spatial descriptions for the complex array, which were as long as, "Get me the jellybean that's between the one on the top and the one on the right." Moreover, the color+space requests were naturally longer than all the other request types because they contained both color and spatial descriptions of the target object. To examine whether reaction time differences were due to the length of the requests, we subtracted out the length of the wave file from participants' reaction time for each trial. This left us with the amount of time participants took to respond, after listening to the entire description. (Negative values, because participants responded before the utterance was completed, were not included in this analysis.) The results showed a pattern comparable to the data obtained from the original analysis. Furthermore, participants were told to click on the object as soon as they knew which one was being requested, and in fact, participants clicked on the correct object before the end of the description on 31.94% of the trials. Thus, the significant results we have obtained cannot be attributed to the length of the descriptions. These findings suggest that cognitive and perceptual processing time, not simply listening time, increases when spatial descriptions contain redundant color information.

General Discussion

The purpose of this study was to explore how speakers and addressees utilize spatial location information in a referential communication task. The production data collected in Experiment 1 suggest that speakers generate spatial descriptions to disambiguate a target object only when no other option for disambiguation is available. Speakers exhibited an overwhelming preference for producing color requests, and color information was often redundantly added, even in cases when it did not serve to disambiguate the target object (i.e., all objects were the same color). These results are in line with previously reported findings that suggest color information has a privileged status in referential object descriptions (Belke & Meyer, 2002; Deutsch & Pechmann, 1982; Hermann & Deutsch, 1976; Whitehurst, 1976).

The comprehension data collected in Experiment 2 reflect a pattern that is somewhat, but not entirely, complementary to the production data. Speakers exhibited a preference for producing color requests, and similarly, addressees were generally fast at comprehending color descriptions. The pre-attentive, parallel processing of color (Treisman & Gelade, 1980) seems to serve speakers and addressees equally well.

However, other comparisons between speakers and addressees suggest that strategies are not entirely complementary between the two. Contrary to production data, in which spatial descriptions were used only when absolutely necessary, addressees were faster at comprehending simple spatial descriptions than color descriptions. One explanation for the ease in which simple spatial descriptions are comprehended is the consistency of the cue. Terms such as 'right,' 'left,' 'top,' and 'bottom' always reference the same area of space, and cue visual attention to only this area. Whereas an addressee must wait for the color term and then conduct a visual search on the items in the array, spatial terms direct an addressee's attention to the area of the array in which only one object, the target object, is located. The results suggest that although simple spatial relations may be harder to conceptualize than color, their reliable cueing make them faster to comprehend than color information. This suggests

that features that aid the addressee's comprehension are not always the easiest for the speaker to conceptualize.

Another mismatch between the production and comprehension data indicates that speakers and listeners utilize referential information differently. Redundant color information, although used gratuitously by speakers, contributed to longer processing times for addressees. Obviously, processing uninformative information takes longer because the addressee must try to process it as informative information before it is determined to be unhelpful. Thus, from an addressee's perspective, redundant information of any type is not preferred. However, as has been suggested by others, the suppression of color information takes more cognitive effort for the speaker than the inclusion of color specification (Belke & Meyer, 2002; Whitehurst, 1976). This creates a mismatch between what is easier for the speaker and what is easier for the addressee.

What do these differences between speakers and addressees suggest about the cognitive influences on linguistic communication? One interpretation is that speakers do what is easier for them, and addressees must process the information as it is received, or request that the speaker reformulate the message in a different way. In fact, recent studies in spatial language production suggest that speakers tend to do what is easiest for them, rather than, for instance, describing an environment from an addressee's point of view or a perspective explicitly requested by a conversational partner (Buhl, 2001; Kriz, 2006). However, before jumping to conclusions about whether speakers' cognitive processes guide the structure of communication, we would like to point out that our study did not evaluate naturalistic communication. We have isolated the processes of production and comprehension in order to how spatial and color information is utilized in both. Although we consider both experiments to be quasi-communicative, they do not quite capture what speakers and addressees do when they participate in time-linked interactive communication. Future studies need to evaluate fully communicative situations in order to assess how speakers and listeners determine who must exert more cognitive effort. Additionally, future studies in referential communication should address the competition between spatial location information and other (non-color) object-based features. We may have unfairly biased this study by choosing color as an alternative to spatial language. An interesting follow-up would be to conduct a similar experiment using size of objects, rather than color, as an alternative to spatial location.

Acknowledgments

References

Belke, E. & Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during 'same'-'different' decisions. *European Journal of Cognitive Psychology, 14*(2), 237-266.

Buhl, H. M. (2001). Partner orientation and speaker's knowledge as conflicting parameters in language production. *Journal of Psycholinguistic Research, 30*(6), 549-567.

Carlson, L. A. & Logan, G. D. (2001). Using spatial terms to select an object. *Memory and Cognition, 29*(6), 883-892.

Deutsch, W. & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition, 11*, 159-184.

Emmorey, K. & Casey, S. (2002). Gesture, thought, and spatial language. In K. R. Coventry & P. Olivier (Eds.), *Spatial Language: Cognitive and Computational Perspectives* (pp. 87-101). Dordrecht: Kluwer Academic Publishers.

Gibson, B. S. & Kingstone, A. (2006). Visual attention and the semantics of space. *Psychological Science, 17*(7), 622-627.

Hermann, T. & Deutsch, W. (1976). *Psychologie der Objektbenennung*. Bern: Huber Verlag.

Kriz, S. (2006). Perspective in Spatial Perception, Representations, and Descriptions. Unpublished doctoral dissertation, University of California, Santa Barbara.

Quinn, P. C. (2005). Developmental constraints on the representation of spatial relation information: Evidence from pre-verbal infants. In L. Carlson & E. van der Zee (Eds.), *Functional Features in Language and Space: Insights from Perception, Categorization, and Development* (pp. 293-309). Oxford: Oxford University Press.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes, 20*, 219-247.

Tenbrink, T. (2005). Identifying objects on the basis of spatial contrast: An empirical study. In C. Freska, M. Knauff, B. Krieg-Bruckner, B. Nebel, T. Barkowsky (Eds.), *Spatial Cognition IV: Reasoning, Action, and Interaction* (pp. 124-146). Berlin: Springer Verlag.

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Whitehurst, G. J. (1976). The development of communication: Changes with age and modeling. *Child Development, 47*, 473-482.