Selective Attention by Structural Alignment: An Eyetracking Study

Aaron B. Hoffman (aaron.hoffman@mail.utexas.edu)
Department of Psychology, 1 University Station A8000
Austin, Texas 78712-0187

Bradley C. Love (brad_love@mail.utexas.edu)
Department of Psychology, 1 University Station A8000
Austin, Texas 78712-0187

Arthur B. Markman (markman@psy.utexas.edu)
Department of Psychology, 1 University Station A8000
Austin, Texas 78712-0187

Abstract

A potential determinant of people's selective attention is offered by the structural-alignment view of comparison. This view holds that objects are compared via structured representations that align sets of features that share relational roles. A central claim of this account is that the comparison process directs attention towards alignable features. This prediction has been supported by offline measures by Markman and Gentner (1997), who showed that alignable features serve as better cues for recall than nonalignable features. The present study provides the first online test of the structure-alignment theory's claim that alignability drives selective attention. Consistent with this, we show that in addition to serving as better cues for recall, alignable differences are attended more than nonalignable differences. Within-trial attention dynamics revealed that attention to alignable differences increases over the course of the comparison process.

Keywords: comparison, alignment, attention, recall, eye movements, eye tracking

Introduction

The amount of information that inundates people's perceptual systems creates a significant challenge. As people move through their environment, they are faced with thousands of decisions about which information they should selectively attend and which they should filter out. They must decide that certain things are worth remembering and that others are not. How are such decisions made?

There are a variety of factors that influence selection of parts of the stimulus stream. Early work examining how people attend to complex visual scenes showed that people will fixate the most informative elements (Buswell, 1935; Mackworth & Morandi, 1967; see Henderson & Hollingworth, 1999 for review). Subsequent work explored people's tendency to attend to the most perceptually salient features (e.g., Henderson, Weeks, & Hollingworth, 1999; Parkhurst, Law, & Niebur, 2002). Work on schemata and memory suggests that semantic consistency with a schema determines what is later recalled (Bransford & Johnson, 1972; 1973; Brewer & Dupree, 1983; Rummelhart, 1980). Finally, recent eye tracking work in categorization (Rehder, Colner, & Hoffman, 2009) and in natural scene perception (e.g., Hayhoe, Shrivastava, Mruczek, & Pelz, 2003) proposes that the information demands of the task are the biggest influences on what people selectively attend.

In the present study we test the idea that yet another determinant of people's selective attention is the comparisons they make. We will first review comparison processes and then evidence from Markman and Gentner (1997) showing that people have better recall when they are cued by elements from scenes that are part of structural alignment. Then, by replicating Markman and Gentner (1997) with an eyetracker, we provide an online test of the idea that structural alignment can drive selective attention.

Comparison

The ability to compare is an integral part of human cognition. Category membership is determined by the degree of similarity to category representations (Medin & Schaffer, 1978; Nosofsky, 1984). In problem solving, people find solutions by comparing new problems to previously solved problems (Chi, Feltovich, & Glaser, 1981; Ross, 1987). In episodic memory, probes are compared to memory traces (Hintzman, 1986). In analogy people compare base and target domains. (Falkenhainer, Forbus, & Gentner, 1989; Gentner, 1983; Hummel & Holyoak, 1997)

There have been a few approaches to modeling the comparison process, including computing distances in multidimensional space using feature vectors, (Shepard, 1962) or comparing features using set operations, Tversky (1977). And yet to account for human comparison of complex stimuli with relational structure, a third approach has been used. Borrowing from models of analogy (Falkenhainer et al., 1989), the structure-alignment account (Gentner, 1983) represents objects as features inside structures of relations. For example, structure-alignment theory posits that people will encode features (e.g., the people and objects in Figure 1A) as arguments to relational predicates: smokes(man, cigar) or paints(painter, model). On this account, significant processing is applied to building a representation of the relations between features in a scene or object, and into determining which objects match on the basis of shared roles.

With structure alignment, a great deal more information is both represented and processed than what is proposed by simpler accounts. Rather than comparing sets of features alone, comparisons are made over features and their relations. To accomplish this, objects with the same relational role in both scenes are placed in correspondence, while objects with different roles in their respective scenes are not.

Alignable Differences and Attention

Structure alignment has the ability to represent and calculate similarity over structured representations. However, this ability comes at a processing cost; the alignment process must build structurally consistent matches that satisfy parallel connectivity and one-to-one mapping. Parallel connectivity requires that matching relations have matching arguments. For example, in Figure 1A and 1B, if the photographer is aligned with the painter, then the man with the backpack is aligned with the model. One–to-one mapping states that across representations each object can be aligned to at most one other object—the boy with the backpack cannot also be aligned with the man and the cigar. Thus, the mapping process in structural alignment involves more than simple feature comparisons.

As a result of the more extensive processing involved in structural alignment, three different kinds of output are produced (Markman & Gentner, 1993). Whereas the feature-based approaches distinguish only between commonalities (matching features) and differences (mismatching features), structural alignment produces commonalities on one hand, and two types of differences. Differences that are linked to the commonalities, or *alignable differences*, and those that are not, *nonalignable differences.* For example, the female figure in Figure 1A is an alignable difference with the boy in Figure 1B. However, the man in the chair is a nonalignable difference, since there is no corresponding object in 1B. Thus, instead of just two kinds information used in the similarity calculation, the structural alignment approach has three.

The three types of output allow structure alignment to make the unique prediction that comparisons will focus people's attention on alignable differences. There are two reasons for this. First, it has been shown that people tend to weigh commonalities more heavily than differences in similarity judgments (Tversky, 1977). Since alignable differences are a type of commonality (on the basis of the relational structure) they should receive more attention.

The second reason for additional focus on alignable differences is that the entire alignment process is geared towards building up relational structure. Since alignable differences are what compose that structure, they should receive a significant amount of attention.

Over the last decade there has been a growing amount of evidence that alignable differences in fact receive more weight than nonalignable differences. Markman and Gentner (1996) showed that when given a choice, subjects were more likely to select scenes with nonalignable differences as being more similar to a base scene than scenes with alignable differences. In a second experiment they showed that similarity ratings were more affected by variability in alignable differences than by variability in nonalignable differences. Markman and Gentner (1993) showed that people tend to list more alignable differences than nonalignable differences.

In another demonstration of the importance of alignable differences, Markman and Gentner (1997) had subjects rate the similarity of ten pairs of scenes, like those in Figure 1. Later, subjects were either given probes that were part of an alignable or nonalignable difference, as in Figure 2. They found that on average, subjects recalled 2.35 pieces of information when memory probes were part of an alignable difference versus just 1.3 when the probes were part of a nonalignable difference. Thus, across a range of studies, people seemed to place more weight on alignable differences.

The critical implication of these findings is the idea that structural alignment can be one of the determiners by which people select relevant aspects of their environment. The most direct test of this idea is an online measure of people's selective attention behavior as they make comparisons.

Eyetracking and Selective Attention

It has been well established that eye movements and selective attention are closely linked. For example, Shepard, Findlay, and Hockey (1986) demonstrated that although attending without making corresponding eye movements is possible, it is not possible to make an eye movement without shifting attention. Since high quality visual information is acquired only from a limited spatial region surrounding the fovea, we move our eyes three times each second through high-velocity saccades to position the fovea on what seems important.

It is no surprise then that eye tracking has enjoyed success in numerous research areas that appeal to the construct of selective attention. For example, Rehder and Hoffman (2005a) showed that learning a category corresponded to abrupt shifts in fixations towards relevant information. Later, Rehder and Hoffman (2005b) replicated Medin and Schaffer's (1978) 5-4 category structure with an eye tracker and found that fixation times to stimulus dimensions matched the decisions weight estimated from behavioral responses.

More recently, researchers have begun to leverage the flexibility that eye movement analysis offers in terms of experimental design. It is now possible to examine how attention is allocated across different kinds of tasks (Rehder, Colner, & Hoffman, 2009) and across different stimuli and categories (Blair, Watson, Walshe, & Maj, 2009). The close link between attention and eye movements has been shown across a variety of cognitive tasks (see Liversedge & Findlay, 2000 and Rayner, 1998 for reviews).

Of course, the key advantage to using eye tracking for the present purposes is that it provides an online measure of what people attend to during the comparison process. While recall behavior, verbal protocols, and similarity ratings all point to the conclusion that alignable differences have a greater impact than nonalignable differences on comparison, these are all offline measures. Testing recall performance, for example, occurs well after the comparison process has taken place. Although offline measures can indicate what subjects preferred to encode, they can't tell us about processing dynamics as they unfold over time.

Finally, one of the key claims of structural alignment is that the comparison process can help people determine what information is worth attending to. If in fact alignable differences do not receive more attention than nonalignable differences, then the validity of this claim is called into question. The present study will provide an online test of whether people allocate more attention to alignable features than to nonalignable features.

Experiment

The goal of the present experiment is to use eye tracking as a source of data to measure how comparison processes direct people's attention to important pieces of information, and how that in turn relates to recall of that information. According to the structural-alignment approach, the process of comparison should lead people to attend to alignable over nonalignable differences. As a result of this boost in attention, alignable differences should serve as better cues for recall later on. To test this, we replicated Markman and Gentner (1997), using an eyetracker to monitor subjects' attention allocation. Subjects were fit with a head-mounted eye tracker and we recorded their eye movements to alignable and nonalignable differences as they rated the similarity of ten pairs of scenes.

The main result of interest is whether subjects tend to allocate a greater amount of attention to alignable differences than to nonalignable differences. The structure-alignment approach predicts that subjects' fixation times will be greater on average for alignable differences than for nonalignable differences. Such a finding supports the idea that comparison via structural alignment helps focus people on what's important in the environment.

We will also examine how attention to alignable differences unfolds over the comparison process. Such dynamics will have implications for models of comparison.

Method

Participants Twenty-eight University of Texas students participated for course credit. They were tested individually and assigned to a random order of items. For each item, half of the subjects saw one comparison scene, and half saw the other. At the same time, the assignment of aligned and nonaligned recall cues to each comparison scene was counterbalanced across subjects. This designed allowed us to separate out effects of alignability on attention allocation and memory from any specific object-salience effects, or differences in subjects' ability to recall particular objects from the scenes.

Materials The stimuli in the current study were based on the Markman and Gentner (1997) materials, but were made more suitable for eyetracking by (1) removing unnecessary textures and (2) increasing the distances between objects to more clearly distinguish which were fixated.

Figure 1 shows an example stimulus. As in the original study, there were ten sets of picture triads (one base, and two comparison pictures). The base picture had two relational scenes within it and each comparison picture matched one of the relational scenes. For example, Figure 1A is a base picture. It contains a portrait relation (the artist is painting a portrait of the model on the right), and there is a burning-dropping relation on the left (the man is dropping ash from a lit cigar) on the left. Each comparison matched one of the relational scenes. For example, Figure 1B matches the portrait relation, and Figure 1C matches the base picture on the burning-dropping relation. On a given trial, the base scene and (one of the) comparison scenes are presented together on screen. Later, one object from each relational structure in the base scene was used as a recall cue. For example, as shown in Figure 2, the painter and the man in the chair from Figure lA were used as recall cues.

The eye tracker was an SMI Eyelink II, which was set to track one eye at 250 Hz.

Procedure Subjects were first fitted and calibrated to the eye tracker. Items (i.e., a pairing of a base and one comparison scene) appeared on the screen. At their own pace, subjects rated the similarity of the base picture to the comparison picture (on a 1-to-9 scale). Before each item

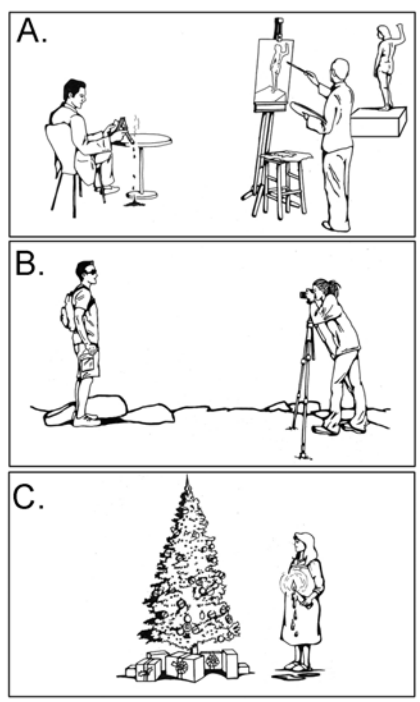

Figure 1. Example stimuli. Panel A is the base picture. Panels B and C are the two comparison scenes.

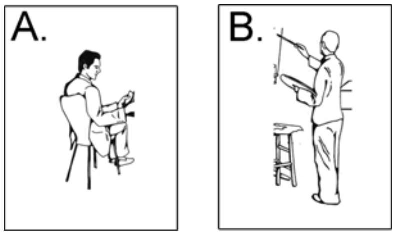

Figure 2. Two example recall cues. Depending on the comparison picture, either cue can be an alignable or nonalignable difference.

presentation subjects were asked to fixate a small circle in the center of the monitor. This was used both as a drift correction and as an indication that they were ready for the next trial. Subjects recorded their rating by typing one of the corresponding numbers keys on the keyboard.

After subjects provided the ten ratings they engaged in a reading task for 30 minutes.

During the recall phase subjects were presented with one of the recall cues. Half of the recall cues were from alignable differences and the other half were from nonalignable differences. Subjects' verbal responses were recorded by a computer microphone.

Results

Recall We first set out to test whether we replicated the basic finding from the original Markman and Gentner (1997) study that alignable cues yield better recall than nonalignable cues with the revised stimuli. Therefore, we examined the effect that alignability had on subjects' recall of the scenes, by counting the number of pieces of information recalled from the base scene as a function of whether they received an alignable or nonalignable cue during recall. The data were first transcribed from the voice recordings and then rated by a single rater. The instructions to the rater were that each proposition (adjective, noun, or verb) about the scene counted as a piece of information.

The average number of correctly recalled pieces of information for the alignable cues (M = 1.8, SD = 1.2) was reliably greater than the number of pieces of information recalled for the nonalignable cues (M = 1.3, SD = 0.92), $t(27) = 2.44$, $p < .05$. The analysis was also carried out by item, and the result was marginally reliable $t(19) = 1.84$, $p = .081$. Thus, the basic findings found by Markman and Gentner were replicated here.

Fixations For our initial analysis, we constructed heat maps of eye fixations to get a sense for where people were looking while judging picture similarity. Figure 3 shows heat maps of fixations superimposed over one of the items, with both comparison scenes. To construct these heat maps, each x-y coordinate of the fixations were weighted by their total fixation time and summed over all subjects for each item. The weighted fixation coordinates were then processed by a Gaussian kernel density estimator, with bandwidth estimation (Jones, Oliphant, & Peterson, 2001). The red spots of the heat map reflect greater average amounts of fixation time, and as a result, where subjects were attending. Overall, and as expected, in both panels of Figure 3 fixations were centered directly over the objects in the scenes. However, the heat maps also show that the allocation of attention is very different depending on which comparison scene the subjects saw.

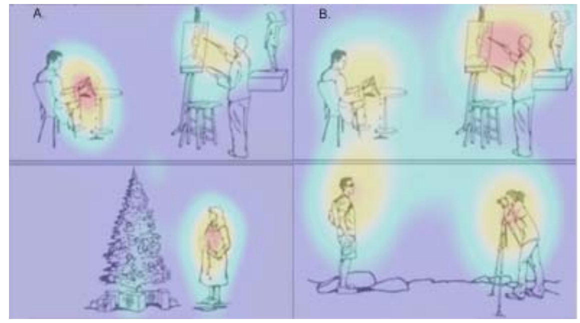

Figure 3. Example heatmap of fixations to an item, averaged over subjects for one of the ten items, with comparison scenes. In Panel A there are more fixations to the man smoking, but in Panel B, there are more fixations to the painter and model.

According to structure-alignment theory, more fixations should land near the objects that align with the comparison picture. For example, the comparison scene in Figure 3A aligns with the man smoking in the left half of the base picture whereas the comparison scene of Figure 3B aligns with the portrait relation on the right hand side of the base image. In fact, the heat maps in Figure 3 show the result predicted by structural alignment. There are more intense and concentrated hot spots over the man in the chair in Figure 3A, and lesser hot spots over the painter and the model. The reverse is true for Figure 3B, there are more intense hot spots over the painter and the model, and weaker hot spots over the man in the chair. The heat map presented in Figure 3 provides a clear illustration of how subjects allocate greater attention allocation to alignable differences in the scene.

Next, we extended the above analysis to all items. For this purpose we coded fixations according whether they were to an alignable difference or to a nonalignable difference in the base picture. We then computed the total fixation time for alignable differences across all items, for each subject. The average total fixation time to alignable differences (M = 1473, SD = 814) was greater than that for nonalignable differences (M = 1272, SD = 690), $t(27) = 2.25$, $p < .05$. (Although item analysis was not statistically reliable $t(19) = 1.2$, $p = .28$., seven out of ten of the items showed the effect in the expected direction). Thus, as structure-alignment predicts, subjects allocated more fixation time to alignable differences as compared to nonalignable differences.

The above results showed that overall, the comparison process engaged by subjects in determining the similarity of two images caused them to fixate alignable differences over nonalignable differences. But how does the comparison process direct attention to important features in a scene, and at what point are people drawn to alignable differences? Figure 4 shows the probability of fixating alignable differences, nonalignable differences, and to the comparison scene as a function of time, for ten seconds of the trial.

To construct Figure 4 we determined, for each 50-ms interval, whether a subject was fixating one of those three

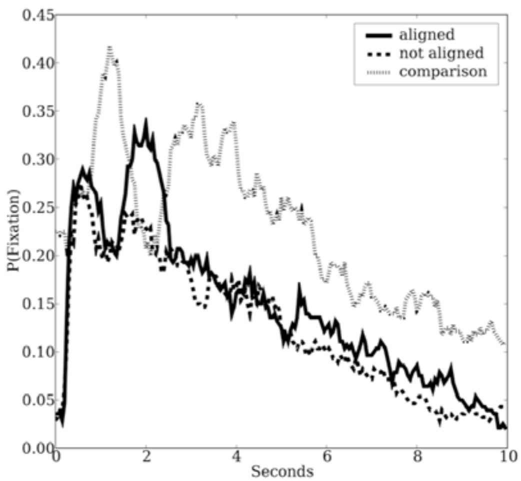

Figure 4. Probability of fixating the aligned, nonaligned, and comparison objects in the scene, as a function of time (seconds).

locations. We then averaged over all trials and subjects to examine attention allocation over the course of the trial.

The figure shows that as expected, subjects showed no immediate preference for the alignable or nonalignable differences in the base scene. (The initial preference for the comparison scene in the first 50 ms reflects that the comparison scene was a much larger area of interest than the individual alignable and nonalignable differences, and there's a greater baseline chance that eye fixations will happen to be there first.)

Figure 4 then shows that during the next second, there was a dramatic increase in fixations to all three locations, but especially to the comparison scene. In fact, after one second, there is a sudden decrease in fixations to the alignable and nonalignable differences. Fixations then shift from the comparison scene to the alignable differences in the base picture, until fixations to alignable differences peak, at around the two-second mark. After this, fixations gradually dropped off for all locations (as more and more subjects have already responded), with the most fixations allocated to the comparison scene. On average, subjects did not allocate more fixations to the nonalignable differences at any point in the trial.

Discussion

Markman and Gentner's (1997) result that people have greater recall performance when cues are part of alignable differences replicated in the present study. These results were consistent with other previous work showing that alignable differences have a greater impact than nonalignable differences on people's comparison behavior.

The main contribution here was that we were able to observe the structural alignment process online. The predictions for the eyetracking results, that more fixations should be allocated to the alignable features obtained. The unfolding of attention allocation over the course of the comparison process also appeared to make sense. As soon as subjects allocated a significant amount of attention to the comparison and base scenes, attention was allocated to the alignable differences, as predicted.

Our results have clear implications for cognitive models. First, mechanisms of comparison need to represent relational structure to explain selective attention behavior towards stimuli with any high level of complexity. Standard models in category learning that contain geometric (Kruschke, 1992) or feature-based (Lee & Navarro, 2002) similarity metrics need to be modified to account for people's ability to represent and attend to relational semantics.

Models that already have the ability to represent relations are consistent with the eye tracking results from the present study. For example, Hummel and Holyoak's, (1997; 2003) LISA and Larkey and Love's (2003) CAB models look for surface-feature similarities between items and only later try to match lower- and higher-order relations. Such mapping patterns reflect the selective attention behavior of our subjects because subjects required two seconds on average to focus primarily on alignable differences.

That subjects in our experiment attended differentially to objects according to their placement in the relational structure provides a proof of concept for using eye movements for more detailed tests of computational models, including those that already have the ability to represent relational structure. Additional eyetracking data can be collected to constrain the various components, for example, by having people make comparisons over objects that with different levels of relations (e.g., higher order versus lower order), or by manipulating subjects working memory, models' changes in selective attention can be related changes in selective attention to humans directly.

One of the most interesting implications for our results is derived from considering the working memory constraints of models like CAB and LISA. Working memory functions in such models to constrain the types of relations considered. With less working memory only lower-order relations or superficial feature matches will be represented by the model. This predicts that the details of the relational structure that people can maintain will also be influenced by working memory constraints. As a result, another potential determiner of what people selectively attend to in a scene is their working memory. If their working memory is compromised, they will not be able to use relational structure to guide their selective attention. Thus, the present data provide clear predictions for future eyetracking studies.

The rich source of data provided by eyetracking was able to confirm predictions of structural alignment and shows promise for constraining and developing more detailed processing accounts of existing computational models of comparison. In addition, there are potential future directions for empirical studies that follow from the present work to explain how it is that people decide what to selectively attend in an information-rich world.

References

Blair, M. R., Watson, M. R., Walshe, R.C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies on dynamic attentional allocation to stimulus features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1196-1206.

Bransford, J.D., & Johnson, M.K. (1972). Contextual prerequisites for understanding some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Bransford, J.D., & Johnson, M.K. (1973). Considerations of some problems of comprehensioin In W.G. Chase (Eds.), *Visual information processing* (pp. 383-438).

Brewer, W. F. , & Dupree, D. A. (1983). Use of plan schemata in the recall and recognition of goal-directed actions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 117–129.

Buswell, G. T. (1935). How people look at pictures: A study of the psychology of perception in art.Chicago: University of Chicago Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Falkenhainer, B., Forbus, K., & Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence, 41,* 1-63.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science,* ***7*,** 155-170

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Hayhoe, M., Shrivastava, A., Mruczek, R., & Pelz, J.B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*, 46-63.

Henderson, J, M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50, 243-271.*

Henderson, J, M., Weeks, P.A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during scene viewing. *JEP: Human Perception and Performance, 25,* 210-228.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411-428.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.

Jones, E., Oliphant, T., Peterson, P. and others (2001). *SciPy*. http://www.scipy.org/SciPy.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44.

Larkey, L.B., & Love, B.C. (2003). CAB: Connectionist Analogy Builder. *Cognitive Science, 27,* 781–794.

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Science, 4*, 6-14.

Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics,2,* 547-552.

Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, *21*, 373–418.

Markman, A. B. & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journalof Memory and Language*, *32*, 517–535.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and Cognition*, *24*, 235–249.

Markman, A. B., & Gentner, D. (1997). The effects of alignabilty on memory. *Psychological Science, 8*, 363-367.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 10, 104-114.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*, 107–123.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, 51, 1-41.

Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 811-829.

Rehder, B., Colner, R.M., & Hoffman, A.B. (2009). Feature inference learning and eyetracking. *Journal of Memory & Language, 60*, 394-419

Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Erlbaum.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *The Quarterly Journal of Experimental Psychology, 38*, 475-491.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function, I. *Psychometrika, 27,* 125-140.