The role of transformations and structure in the same-different paradigm

Carl J. Hodgetts (hodgettscj@cf.ac.uk)
School of Psychology, Cardiff University, Park Place,
Cardiff, CF10 3AT, UK

Ulrike Hahn (hahnu@cf.ac.uk)
School of Psychology, Cardiff University, Park Place,
Cardiff, CF10 3AT, UK

Nick Chater (n.chater@ucl.ac.uk)
Cognitive, Perceptual and Brain Sciences, UCL,
London, WC1e 6BT UK

Abstract

This paper investigates the role of transformations and similarity in a perceptual task, the *same-different paradigm*. Representational Distortion (RD) theory measures similarity by the complexity required to 'distort' compared object representations. In an experiment, participants compared pairs of geometric shapes that varied across two dimensions (shape and color). We then modelled data using a variety of model of similarity. RD yielded accurate fits and compared favourably with various models of Structural Alignment. Results highlighted the relationship between transformations and low-level stimulus properties.

Keywords: similarity; transformations; complexity; alignment; representation.

Introduction

Acting intelligently in our environment will often involve similarity. This may be in determining whether an object is a particular sort or not, thus, allowing us to interact appropriately with it; or inferring that a particular property of an object will be shared by other objects that are similar to it. Broadly speaking, similarity between past instances and novel experiences helps render the mass of information inherent to the world more coherent and manageable. The significance of similarity in cognitive psychology and perceptual theory is reflected in the many phenomena that assume a central role for similarity: categorization, induction, linguistic knowledge, memory retrieval and object recognition, and more.

Traditionally, similarity theory has been dominated by two opposing accounts: the spatial account and the featural account. On the spatial account (Shepard, 1957), objects are represented as points within a multidimensional space. The distance between objects in this internal, psychological space reflects their similarity. Alternatively, the contrast model (or featural account), proposed by Tversky (1977), conceptualizes objects as mentally represented feature-sets. Here, similarity is a function of the common and distinctive features possessed by the comparison objects.

Despite empirical support, particularly in similarity-based models of categorization (Nosofsky, 1986; Osherson, 1989), these traditional models have fundamental limitations: the simple and very specific representations they posit make them seem incompatible with the inherent structure of real world objects. Characterizing objects and scenes by feature sets or dimensions may provide some basic level of description but will fundamentally underplay the relations between these attributes, that is, their structural properties (Biederman, 1985; Gentner, 1983, 1989; Hahn, Chater & Richardson, 2003; Markman and Gentner, 1993a). This limitation has given rise to a number of novel similarity accounts that are able to tolerate a wider range of representations, including *structured* representations. In the current paper we investigate one such account, Representational Distortion (henceforth RD; or the Transformational Approach).

RD suggests that similarity between two objects is best understood in terms of the complexity of 'transforming' or 'distorting' our representation of one into our representation of the other (Hahn et al. 2003). Objects that are similar will require simpler transformations, objects that are dissimilar will require complex ones. In experimental contexts, transformational complexity has typically been operationalized simply by the number of instructions required to complete the overall transformation (Hahn et al. 2003; Hodgetts et al. 2009).

In support of the account, Hahn et al. (2003) found that transformation distance predicted similarity ratings across a range of materials (dot patterns, simple geometric shapes and Lego bricks). Hodgetts et al. (2009) presented participants with objects that consisted of a pair of shapes that varied on one dimension or two dimensions (color and/or shape; see Figure 1).

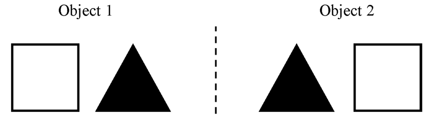

Figure 1: An example of the stimuli used by Hodgetts et al. (2009).

The set of transformations governing these materials, posited by Hodgetts et al., accurately predicted similarity ratings (r = -0.86) and forced choice responding (r = -0.95). Furthermore, these fits compared favorably to competitor models: Goldstone's (1994) Similarity Interactive Activation and Mapping model (SIAM) and Gentner's (1989) Structure Mapping Engine (SME). Finally, Hahn, Close and Graf (2009) manipulated transformation direction and found attendant effects on ratings of similarity. They presented participants with short animations of one object morphing into another. Subsequently, similarity ratings for identical comparisons of objects drawn from this morph continuum were higher when the direction of the comparison (i.e., "how similar is A to B?" vs. "how similar is B to A?") was congruent with the direction of the preceding animation than when it was in opposition to it.

This evidence for RD has been based wholly on explicit similarity judgments (direct ratings or forced choice). Here, we provide the first test of RD on an implicit similarity task. Both explicit measures such as ratings, and implicit measures, such as confusability, reaction times, or matching errors have their benefits. While ratings offer, in quite unambiguous terms, an individual's subjective assessment of similarity, they less readily tap into similarity assessment as a process, and it is unclear how results obtained with ratings extend to similarity as it functions in a wide range of cognitive tasks such as categorization, inductive reasoning and so on. Consequently, it is important to determine whether there is a place for structure and transformations above and beyond direct measures of similarity.

Parallels between similarity research and perceptual theory provide some grounds for believing that there is. For example, the speed and ease of object recognition has been commonly associated with the transformational relationships between objects (Bundesen & Larsen, 1975; Graf, 2007; Tarr & Pinker, 1989). As we navigate our environment, the pattern of retinal stimulation undergoes constant transformation. Under these circumstances we manage to maintain object constancy regardless of such changes, suggesting that our visual system has mechanisms and representations attuned to visual transformations (e.g., rotation, dilation). As similarity in RD also refers to a distance from an original object's identity (particularly in a directional similarity comparison), this parallel seems both relevant and justified. Studies into apparent motion have also provided evidence that, along with spatio-temporal proximity parameters, similarity and transformation distance are key factors in facilitating motion correspondence (Bundesen, Larsen & Farrell, 1983; Farrell, 1983; Shepard & Judd, 1976).

However, it would be desirable to manipulate structure-based transformations more directly. To this end, we used a speeded same-different (or 'perceptual matching') task for assessing similarity implicitly. Previous studies that used the same-different task have observed response patterns that correspond to underlying object similarities. For example, the speed of a *different* judgment is considered to correspond to the underlying similarity between compared items in the sense that a fast different response indicates low similarity and vice versa (Cohen & Nosofsky, 2000).

Despite the prominent similarity-element of this paradigm, the relationship between specific similarity models and same-different performance has been relatively under-explored (for exemplar-retrieval see Cohen & Nosofsky, 2000; for a theoretical contrast see Frost & Gati, 1989; Goldstone & Medin, 1994). The study most relevant to the present paper was conducted by Goldstone and Medin (1994) who used a same-different task to test the dynamic, time-course characteristics of the similarity model SIAM. SIAM belongs to the class of Structural Alignment models (henceforth SA; Gentner, 1983; Markman & Gentner, 1993a, 1993b). The SA framework assumes that similarity is determined in a manner akin to analogical mapping where features *and* relations are placed into correspondence. Specifically, SIAM calculates similarity by aligning features, objects and relations in one scene with those in another through a process of interactive activation. There are two kinds of matches in SIAM that dynamically govern the model's behavior in this context: 1) matches in place (MIPs) and 2) matches out of place (MOPs). A match in place is a feature match between objects that correspond, whereas a MOP is a feature match between objects that do not correspond. Optimally, and with sufficient time, SIAM will make correspondences that maximize the number of MIPs, that is, make correspondences that are globally consistent with other correspondences. Initially, however, MIPs and MOPs are equally salient meaning that locally consistent matches will strongly influence similarity at short deadlines; over time MIPs will grow in salience and principally determine similarity.

Goldstone and Medin's (1994) data supported SIAM's predictions for thirteen stimuli that varied along four dimensions. Specifically, MIPs had an only marginally larger influence on similarity than MOPS at shorter deadlines but a much larger influence at longer deadlines. This provides the first evidence for a structural model of similarity using a speeded same-difference task.

In the present study, we used the same basic paradigm to test RD.

Experiment

For our experiment, we used the stimulus set and coding scheme of Hodgetts et al. (2009; for example see Figure 1). These stimuli have a long tradition as a tool for developmental, and comparative (non-human animal) research into feature binding and the representation of structure (Cheries, Newman, Santos & Scholl, 2006; Kaldy & Leslie, 2003; Larkey & Markman, 2005). Specifically, comparisons within this domain consist of two pairs of geometric shapes that can vary on two dimensions (shape and color).

For each comparison, features can be arranged in 14 ways on each dimension and combined to create all possible comparisons (resulting in 196 -14 x 14- possible comparison pairs).

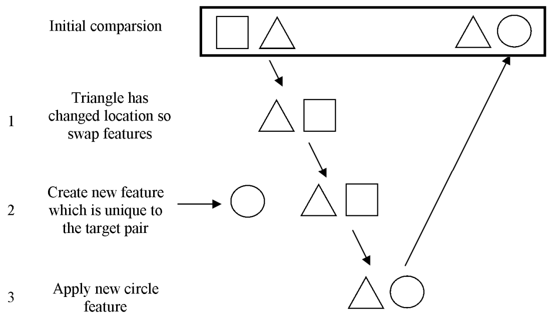

Figure 2: An example of transformations being carried out. This comparison has a transformation distance of 3.

As in Hodgetts et al. (2009), we used a random subset of all possible combinations (78 comparisons). This should provide a sufficient estimate of the entire population. The three transformations that govern similarity in this domain take the base pair and modify it as follows[1]: 1) *Create* – taking the base pair we apply this operation to create a new feature that is unique to the target pair; 2) *Apply* – this operation takes an object or entity that is currently available (by being present in the base or by having been created via step (1) and applies it to *one or both* of the objects in the target pair. 3) *Swap* – this swaps features between a pair of objects *or* swaps the object in its entirety (i.e. on both dimensions).

Figure 2 provides a schematic demonstration of this coding scheme. As stated above, these three operations have provided compelling fits on two explicit measures. Furthermore, superior fits were found relative to SIAM and SME, without the need to resort to free parameters (for a detailed account of this coding language see Hodgetts et al. 2009).

In the experiment, participants took part in a sequential same-different task whereby the entire random subset of 78 comparisons made up the different trials. Unlike Goldstone and Medin (1994) we did not impose response deadlines; instead participants were urged to respond as quickly as possible. Primarily, this experiment will indicate whether the coding language, that has performed well with ratings data, will successfully extend to an implicit measure of similarity i.e., response-time.

Participants 30 participants took part in the experiment.
Materials Trials were sequentially presented on a 19" LCD monitor with a refresh rate of 60 Hz. The shapes were created using the AutoShape function on Microsoft Publisher. Each shape was 2.5cm wide x 2.5cm tall. Shapes within a pair were separated by a horizontal distance of 0.5cm. The screen location of pairs on a given trial was determined by randomly combining predetermined values on each screen axis (i.e., 10, 20, 30, 40, 50, 60, 70, 80 and 90). The stimulus duration for a given pair was 833ms (50 frames) with an Inter-stimulus Interval (ISI) of 16ms (1 frame). A response could be given at the onset of the second stimulus pair.

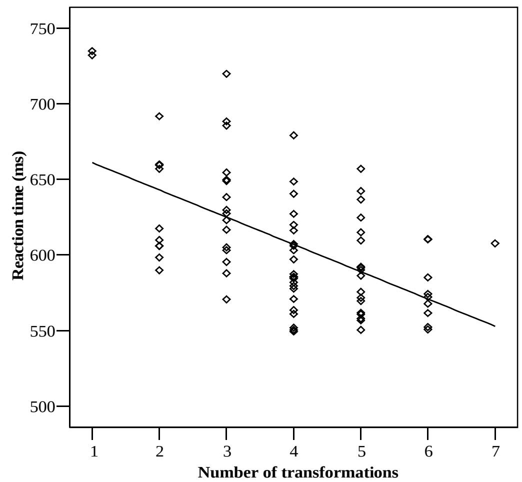

Figure 3: Graph depicting the relationship between transformation distance and reaction time (r = -0.55).

All stimuli were presented in both directions (i.e., each participant saw both object 1 in Figure 1 followed by object 2, and vice versa) resulting in 156 different trials. The 'same trials' were generated by pairing each composite pair with itself resulting in 128 same trials. Trial order was randomized. Participants could indicate 'same' by pressing Z on the keyboard and 'different' by pressing M. A 'different' response was given if pairs differed on a single dimension or on both. After a response was given, the screen was erased and a new pair was randomly selected for the next trial. No response deadline was imposed but participants were urged to respond as quickly as possible.

Results

For analysis, we looked at the correct responses for the different trials, averaged across the two directions. Reaction times three standard deviations above and below the overall mean were removed. We then correlated reaction time with transformation distance. As similarity is a decreasing function of RT, RTs should decrease with increasing code length, as longer codes reflect greater dissimilarity.

The graph in Figure 3 depicts the expected negative relationship between reaction time and transformation distance. A bivariate correlation between the number of transformations and reaction time for the *different* trials was found to be significant using Pearson's r (r = - 0.55, p <0.01). Without free parameters, transformational distance, as specified by the coding scheme, accounted for 31% of the variance in reaction time for *different* responses (R^2 = 0.31).

Comparing models of structural alignment

To correspond with Hodgetts et al. (2009), we also modeled this data using two models of SA: SIAM (Goldstone,

[1] In a directional similarity comparison (i.e., 'how similar is *B* to *A*' as opposed to '*A* and *B* are similar'), the term 'base' refers to the referent object or *A*.

1994b) and SME (Gentner, 1983). Previous tests of SIAM and SME indicate moderate to good fits for similar stimuli in ratings tasks (Larkey & Markman, 2005).

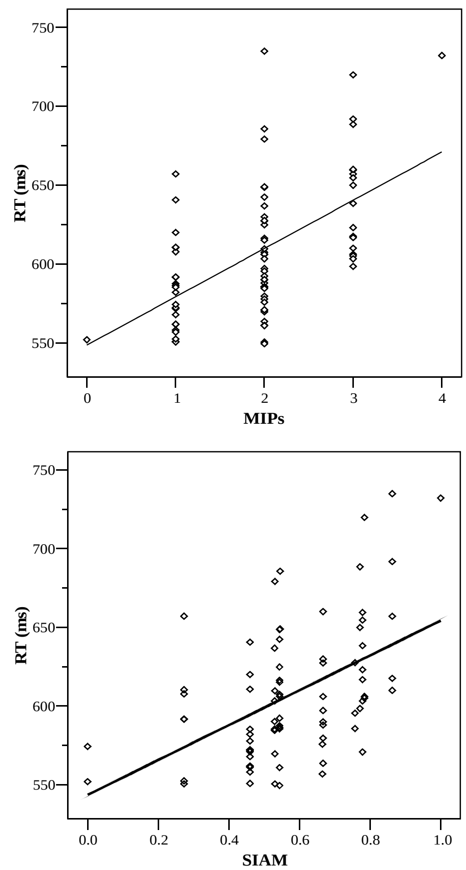

Figure 4: a) The relationship between the number of MIPs and reaction time and b) reaction time and SIAM.

For SME we correlated the number of matches in place (MIPs) with the similarity data[2]. The graph in Figure 4 shows the relationship between RT and the number of MIPs. A positive relationship is clearly evident and this relationship is born out statistically ($r = 0.55$, $p < 0.05$; Pearson's r). The variance accounted for by MIPs alone is 31% ($R^2 = 0.305$). The *r* value and the accounted variance are identical to RD for a basic MIP counting approximation of SME.

As in previous experiments, SIAM was modeled using its default parameters. Contrary to Hodgetts et al. (2009) and Larkey and Markman (2005), SIAM provides a poorer account of the data than SME (a more constrained SA model). The graph in Figure 4 illustrates the significant positive relationship between SIAM's predictions and RT for correct *different* trials ($r = 0.49$, $p < 0.01$; Pearson's r). With an R^2 value of 0.24, it also fails to fit the data as well as RD. It is possible that SIAM could achieve comparable fits with different parameter settings but such a gain would likely be offset when the number of free parameters is taken into account.

[2] The number of MIPs is an approximation of SME performance (as used by Larkey & Markman, 2005). This method is understandable given the model's strict adherence to the one-to-one constraint. This, however, will slightly underestimate the influence of MOPs in the similarity computation.

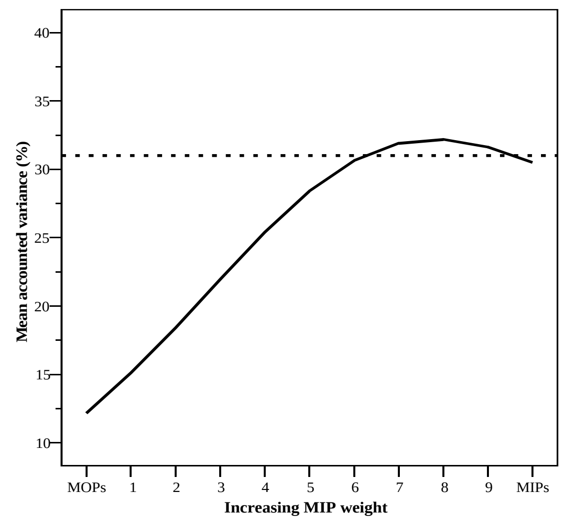

Figure 5: Variance accounted for by MIPs/MOPs when they are weighted differentially. Increments on x-axis correspond to percentage weight increases in steps of 10, i.e., tick mark '3' corresponds to a weighted average that is 30% MOP/70% MIP. The dotted line represents the variance accounted for by RD.

MIP/MOP weighting

SIAM's model outputs are optimal alignments that reflect behavior in the absence of time pressure and therefore may underestimate the relative effects of MIPs and MOPs (these are outputs after 100 cycles, once the model has settled). SIAM, when response deadlines are considered, differentially weights MIPs and MOPs over the time course, with MOPs having a comparable influence at shorter deadlines when compared with MIPs (Goldstone & Medin, 1994). In essence, this is achieved by running fewer cycles and stopping the model earlier in the computation when MOPs have a greater influence on similarity. Even though time course is not manipulated here, it seemed valuable in this comparison to see how, in general, *any* MIP/MOP model would fare against these data. Figure 6 demonstrates the variance accounted for by a model that counts differentially weighted MIP and MOPs, considered over the entire range of possible weightings between the two. The marker on the *y*-axis signifies the variance accounted for by transformation distance alone. Based on our earlier approximation, SME corresponds to the far right of Figure 5, that is, the region where MIPs alone influence the comparison.

As seen in the graph, differentially weighting MIPs and MOPs has a profound effect on overall fit. For this task, the optimal fit is where MIPs are weighted at 80% and MOPs are weighted at 20%. At this point, the weighted MIP/MOP model exhibits a slightly better fit than that provided by RD. Again, however, model complexity needs to be factored into the comparative evaluation, given that the weighted MIP/MOP model has one free parameter.

Discussion

The results of the current experiment provide moderate support for the notion that similarity is determined by the complexity to transform object representations. The RD model, based on three simple operations, achieved significant data fits when correlated with reaction time. Furthermore RD managed to account for 31% of the variance. When compared with models of SA, RD compares favorably but not superiorly; RD and SME fitted the data equally well. This contrasts with the findings of Hodgetts et al. (2009) who observed vastly superior data fits for RD compared to SME for both similarity ratings and forced choice tasks.

To date, evidence for RD has solely been based on direct similarity measures; whilst these provide important insight into transformations, it is crucial that these effects extend beyond direct measures, particularly given the prominence of transformations in the perception literature (though we do not assume that subjective assessments are poor reflections of perceived similarity; they merely represent one out of a number of possible similarity measures - for comparisons of both implicit and explicit measures, see Desmarais & Dixon, 2005). As implied above, the significant fits for RD are consistent with vision research that has long assumed a central role to transformations in the speed and ease of object recognition (see Graf, 2007). Given the evidence that the visual system is sensitive to transformational relationships, it seems plausible that the cognitive system may share this sensitivity when dealing with object identities via similarity comparisons. The conclusions are, however, necessarily tentative until the matching fits of SME and RD are disentangled.

The fact that SME performs comparably is an issue that must be addressed. Even though MIP counting is commonly used a measure of SME (largely because of its simplicity), it underestimates the influence of MOPs. Therefore, a more complex version of SME could yield different fits, for better or for worse. The graph in Figure 5 emphasizes the complexity of this issue; differential weightings of MIPs and MOPs profoundly affect model performance. In one area of the parameter space, MIP/MOPs do better than RD, indicating that a weighted instantiation of SA could yield superior fits; although none of the formal models do so. However, it does go in RD's favor that it provides accurate fits without free parameters.

Although no response deadlines are imposed, there is still pressure to respond quickly. For SA, certain weightings may well simulate responding under certain time course conditions. For example, the 0.2 weighting of MOPs may suit the time conditions of this task specifically, i.e., non-optimal but without time pressure. RD, in this context, has no time course prediction in that the coding language only provides a static, endstate prediction, albeit a fairly accurate one. However, identifying transformations clearly has a time course associated with it.

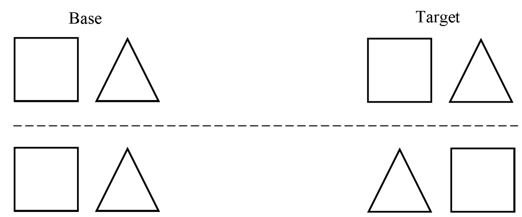

Figure 6: Identical vs. swapped case.

Crucially, for a given transformation, all the features to-be-transformed must be initially identified. If task parameters, or participant factors, do not allow for the sufficient processing of 'what' and 'where' information of both objects, then certain transformational relationships may not be recognized at all. MOPs, being non-optimal alignments, will be equally salient to MIPs when time pressure is high. Participants are also likely to quicken responding given the fast rate of stimulus presentation. Development of a fully specified process model to supplement RD as a similarity metric consequently seems an interesting question for future research.

Considered purely at the level of similarity metric, counting MIPS, however, also has some fundamental shortcomings. MIPs, in this domain, are not assessed with reference to location, just shape and color[3]. This gives rise to a number of counter-intuitive predictions where the relational attributes are manipulated but the similarity, according to MIPs, is preserved. This is epitomized by the two comparisons in Figure 6. According to SME, these two comparisons share an equal number of MIPs (2), as the matching squares and triangles align in each case. For RD, one involves a swap transformation and is therefore less similar. Therefore, these swapped cases should be associated with faster reaction times, as they are less similar under a transformational model.

In Figure 7 we present a bar graph highlighting these problematic items for SME. The left hand bar shows the mean reaction time for an item that RD predicts to be highly similar (AB/AB, color; AB/BA, shape) as only one transformation is required to change one into the other (i.e., swap (shape)). SME, however, predicts only moderate similarity. The remaining two bars show reaction times for items where RD and SME make matching predictions (middle bar = both predict moderate similarity; right bar = both predict low similarity). The graph highlights SME's problems with these swapped cases. In other words, even though the performance of SME and RD were quantitatively

[3] The version of assessing MIPS and MOPS used here follows Larkey & Markman (2005). We also considered a MIP model that referred to structure i.e., features were only aligned if they occupied the same relative position in a pair. This variation fared worse than the model used here.

equal overall, there do seem to be systematic mismatches between SME's predictions and the data.

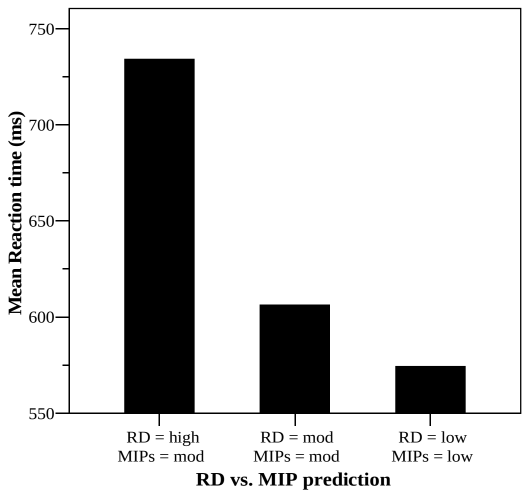

Figure 7: Graph showing RTs for three items when MIPs and RD make different predictions (left bar) and matched predictions (middle/right bar).Finally, this particular evidence for the role of swaps suggests that structural transformations are being represented, even in this perceptual-matching task.

Summary

In this paper, we have presented promising evidence for a transformational account of similarity. Although model accuracy was equal to SME, significant fits were still found in an implicit task. This provides an important extension to prior research that has consistently related transformation complexity and similarity for direct similarity judgments.

Acknowledgments

This research was supported by European Commission grant 51652 (NEST). Nick Chater was supported by a Leverhulme Major Research Fellowship.

References

Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing, 32,* 29–73.

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 214-220.

Bundesen, C., Larsen, A., & Farrell, J. E. (1983). Visual apparent movement: Transformations of size and orientation. *Perception, 12,* 549-558.

Cheries, E. W., Newman, G. E., Santos, L. R., & Scholl, B. J. (2006). Units of visual individuation in Rhesus Macaques: Objects or unbound features? *Perception, 35(8)*, 1057 - 1071.

Cohen, A. L., & Nosofsky, R. M. (2000). An exemplar-retrieval model of speeded same-different judgments. *Journal of Experimental Psychology: Human Perception & Performance, 26(5),* 1549-1569.

Desmarais, G., & Dixon, M.J. (2005). Understanding the structural determinants of object confusion in memory: An assessment of psychophysical approaches to estimating visual similarity. *Perception and Psychophysics*, *67*, 980–996.

Frost, R., & Gati, I. (1989). Comparison of the geometric and the contrast models of similarity by presentation of visual stimuli to the left and the right visual fields. *Brain & Cognition*, 9, 1-15.

Gentner, D. (*1983*). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7* (2), 155-170.

Goldstone, R. L. (1994a). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 3–28.

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 29-50.

Graf, M. (2007). Coordinate Transformations in Object Recognition. *Psychological Bulletin, 132*, 920-945.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition, 87,* 1–32.

Hahn, U., Close. J., & Graf. M. (in press). Transformation Direction Influences Shape Similarity Judgments. Psychological Science.

Hodgetts, C. J., Hahn, U., & Chater, N. (2009). Transformation and alignment in similarity. Submitted.

Kaldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: integrating 'what' and 'where' information. *Developmental Science, 6,* 360-373.

Larkey, L. B., & Markman, A. B. (2005). Processes of similarity judgment. *Cognitive Science, 29*, 1061-1076.

Farrell, J. E. (1983). Visual transformations underlying apparent movement. *Perception and Psychophysics*, *1*, 85-92.

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

Markman, A. B., & Gentner, D. (1993b). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.

Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.

Shepard, R. N., & Judd, S. A. (1976). Perceptual illusion of rotation of three-dimensional objects. *Science*, *191*, 952-954.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Science*, *21*, 233-282.

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