Comparison, Categorization, and Perceptual Dimensions: A Connectionist Model of the Development of the Notion of Sameness

Michael Gasser and Linda B. Smith

1 Comparison and Categorization

1.1 Comparison and Cognition

The process of comparison is fundamental to behaving organisms. Generalization from past to present experience involves a measure of the similarity of present perceptual input to what has been perceived before. The likelihood that we call some object a *dog* is a function of the perceptual similarity of that candidate object to other objects known to be *dogs*. But humans do far more than simply categorize objects into a single basic level. At various times and for various purposes, we compare objects in terms of their overall similarity to one another and along a wide array of perceptual dimensions. For example, we might judge a dog to be brown, to be the same color as our cat, to be large for dogs in general. Indeed, what we consider higher mental functioning—metaphor, poetry, science itself—involves pointing to and discovering novel kinds of similarity.

The problem of how a child develops a system of multiple kinds of perceptual similarity together with devices for linguistically communicating about similarity is clearly of great importance to cognitive science. This is an area in which there is rich and detailed data about human development, but no current theory that adequately explains it.

In this paper we describe a connectionist model of some of the basic facts of comparison along perceptual dimensions. The workings of the model are based on the idea that categorization along dimensions (*what color is it?*) and comparison along dimensions (*is X the same color as Y?*) make use of the same dimension representations and the same internal representations for objects. We propose that these representations develop in response to the demands of the two kinds of tasks. Our model should yield behavior which qualitatively fits the developmental data from children.

1.2 Implicit Comparison and Categorization

Generalization means recording the similarities among things that are relevant for a particular type of behavior. Making use of a generalization means comparing a stimulus to representations of previously encountered stimuli to generate an appropriate response. A simple pattern associator performs this implicit comparison through the connection weights that make up its long-term memory. We shall call this basic implicit form of comparison **categorization**. It consists of a comparison between a stimulus in short-term memory and others in long-term memory. Categorization can either be in terms of complex categories such as DOG and CHAIR or in terms of dimensional attributes such as RED and BIG.

1.3 Explicit Comparison

But the comparison between a stimulus in short-term memory and a host of others in long-term memory may be quite different from the comparison of two items in short-term memory, a process which, if we take the evidence from language seriously, goes on often in human cognition. A sentence such as *this marble is the same color as that one* requires speaker and hearer to maintain representations of both objects in short-term memory, where they can be compared. Language gives people the capacity to make use of explicit comparison of this sort.

Explicit comparison involves looking for and pointing out similarities or differences between the compared entities. In this paper we will only be concerned with that subtype of explicit comparison which is usually signalled in English by the word *same* along with a dimension noun such as *size*, and within this category we will only deal with comparison along simple perceptual dimensions such as color and size.

Given an abstract comparison device which looks for symmetry in its two input patterns, what is required to have it make judgements of “same thing,” “same color,” and “same size”? Such a device cannot compare the two objects themselves; it must have access to representations of the objects. Moreover, to make same-color and same-size comparisons, the device must have access to representations in which only the relevant dimension manifests itself. Irrelevant dimensions need somehow to be “filtered” out.

1.4 Developmental Facts

There is a well-documented trend in the development of object and dimensional comparisons and object and dimensional language:

1. Early object categorizations are principally across all dimensions at once (see [Smi89b, Smi89a] for reviews of relevant literature). The sensory features are somehow compressed into a single representation in which all constituents are weighted more or less equally.

2. The comparison of objects by overall similarity—the judgement that two identical cups are alike in the same way as two identical dogs—also appears very early in development [Smi84]. Moreover, children comment on the overall similarity (or absolute identity) of objects by 24 months through such devices as iterative naming (*dog, dog*), counting, and use of the plural [Sug83]. These facts suggest that there is a comparison component that operates early and that is independent of specific perceptual properties. That is, the formation of a productive rule of the sort “when there are two objects that are the same use the plural” requires a sameness judgement that is independent of the specific features that make the individual objects a dog, a chair, red, or big.

3. The ability to make abstract judgements of sameness along a single dimension—to know that two green objects are alike in the same way as two blue objects—develops sometime later, after the acquisition of the words by which we talk about the perceptual properties of objects [Smi84, Smi89b].

4. Early judgements of sameness along a dimension appear to be contaminated by overall similarity. That is, 3- and 4-year-old children will call a big red square and a big orange square *the same size* but will refuse to affirm that a big red square and big blue square are *the same size* [Kem82].

It is these facts that were are specifically interested in accommodating or accounting for.

2 The Model

2.1 Overview

We model the phenomena outlined above using a connectionist network which takes inputs in the form of “pre-perceived” visual images together with dimension words such as *color* and *size* and yields lexical outputs such as *big*, *red*, *same*, and *different*. The main features of the model are the following:

1. There is a single general comparison network which can compare objects along various dimensions as well as for global similarity.

2. There is a separate network for categorization, both by dimension and in terms of complex categories such as DOG. The dimension and the internal object representations used by this network are shared by the comparison network.

3. Input from a DIMENSION WORD layer acts as a “filter” on the internal representations of objects once the system has been trained on appropriate categorization and comparison tasks. Filtered representations contain information mainly about the relevant dimension; irrelevant dimensions are factored out.

2.2 Categorization

The CATEGORIZATION component is composed of a simple pattern associator; it is illustrated in Figure 1. Input to this component comes in in the form of a "pre-perceived object", corresponding to the output level in a theory such as Treisman and Gelade's [TG80] (see also [Smi89b]). The pre-perceived object has been segregated from background objects as an individual in the visual field and contains information about its perceptual features. There are separate sets of units for each of several "pre-dimensions" at this level. However, the system does not recognize these dimensions as such; they cannot be used in making categorizations or comparisons along particular dimensions.

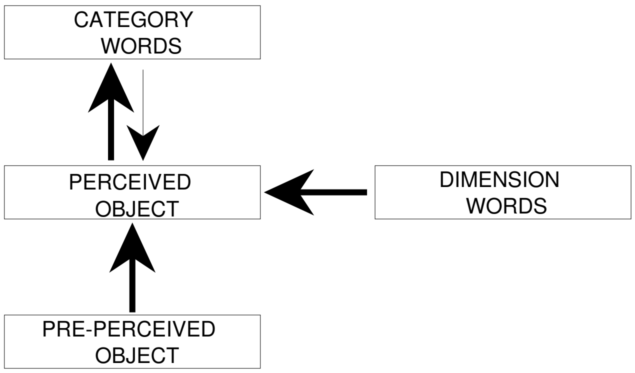

Figure 1: Categorization Component

The PERCEIVED OBJECT layer corresponds to what is perceived or experienced. A pattern on this layer is a compression of the pre-perceived object pattern and may be influenced by input from other layers, in particular from the DIMENSION WORDS layer. This influence takes the form of the focusing of attention on one or more dimensions in the perceived object representation. For example, the question *what color is it?* should cause the color information to dominate the representation.

The CATEGORY WORD level corresponds to the internal representation of words such as *red*, *big*, and *chair*. Each word is represented by a single unit. In categorization this layer is the output of the network; the system "sees" an object and names it or assigns it an attribute. Though it will not concern us in this paper, the CATEGORY WORDS layer may also function as an input layer (indicated by the thin arrow in the figure), e.g., in modeling the system's response to an utterance like *the marble is green*. In this case, it influences the pattern of activation on the PERCEIVED OBJECT layer.

While the network is designed to learn both complex categories such as DOG and CHAIR as well as attributes such as RED and BIG, we will be concerned only with the latter in the paper.

During training and subsequent testing on categorization, the network is presented with a pre-perceived "object" on its input layer, consisting of a pattern of features on the various pre-dimensions, together with a pattern on the DIMENSION WORDS layer. On the DIMENSION WORDS layer, either a single unit is on, corresponding to a question about one dimension, e.g., *what color is the object?*, or no unit is on, corresponding to instructions or intent to simply describe the input object. The network is trained using backpropagation [RHW86]. However, in an effort to make the degree of supervision more realistic, output targets are provided only for those units which are above a response threshold, unless no unit goes above the threshold, in which case a target (1.0) is provided for a single appropriate unit.

2.3 Comparison

Explicit comparison can also be implemented in a pattern associator which takes the two representations to be compared as inputs and associates them with a degree of similarity. Rumelhart et al. [RHW86] described a network that used backpropagation to do just this. For a similiarity-detection network like this to be applied to the problem at hand, the system must have the capacity to maintain representations of two items in short-term memory at once. The implementation of this mechanism in the model is discussed below.

An important question is the degree of generality of the similarity-detection mechanism. One possibility would be separate components of this type for different purposes, for example, one for comparing objects, another for processes; one for overall similarity, others for similarity along particular dimensions. We believe the evidence is in favor of a general, "dumb" similarity-detection mechanism which does not have to know what sorts of things it is comparing and along what dimension. If this were not the case, we would not expect knowledge of one sort of explicit comparison to generalize to another. For example, the fact that a child could comprehend and produce appropriately sentences with *same color* would have little to do with her ability to comprehend and produce sentences with *same size*. The fact that languages tend to have a single word for the notion SAME is perhaps the strongest evidence that there is a single general mechanism.[1]

Thus we posit, alongside the familiar implicit comparison mechanism that implements categorization, a COMPARISON component. This is just another pattern associator, with a hidden layer to handle inputs which are not linearly separable. Run in one direction, the COMPARISON component takes representations of two items and determines their similarity. Representations of all types, including those for which only a particular dimension is relevant to the comparison, are compared using the same units and connections. Run in the other direction, this component implements the system's response to sentences like *my sand castle is the same as yours*. In this case the inputs consist of a complete representation of one object (*your sand castle*), a partial representation of another (*my sand castle*), and the asserted degree of similarity, and the output is a complete representation of the lesser-known object.

The fact that languages refer to dimension categorization (*what color?*) using the same nouns as are used for sameness comparison (*same color*) is an indication that both processes make use of the same internal representations for the dimensions.[2] Our model incorporates this claim by using the same dimension word and perceived object patterns for comparison as for categorization.

The COMPARISON component of the model is shown in Figure 2. The compared patterns appear on separate groups of units. One is just the PERCEIVED OBJECT layer, which also participates in the CATEGORIZATION component. The other input group is a short-term memory buffer which contains a copy of a recent pattern from the PERCEIVED OBJECT layer.

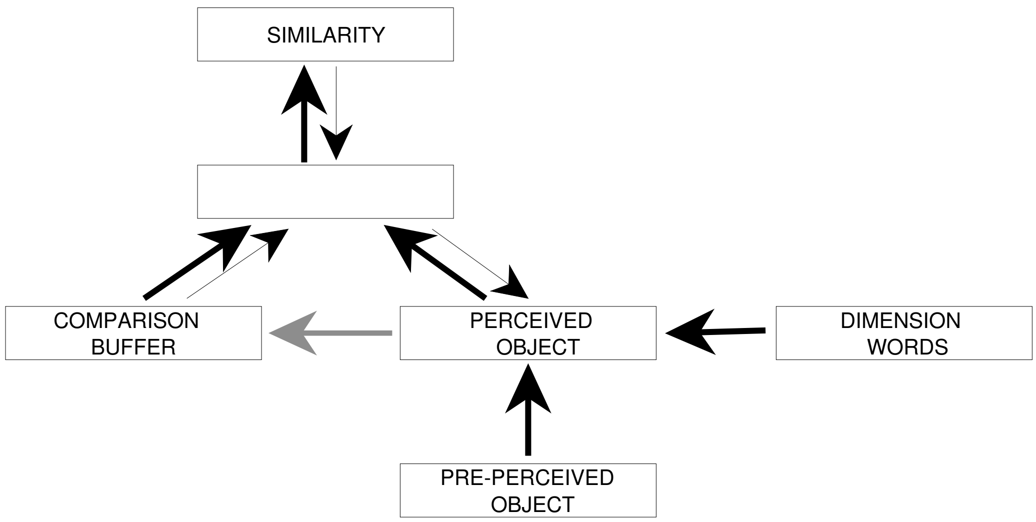

Figure 2: Comparison Component

Run in the direction of the thick arrows in the figure, the network compares two input objects. Input from the PRE-PERCEIVED OBJECT and DIMENSION WORDS layers produces a pattern on the PERCEIVED OBJECT

[1] In any case, this is an empirical question; it remains to be established that there is a single sameness comparison mechanism.

[2] Again, this is an empirical issue: it needs to be established that children know the use of words such as *color* in comparison once they have acquired their use in categorization.

layer. If a dimension word unit is on, the perceived object pattern is a "filtered" version of the object. This pattern is copied to the COMPARISON BUFFER group, and a second pre-perceived object is fed to the network together with the same dimension word pattern. Finally the two (possibly filtered) object representations are compared at the SIMILARITY layer, which consists of a single unit.

Run in the direction of the thin arrows in the figure, the network models responses to assertions about the sameness of objects. The input is a pattern on the COMPARISON BUFFER units representing one object, a pattern on the SIMILARITY unit representing sameness, and a pattern representing (possibly incomplete) knowledge about the second object on the PERCEIVED OBJECT layer. The output is a completed (or corrected) representation on the PERCEIVED OBJECT layer.

2.4 Dimension as a Filter

The CATEGORIZATION and COMPARISON components share the perceived object representations, which are subject to the filtering effects of dimensional input. In order for the CATEGORIZATION subnetwork to succeed on a dimensional categorization task (*what size is it?*), the input on the DIMENSION WORDS layer should highlight the relevant dimension and attenuate the other dimensions to the extent that only the appropriate output unit reaches the response threshold. It is thus possible for the network to produce an appropriate response even with some contamination from irrelevant dimensions. That is, training on the categorization task may not result in weights on the connections from the DIMENSION WORDS to the PERCEIVED OBJECT layer which completely eliminate dimensions other than the input dimension from a representation.

The comparison task is more demanding, however. Consider the case of two objects which are the same on the dimension in question but significantly different on all others. Any contamination from the irrelevant dimensions at all would adversely affect the output on the SIMILARITY unit.

How might we expect the system's performance on the comparison task to vary with time? We assume the COMPARISON component is trained initially simply to detect similarity between pairs of real-valued input patterns. At this point, assuming random initial weights on the connections from the DIMENSION WORDS units, the system would be completely unable to make use of dimension information. Next, training on the categorization task would result in some filtering out of dimensions other than the one that is input from the DIMENSION WORDS layer. Now the network should also begin to be able to detect similarity between two objects along a given dimension. But, as in children, similarity judgements at this stage should still depend to a large extent on the overall similarity between the objects. Training on the comparison task itself would then refine the behavior of the dimension filter. Given two objects and the assertion that they are the same on a given dimension, their filtered representations should be identical. Thus the filtered representation of the first could be used as a target for the filtered representation of the second. Together with continued training on the categorization task, this should result in an adequate representation of dimension.

3 Experiments

To test the hypothesis that our model could account for the effects discussed above, we ran an experiment in which the same network was trained on categorization and comparison tasks. The procedures described in the following sections were repeated 6 times with different initial random connection weights.

3.1 Categorization Task

A network of the type shown in Figure 1 was first set up with random initial weights. The PRE-PERCEIVED OBJECT layer consisted of 28 units, 7 for each of 4 simple linear "pre-dimensions". Each of these was modeled on the psychological dimension of length. The PERCEIVED OBJECT layer consisted of 20 units; that is, there was some compression of the patterns from the input layer. The DIMENSION WORDS layer contained 3 units, one for each of the output dimensions, that is, those for which there were target categories. The CATEGORY WORDS units consisted of 9 units, one for each of the target categories.

The network was trained to perform dimension categorization on 2500 randomly generated "pre-perceived" input objects. Input objects were constrained in ways designed to model in a gross fashion the structure that is present in the world; the details need not concern us here. Also given to the network was input from a single unit in the DIMENSION WORDS layer. Thus the network's task corresponded to a question such as *what*

Table 1: Comparison of "Same" and "Different" Pattern Distances

	Same	Different
Before training	1.464	0.966
After categorization training	1.172	1.211
After comparison training	0.314	0.519

color is this object?. Output targets were provided using the procedure described in 2.2. That is, targets depended on the system's own output, in ways that seem to correspond to what goes on in actual language acquisition contexts.

The performance of the network on categorization improved overall, as would be expected, though with the output-generated targets, improvement was not as smooth as it would have been with completely supervised learning. For most of the runs, the network succeeded in correctly categorizing at least 25 consecutive input objects at some point during the training.

What interests us for the purposes of this paper, however, is how well the network factors out dimensions other than the one which is input on the DIMENSION WORDS layer. To determine this, we created a set of 45 test pattern pairs. These were of two types, those in which the objects were the same on the input dimension and different on the other three dimensions (hereafter referred to as the "same" pairs) and those in which the objects were different on the input dimension and the same on the other three dimensions (hereafter referred to as the "different" pairs). Testing the network consisted in running it with the objects in the test pairs as inputs and determining the Euclidian distance between the PERCEIVED OBJECT responses to the inputs for each pair. Of interest is the relative distance between the pairs. To the extent that the dimension input is behaving like a filter, as described in 2.4, the distance between the hidden-layer patterns for the "same" pairs should be smaller than that between the "different" pairs.

We made these comparisons before the network was trained, after the categorization training on 2500 inputs, and again following the second, comparison phase of training (described below) on 2500 additional inputs. Table 1 shows the results of the comparisons, averaged over the 6 runs.

Not surprisingly, the "same" pairs start out considerably further apart than the "different" ones because they differ on three out of four, vs. one out of four input dimensions. The effect of training on the categorization task, as predicted, is to significantly ($p < .01$) diminish this difference, though the "same" pairs are only slightly closer than the "different" pairs. However, despite the fact that the dimension filter is not doing a very good job of eliminating irrelevant dimensions from the input, the network has learned to categorize quite well. Recall that good categorization by overall similarity together with an inability to filter out irrelevant dimensional information is precisely the the performance found in very young children.

3.2 Comparison Task

During the second phase of training, half of the time, on the average, the task was that of the first phase, categorization of an input along a given dimension. For the remaining trials, the task represented the response to an assertion such as *X is the same color as Y*. That is, the task was that indicated by the thin arrows in Figure 2. Given an input object which is red, big, round, and smooth, and another which is red, small, square, and rough,[3] the system was expected to use the information that the objects were the same color so that it later could make sameness judgements of its own. Though the network itself was not comparing the objects, for the sake of simplicity this will be referred to as the "comparison" task.

The comparison task was implemented in the following manner. The input pattern for one object was presented to the PRE-PERCEIVED OBJECT layer together with one lexical dimension on the DIMENSION WORDS layer, just as for the categorization task. The pattern this yielded on the PERCEIVED OBJECT layer was then saved. Next the second input object, identical to the first on the input dimension, was presented in the same way. Now the stored pattern was treated as a target for the PERCEIVED OBJECT layer. Note that the idea is not that the response to the first object is somehow superior to the response to the second, only that

[3] Labels for the various dimensions are used for convenience only; as noted above, there was no intent to model the structure of dimensions such as COLOR and SHAPE

training in this manner should bring the patterns closer together. The important point is that an effective filtering mechanism is learned via the explicit comparison of objects along single dimensions.

Note that we did not actually use the COMPARISON component for the implementation of this task. We assumed that the COMPARISON network, given a dimension-filtered representation of one object (in the COMPARISON BUFFER layer) and an indication that it was the same as another object (resulting in high output on the SIMILARITY unit), could generate its own internal target for the filtered representation of the second object by simply copying the pattern for the first object.

Results of comparisons between the "same" and "different" pairs following the second phase are shown in Table 1. As predicted, the "same" distances have decreased significantly ($p < .01$) relative to the "different" distances. At the same time, there is a significant overall decrease in distances for both pairs. Because we trained the network only on comparison of objects that were meant to be the same, this is not surprising. Though it learned to treat some as more similar than others (those that are the same on the input dimension), in general it moves object representations closer together.

4 Discussion

Our model offers a way in which dimension words such as *color* can have the same internal representations whether they apply to categorization or to comparison of objects. The two tasks place the same sort of demands on the dimension representations: within the distributed PERCEIVED OBJECT representations, features of the input dimension must be played up and features of other dimensions played down. The comparison task, however, is more demanding in this regard. Some contamination from irrelevant dimensions does not prevent correct responses to the categorization task. This aspect of our model may help us understand why young children are able to categorize objects even seemingly by a single attribute well in advance of their ability to make explicit comparisons along single dimensions.[4]

Importantly, training on categorization is insufficient for the formation of an effective dimensional filter. Following categorization training only, the distance between pairs of dimension-filtered representations is about the same when the two objects are the same only on the input dimension as it is when they differ only on the input dimension. We still find a contamination effect like that observed in preschool children [Kem82]. It is training on explicit comparison in our model that gives rise to effective dimension filters. The developmental implications of this finding are clear. Training children on explicit comparison tasks—teaching them the language of dimensional comparison—may be a causal force in the emergence of the ability of children in the late preschool period to selectively attend to single dimensions.

One possible criticism of our experiments is the sequencing that we imposed on the learning. The network was trained on categorization and only then given the comparison task. This order fits the developmental facts [Mac82]. Nonetheless, determining whether (and how) our results depend on the sequencing of training will be an important aspect of future research.

This research makes three contributions. First, it provides a model of one of the major trends in human development, from wholistic object comparison to dimensional comparisons. This trend is so pervasive across cognitive tasks that it has the status of a principle in developmental psychology (see, e.g., [Gib69]).

Second, our model distinguishes categorization and comparison in ways which clarify the theoretical issues and suggest new experiments. We consider one example here. Alongside the sameness comparison that has concerned us in this paper, there is comparison signalled by the plural and by the adjective *both*. With these expressions, the speaker makes reference to an explicit category rather than just the dimension along which the two objects are similar. Within the terms of our model, the correct adult use of these forms requires that the lexical outputs of the categorization of the objects be themselves compared, a relatively sophisticated procedure. The fact that children use the plural and iterative naming patterns relatively early on [Sug83] leads one to suspect that there is a simpler (but perhaps fallible) route to the use of these expressions. There is such a route in our model: children may be responding with the plural only when the objects are globally similar. This suggests experiments in which children at different ages are given the opportunity to use the plural for objects which belong to the same category but differ in terms of overall similarity.

Third, our model may bring insights to connectionist modeling, particularly as it is applied to cognitive development. For example, our model relies on the use of one internal representation as a target for another.

[4]This fact of the model also fits with one current mathematical model of similarity judgments in young children, [Smi89b].

We believe this idea may be applicable generally when there is reason to posit representations that are shared by processes which constrain them in particular ways.

5 Conclusions

The central problem in understanding development is understanding how new behaviors emerge. The inherent difficulty of this problem has led much of the best empirical and theoretical work in cognitive development to be essentially adevelopmental. The dominant empirical strategy consists of describing behavior at different developmental points. Thus, we know for example that 5-month-olds can discriminate colors, that 2-year-olds have severe difficulty learning color words relative to other words, that 5-year-olds have a rudimentary mapping of color words to the color space, and that adults have a sophisticated and highly structured organization of color in language, perception, and knowledge. But we do not know how the abilities of babies translate into the difficulties of toddlers, the minimal competence of children, and the sophistication of adults (see [AS88] for review).

We believe that the model described in this paper provides a starting point for understanding developmental <u>change</u>, at least for the crucial area of comparison and categorization. While the model is still quite primitive, it already provides an account of how a system can get from a stage at which it judges two objects to be the same color only if they are similar overall to a stage at which it can make the judgement without paying attention to irrelevant features of the objects. It does this by adjusting its connection weights in such a way that dimensional input has the effect of filtering out the irrelevant features in its internal representations for objects. Thus in this connectionist model, we have an account of how change might occur to cause the trend from overall similarity to selective dimensional comparisons.

References

[AS88] R. N. Aslin and L. B. Smith. Perceptual development. In M. R. Rosenzweig and L. W. Porter, editors, *Annual Review of Psychology*. 1988.

[Gib69] E. J. Gibson. *Principles of Perceptual Learning and Development*. Appleton-Century-Crofts, New York, 1969.

[Kem82] D. G. Kemler. The ability for dimensional analysis in preschool and retarded children: Evidence from comparison, conservation, and prediction tasks. *Journal of Experimental Psychology*, 34:469–489, 1982.

[Mac82] J. Macnamara. *Names for Things: A Study of Human Learning*. MIT Press, Cambridge, MA, 1982.

[RHW86] D. E. Rumelhart, G. Hinton, and R. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel Distributed Processing*, volume 1, pages 318–364. MIT Press, Cambridge, MA, 1986.

[Smi84] L. B. Smith. Young children's understanding of attributes and dimensions: A comparison of conceptual and linguistic measures. *Child Development*, 55:363–380, 1984.

[Smi89a] L. B. Smith. From global similarities to kinds of similarities: The construction of dimensions in development. In S. Vosniadou and A. Ortony, editors, *Similarity and Analogy*, pages 146–178. Cambridge University Press, Cambridge, 1989.

[Smi89b] L. B. Smith. A model of perceptual classification in children and adults. *Psychological Review*, 96:125–144, 1989.

[Sug83] S. Sugarman. *Children's Early Thought*. Cambridge University Press, Cambridge, 1983.

[TG80] A. Treisman and G. Gelade. A feature-integration theory of atttention. *Cognitive Psychology*, 12:93–136, 1980.