The Development of the Notion of Sameness: A Connectionist Model

Michael Gasser and Linda B. Smith
Indiana University

Abstract

Comparison is of two types, the implicit sort that is behind all categorization and the explicit sort by which two object representations are compared in short-term memory. Children learn early on both to categorize and to compare explicitly, but they only learn to use dimensions in these processes considerably later. In this paper we present a connectionist model which brings together categorization and comparison, focusing on the development of the use of dimensions. The model posits (1) a general comparison mechanism which is blind to the nature of its inputs and (2) the sharing of internal object and dimension representations by categorization and comparison processes. Trained on the two processes, the system learns to use dimension inputs as filters on its representations for objects; it is these filtered representations which are matched in comparison. The model provides an account of the tendency for early comparison along one dimension to be disrupted by similarities along other dimensions and of the process by which the child might overcome this deficiency.

Background

Comparison and Cognition

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 how similar that object is to other objects known to be *dogs*. But humans do more than categorize objects; we also compare objects along a wide array of perceptual dimensions. For example, we judge a dog to be the same color as our cat or 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 (*what color?*) and comparison (*same color?*) 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 tasks.

Categorization and Comparison

Categorization involves comparing a stimulus in short-term memory to representations of previously encountered stimuli in long-term memory. A simple pattern associator performs this implicit form of comparison through the connection weights that make up its long-term memory. Categorization can be in terms of either complex categories such as DOG and CHAIR or dimensional attributes such as RED and BIG.[1]

But the implicit comparison between an item in short-term memory and long-term representations may be quite different from the comparison of two items in short-term memory. We shall call the latter **explicit comparison**. If we take the evidence from language seriously, this process goes on often in human cognition. A sentence such as *my ball is the same color as yours* requires speaker and hearer to maintain representations of both objects in short-term memory, where they can be compared. In this paper we will only be concerned with that subtype of explicit comparison which is signalled in English by the word *same* along with a perceptual dimension noun such as *size*.

In order for an abstract comparison device which looks for symmetry in its two input patterns to make judgements of "same thing," "same color," and "same size," it must have access to representations in which only the relevant dimension manifests itself. Irrelevant dimensions need somehow to be "filtered" out.

[1] We are concerned here with categorization in the sense of naming an object or an attribute and not in the sense of the underlying meaning of concepts.

Developmental Facts

There is a well-documented trend in the development of object and dimensional comparisons and object and dimensional language. The following specific facts are those that we are interested in accommodating.

1. Early object categorizations are principally across all dimensions at once [Smith 1989a, 1989b]. 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—appears very early [Smith, 1984]. By 24 months, children comment on the similarity of objects through iterative naming, counting, and use of the plural [Sugarman, 1983]. The productive use of a form such as the plural at this age suggests that there is a comparison component that operates early and that is independent of specific perceptual properties.
3. The ability to make judgements of sameness along a single dimension—to know that two green objects are alike in the same way as two blue objects—develops later, after the acquisition of the words by which we talk about the perceptual properties of objects [Smith, 1984, 1989b].
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 call a big red square and a big blue square *the same size* [Kemler, 1982].

The Model

We model these phenomena using a connectionist network which takes as inputs "pre-perceived" images and dimension words such as *size* and yields lexical outputs such as *big* and *same*. The architecture of the model is shown in Figure 1. Boxes denote banks of units and solid arrows complete connectivity between banks. The main features of the model are the following:

1. Global comparison and comparison along various dimensions are handled by the same subnetwork.
2. The dimension and the internal object representations are shared by the comparison and categorization components of the system.
3. Following training on categorization and comparison, input from dimension words "filters" out irrelevant dimensions in the representations of objects.

Categorization

The CATEGORIZATION component, shown on the right side of Figure 1, is composed of a simple pattern associator. Input to this component comes in in the form of a "pre-perceived object" (hereafter "PPO"),

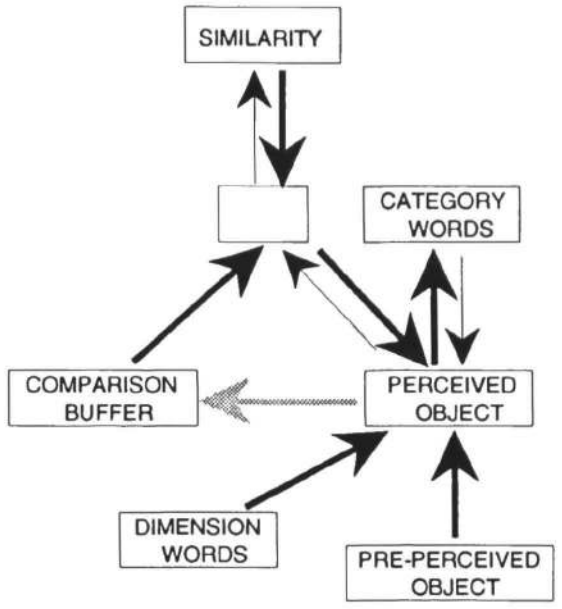

Figure 1: Architecture of the Model

corresponding to the output level in a theory such as Treisman and Gelade's [1980] (see also Smith [1989b]). The PPO has been segregated from background objects 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. The PERCEIVED OBJECT ("PO") layer corresponds to what is perceived or experienced. A pattern on this layer is a compression of the PPO pattern and may be influenced by input from other layers, in particular from the DIMENSION WORDS ("DW") layer. This influence takes the form of the focusing of attention on one or more dimensions. For example, the question *what color is it?* should cause color information to dominate the representation.

On the CATEGORY WORDS level, attribute words like *red* and *big* and complex category nouns such as *dog* and *chair* are each assigned 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. While the network is designed to learn both complex categories such as DOG and attributes such as BIG, we will be concerned only with the latter. 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 PO layer.

During training and testing on categorization, the network is presented with a PPO on its input layer, consisting of a pattern of features on the various pre-

dimensions. A single unit is also turned on on the DW layer, corresponding to a question about one dimension, e.g., *what color is the object?*. The network is trained using backpropagation [Rumelhart *et al.*, 1986]. In an effort to make the degree of supervision 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.

Comparison

Alongside the CATEGORIZATION subnetwork, we propose a "dumb" COMPARISON component which does not know what objects are being compared or the dimensions on which they are to be compared. It simply compares patterns of activation. Selective attention—changes in the dimensions along which the comparison is to made—is accomplished by the same mechanisms that are involved in categorization by dimension. The overlap between these mechanisms is suggested by the fact that languages refer to dimension categorization (*what color?*) using the same nouns as are used for sameness comparison (*same color*).[2]

The COMPARISON component is shown on the left side of Figure 1. This is another pattern associator, similar to the symmetry network described by Rumelhart et al. [1986], with a hidden layer to handle inputs which are not linearly separable. The compared patterns appear on two input groups. One is just the PO layer, which also participates in categorization. The other is a short-term memory buffer which contains a copy of a recent pattern from the PO layer.

This component implements two sorts of processes. Run in one direction, it compares two input objects. Input from the PPO and DW layers produces a pattern on the PO layer. If a DW unit is on, the PO pattern is a "filtered" version of the object. This pattern is copied to the COMPARISON BUFFER layer, and a second PPO is fed to the network together with the same DW pattern. Finally the two (possibly filtered) object representations are compared at the SIMILARITY layer, which consists of a single unit.

Run in the other direction (the thick arrows in the figure), the network models responses to assertions about the sameness of objects (*my doll is the same as yours*). 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 PO layer. The output is an updated representation on the PO layer.

Dimension as a Filter

The CATEGORIZATION and COMPARISON components share the PO 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 color is it?*), the input from the DW layer should highlight the relevant dimension and attenuate the other dimensions to the extent that only the appropriate output unit (e.g., RED) 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 categorization may not result in DW-to-PO connection weights which completely eliminate irrelevant dimensions from a representation.

The comparison task is more demanding. 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 the system's performance on the comparison task vary with time? We assume the COMPARISON component is first trained simply to detect similarity between pairs of input patterns. At this point, the system would be 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 DW 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 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.

Experiment

We ran an experiment in which the same network was trained on categorization and comparison tasks. The procedures described below were repeated six times with different initial random connection weights.

Categorization Task

A categorization network was first set up with random initial weights. The PPO layer consisted of 28 units, 7 for each of 4 simple linear "pre-dimensions". The PO layer consisted of 20 units; that is, there was some compression of the patterns from the input layer. The DW layer contained 3 units, one for each of the output dimensions, that is, those for which there were target categories. There were 9 CATEGORY WORDS units, one for each of the target categories.

The network was trained to perform dimension categorization on 2500 randomly generated "pre-

[2] It remains to be established whether children know the use of words such as *color* in comparison once they have acquired their use in categorization.

Table 1: "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

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 DW layer. Thus the network's task corresponded to a question such as *what color is this object?*. Output targets were provided using the procedure described above. 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 by the end of the training.

The critical question, however, is how well the network learns to selectively attend to single dimensions. 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 PO 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 above, 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.

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 is to significantly ($p < .01$) diminish this difference, though the "same" pairs are only slightly closer than the "different" pairs. Although 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. Good categorization along dimensions and an inability to ignore overall similarity in comparing objects on one dimension are precisely the behaviors exhibited by 4- and 5-year-old children.

Comparison Task

During the second phase of training, the network was trained on more categorization on half of the trials and on explicit comparison on the other half. The comparison task was designed to fit the real task in which there are two objects in front of the child and the adult says *X is the same color as Y*. This task conforms to running the network in the direction indicated by the thick arrows in the COMPARISON part of Figure 1. 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 to help it later make sameness judgements of its own.

This task was implemented in the following manner. The input pattern for one object was presented to the PPO layer together with one lexical dimension on the DW layer, just as for the categorization task. The pattern this yielded on the PO 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 PO 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 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.

We did not actually use the COMPARISON component for the implementation of this task. We assumed that the COMPARISON network, given a filtered representation of one object in the COMPARISON BUFFER and an indication of sameness to another object 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 this phase are shown in Table 1. As predicted, the "same" distances have decreased significantly ($p < .01$) relative to the "different" distances. There is also 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 moved object representations closer together.

[3] Labels for the various dimensions are used for convenience only.

Discussion

In our model dimension words such as *color* have the same internal representations whether they apply to categorization or to comparison of objects. The two tasks place similar demands on the dimension representations: within the distributed PO representations, features of the input dimension must be played up and features of other dimensions played down. As we have seen, however, the comparison task is more demanding in this regard. This aspect of our model, which fits with one current mathematical model of children's similarity judgements [Smith, 1989b], may help us understand why young children are able to categorize objects even seemingly by a single attribute long before they can make explicit comparisons along single dimensions.

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. 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 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 experiment is the sequencing that we imposed on the learning. Comparison training began only after categorization was learned. This order fits the developmental facts [Macnamara, 1982]. 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. Second, the model distinguishes categorization and comparison in ways which clarify the theoretical issues and suggest new experiments. For example, it suggests that children's early use of plural and iterative naming may depend on global similarity, in addition to category identity, a hypothesis that could be tested empirically. Third, our model may bring insights to connectionist modeling of cognitive development. For example, the idea of using one internal representation as a target for another may be applicable generally when there is reason to posit representations that are shared by processes which constrain them in different ways.

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 work in cognitive development to be essentially adevelopmental. The dominant empirical strategy consists of describing behavior at different developmental points. We know for example that 5-month-olds can discriminate colors, that 2-year-olds have 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 exhibit sophisticated and highly structured color concepts. 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.

While the model described in this paper is still primitive, it already demonstrates 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. 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, for the crucial area of comparison and categorization, this connectionist model provides a starting point for understanding developmental change.

References

[Kemler, 1982] 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.

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

[Rumelhart *et al.*, 1986] 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.

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

[Smith, 1989a] 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.

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

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

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