SIMILARITY SCALING AND COGNITIVE PROCESS MODELS

Robert M. Nosofsky

Department of Psychology, Indiana University, Bloomington, Indiana 47405

Contents

INTRODUCTION

A well-known virtue of similarity-scaling techniques such as multi-dimensional scaling, Thurstonian modeling, and clustering is that they reveal hidden structure underlying psychological data. By applying these techniques, complex matrixes of similarity data such as similarity ratings, identification confusions, and same-different errors can be efficiently described, summarized, and displayed, and a deeper insight into the underlying basis of the similarity data can be derived.

Beyond describing and summarizing data, however, scaling techniques can be viewed as psychological models for the mental representation of interobject similarity. In this chapter I review the role of similarity-scaling techniques as components in formal psychological models of perceptual and cognitive processes.

Cognitive models are often conceptualized as representation-process pairs (e.g. Anderson 1976). Objects that are perceived or remembered receive some internal representation. Various cognitive processes are then assumed to act upon that representation. The particular processes that operate are task dependent—they will vary depending on whether subjects are asked to discriminate among objects; identify, categorize or recognize them; supply similarity ratings; make preference judgments, and so forth. Thus, to understand performance in tasks involving similarity data we must specify not only an underlying similarity representation, but also the cognitive processes that act on that representation.

The beauty of deriving a similarity-scaling representation by modeling performance in a given task is that the derived representation can then be used to predict performance in independent tasks involving the same objects and stimulus conditions (e.g. Cliff 1973; Henley 1969; Hutchinson & Lockhead 1977; Monahan & Lockhead 1977). For each task, one needs to specify the cognitive processes that are operating, but key aspects of the underlying similarity representation may be invariant. Thus, similarity-scaling techniques allow one to characterize how performance across independent tasks is related. I believe that the characterization of such invariant relations should be one of the central goals of psychological science.

I also argue that any evaluation of how well a scaling representation accounts for similarity data must occur within the framework of formal process models. Mispredictions involving a proposed scaling representation may reflect inadequacies in the scaling model, but they may also indicate a failure to specify adequately the cognitive processes that operate on the representation.

I organize this chapter into two main parts. Part 1 focuses on models incorporating deterministic multidimensional scaling (MDS) approaches, Part 2 on probabilistic MDS approaches. I distinguish between the two as follows: In deterministic MDS, each object is represented as a single point in the spatial representation, whereas in probabilistic MDS, each object is represented as a probabilistic distribution of points. (Note that by this definition, many deterministic MDS models can still have probabilistic components. For example, although each object is represented as a fixed point, the distance-judgment process itself may be noisy.) Because of space limitations I review only spatial models, although the role of discrete-feature and network approaches as components in cognitive process models is of equal importance.

The last review article on scaling was that of Gescheider (1988), who also emphasized linkages between scaling methods and perceptual and cognitive processes. However, Gescheider's (1988) review was concerned with classic psychophysical tasks involving unidimensional scaling, such as magnitude estimation, whereas the present review focuses on multidimensional cognitive processes involving similarity data. I do not attempt to match the broad and comprehensive reviews on multidimensional scaling provided in the chapters by Carroll & Arabie (1980) and Young (1984a), but rather focus on the intersection between similarity scaling and cognitive-process models. Ashby's (1992) edited volume, *Multidimensional Models of Perception and Cognition*, covers more extensively many of the MDS-based models discussed here.

DETERMINISTIC MULTIDIMENSIONAL SCALING APPROACHES

Universal Laws of Generalization and Similarity

A major recent contribution in the use of MDS techniques for understanding cognitive processes was Shepard's (1987) "Toward a universal law of generalization for psychological science." As argued by Shepard, an organism's generalization from one situation to another must surely stand among the most fundamental psychological processes. The process of generalization is often studied within the context of identification learning paradigms. In these paradigms, subjects learn to associate a unique response with each member of a set of stimuli. Generalization or similarity is measured in terms of the probability of interstimulus-confusion errors. A well-known quantitative measure of generalization or similarity between stimuli i and j is given by $g_{ij} = [(p_{ij} \cdot p_{ji})/(p_{ii} \cdot p_{jj})]^{1/2}$, where p_{ij} is the probability that stimulus i is identified as stimulus j. An intuitive justification for this measure is provided by Shepard (1958a). The major theoretical justification is that this quantity gives an estimate of the similarity parameters in the classic similarity choice model (SCM) for predicting identification confusions (Luce 1963; Shepard 1957). I discuss the SCM extensively in the following section.

The similarity measures obtained in generalization experiments can be used as input to nonmetric multidimensional scaling algorithms, and an MDS solution for the stimuli can be derived. The generalization measures g_{ij} can then be plotted against the corresponding distances d_{ij} between points in the derived solution to discover the form of the gradient relating generalization to distance in the psychological space. Shepard (1987) presents 12 different plots of these derived generalization gradients, the generalization measures having been obtained in experiments involving both human and animal subjects, and both visual and auditory stimuli. As summarized by Shepard (1987:1319), "in every case, the decrease of generalization with psycholog-

ical distance is monotonic, generally concave upward, and more or less approximates a simple exponential decay function. . . ."

It is critical to realize that in discovering this exponential law relating generalization to psychological distance, Shepard (1987) operates at an entirely psychological level of analysis. Distances between objects are derived by using MDS techniques that rely on only the generalization measures themselves. "Psychophysics" is not involved, in the sense that no physical measurements are ever taken on the stimuli. Indeed, Shepard (1987:1318) argues that the invariant law of generalization was "attainable only by formulating that law with respect to the appropriate abstract psychological space."

When more than one dimension is required to describe the similarity structure of a set of stimuli, the generalization data also provide information about the metric structure of the psychological space. Noting the extensive literature on this subject, Shepard (1987) suggests that for stimuli composed of relatively unanalyzable, "integral" dimensions, such as colors varying in lightness and saturation, the distance structure of psychological space is well approximated by the Euclidean metric; whereas for stimuli composed of highly analyzable, "separable" dimensions, such as forms varying in size and orientation, the distance structure is generally best approximated by the "city-block" metric (see Garner 1974 and Shepard 1991 for reviews).

Given these observed regularities in the form of the generalization gradient and the metric structure of psychological space, Shepard then proposes a cognitive process model to account for the laws. Assume that an experience with an object has had some significant consequence for an organism. The organism must decide which new objects are enough like the old one that they are likely to have the same consequence. A class of objects with the same consequence corresponds to a region in the organism's psychological space that Shepard terms a *consequential region*.

In finding a given object to be consequential, the organism learns that there is some consequential region that overlaps the point in psychological space corresponding to that object. Probability of generalization to a new object would be determined by estimating the conditional probability that the consequential region also overlaps the point corresponding to the second object. To determine this conditional probability precisely, one needs information concerning the probability that the consequential regions in an organism's psychological space are of given shapes, sizes, and locations, and then needs to integrate over the hypothesized forms of the consequential regions. Nevertheless, given remarkably weak assumptions, many justified by evolutionary considerations, Shepard (1987) demonstrates that the conditional probability of overlap is always well approximated by an exponential decay function of the distance between the objects in the psychological space. He concludes, "Evidently, the form of [the generalization gradient] is a relatively robust

consequence of the probabilistic geometry of consequential regions" (Shepard 1987:1320). Finally, Shepard's (1987) cognitive process model also provides an account of why the Euclidean and city-block metrics closely approximate the distance structure for integral-dimension and separable-dimension stimuli, respectively.

CHALLENGES   The proposed universal law of generalization has not gone unchallenged. Work reported by Nosofsky (1985a,b; 1986, 1989) raised questions about the universality of the exponential-decay generalization gradient, although these questions have now been largely resolved (Ennis 1988; Ennis et al 1988; Nosofsky 1988b; Shepard 1986, 1987, 1988). Using essentially the same theoretical approach as described by Shepard (1987), Nosofsky found evidence that in some identification-confusion experiments, the plot of similarity against psychological distance was Gaussian in form rather than exponential (see also Ashby & Lee 1991). The main difference between the experiments conducted by Nosofsky (1985b, 1989) and those discussed by Shepard (1987) is that Nosofsky's studies involved protracted identification training involving asymptotic performance with highly confusable stimuli, whereas the studies considered by Shepard involved identification learning of fairly discriminable stimuli. Shepard (1986:60) suggested that the Gaussian similarity functions observed by Nosofsky (1985a,b) reflected limitations on discrimination performance resulting from "irreducible noise in the perceptual/memory system," and not the cognitive form of similarity intrinsic to the process of generalization.

The difference in the situations considered by Shepard (1987) and Nosofsky (1985a,b) can be viewed theoretically in the following way. Because of noise in the perceptual/memory system, presentation of an object does not result in the same internal representation on every trial. Rather, over trials, presentation of an object gives rise to a probabilistic distribution of points in the observer's psychological space (see the second part of this chapter). Nevertheless, in Shepard's situations, the distance between the means of the object representations is so great, relative to the variability of these representations, that each object can be represented as essentially a single point in the psychological space. By contrast, in Nosofsky's situations, overlap among the alternative object distributions is substantial, so each object should really be represented as a distribution of points rather than as a single point. As will be seen in the section on probabilistic MDS, Ennis and his colleagues have demonstrated that such a model reconciles Shepard's proposed universal law of generalization with the findings reported by Nosofsky.

RELATED WORK   Alternative process models to account for the form of the generalization gradient have also been proposed. Staddon & Reid (1990) proposed a simple neural network model in which activation received by

individual units tends, over time, to spread to neighboring units in the network. When a given stimulus is presented for a moderate number of iterations, the activation gradient that it produces is approximately exponential in form. But when the stimulus is withdrawn, the diffusion process eventually produces a gradient that is Gaussian in form. Shepard (1990) notes that this neural network model is formally identical to his earlier proposed trace-diffusion model of stimulus generalization (Shepard 1958a). He argues that a limitation of both diffusion models is that they fail to predict that with continued training, subjects can eventually learn to discriminate perfectly between objects that are members of contrasting consequential regions. Shepard & Kannappan (1991) present a multi-layered, neural-network embodiment of the 1987 cognitive theory of generalization, which successfully predicts the form of the generalization gradient under different conditions of discrimination training. Finally, Gluck (1991) suggests that the configural-cue adaptive network model proposed by Gluck & Bower (1988) produces, in discrete-dimension domains, a generalization gradient that approximates the exponential.

In work related to Shepard's, Blough (1988) observed highly regular results relating reaction time in visual search tasks to distances in multidimensional psychological space. Pigeons were trained to peck at a unique target embedded in a field of identical distractors, and visual search reaction time (RT) was measured. The targets and distractors were drawn from fixed stimulus sets, such as squares varying in size, and rectangles varying in height and width. In most experiments, all possible pairs of forms from each set served as targets and distractors across trials, yielding a complete matrix of mean RTs, one for each pair of forms. MDS solutions were derived for the forms by using these matrixes of mean RTs as input. Blough then plotted the mean RT for each pair of forms against their distance (D) in the derived scaling solution, and found that for all stimulus sets, the RT gradients were well fitted by the function $RT - k = c \cdot exp(- b \cdot D)$, where k, c, and b are estimated parameters. Thus, just as occurs for generalization, there is the suggestion of lawful relations between visual search speed and psychological distance (see also Shepard et al 1975, for reports of lawful relations between discrimination RT and psychological distance).

Identification, Categorization, and Recognition

The MDS-based models of generalization and identification learning discussed by Shepard (1987) have been extended to account for categorization and recognition performance. Whereas in identification each stimulus is to be assigned a unique response, in categorization stimuli are to be classified into groups. Recognition refers to a memory experiment in which subjects judge whether items are "old" or "new."

IDENTIFICATION   One of the classic models for predicting identification performance is the *similarity choice model* (SCM) proposed by Shepard (1957) and Luce (1963), whose formal properties have been further investigated by researchers such as Smith (1980, 1982), Townsend (1971; Townsend & Landon 1982), Nosofsky (1985b, 1990), and Takane & Shibayama (1985). According to the model, the probability that stimulus i is identified as stimulus j is given by

$$P(R_j|S_i) = b_j \ \eta_{ij} \ / \ \Sigma b_k \ \eta_{ik}, \hspace{3cm} 1.$$

where η_{ij} ($0 \leq \eta_{ij}$, $\eta_{ij} = \eta_{ji}$) denotes the similarity between stimuli i and j, and b_j ($0 \leq b_j \leq 1$, $\Sigma b_j = 1$) is often interpreted as the bias for making response j. Although descriptive in nature, a variety of process interpretations for the SCM have been proposed (e.g. Marley 1992; Nosofsky 1990; Smith 1980; Townsend & Landon 1983).

The SCM usually provides excellent descriptions of the detailed quantitative structure of identification-confusion matrixes. However, assuming n stimuli, fitting the model requires estimation of $n(n - 1)/2$ freely varying similarity parameters (one similarity parameter for each pair of unique stimuli), and $n - 1$ freely varying bias parameters. Furthermore, simply fitting the full version of the model provides little insight into the psychological processes and similarity structure that underlie identification performance.

A vast reduction in the number of free parameters and a deeper understanding of identification processes can be achieved by testing and comparing restricted versions of the SCM in which theories of similarity are used to constrain the η_{ij} parameters. One classic approach, initiated in Shepard's (1957, 1958b) original formulation of the model, is to derive an MDS solution for the set of stimuli, and assume that the η_{ij} parameters are functionally related to distances in the derived scaling solution. Systematic comparisons among different versions of this MDS-choice model can provide insights into the underlying dimensions of the stimuli, the metric structure of the psychological space, the extent to which values on different dimensions are perceived independently of one another, and so forth.

As one example, Nosofsky (1985b) collected confusion data in which two subjects identified stimuli varying along two continuous dimensions (size and angle). There were four orthogonally varying values per dimension, yielding a 16-member stimulus set (and, therefore, a 256-cell identification-confusion matrix). Nosofsky (1985b) found that by representing each stimulus as a point in a two-dimensional psychological space, computing similarities between stimuli on the basis of their distance in the space, and substituting these similarities into the SCM response rule (Equation 1), excellent predictions of the identification-confusion data could be achieved. Indeed, the fits of this

MDS-choice model were not significantly worse than those of the full SCM for either subject, suggesting that the MDS solution provided a precise quantitative account of the similarity structure inherent in each subject's data. Moreover, for one of the subjects, a constrained MDS-choice model with only six freely varying MDS coordinate parameters accounted for the data essentially as well as the full SCM with 120 freely varying similarity parameters. In this constrained model, all stimuli with a given physical value of angle were assumed to have the same psychological value on the angle dimension, and likewise for the size dimension. The excellent fits of this constrained MDS-choice model provided evidence that the subject perceived the size and angle dimensions in a separable manner. Other illustrative applications of the MDS-choice model are provided by Shepard (1958b), Getty et al (1979), Nosofsky (1985a, 1987, 1989), Takane & Shibayama (1985), and Heiser (1988).

CATEGORIZATION   A classic issue in cognitive psychology is whether the principles of stimulus generalization and similarity that underlie identification performance also underlie categorization performance. Indeed, perhaps the most straightforward view of categorization, formalized in what are known today as *exemplar models* (e.g. Estes 1986; Hintzman 1986; Medin & Schaffer 1978; Nosofsky 1986), is that classification of an object is determined by how similar it is to the individual members of alternative categories.

Seminal investigations of this idea were conducted by Shepard et al (1961) and Shepard & Chang (1963). These researchers measured similarities among the individual objects in a set in terms of the probability of pairwise confusions in identification learning paradigms. The measured similarities were then used to quantitatively predict the difficulty of learning different category structures. Intuitively, if an exemplar-based generalization view is correct, it should be easier to learn structures in which within-category similarities among objects are large, and between-category similarities are small. In a situation involving relatively unanalyzable, integral-dimension stimuli, Shepard & Chang (1963) found that the difficulty of learning different category structures could indeed be predicted well on the basis of pairwise confusions in identification learning tasks. But in a situation involving highly analyzable, separable-dimension stimuli there were systematic failures of the exemplar-based generalization hypothesis. Shepard et al (1961) attributed these failures to the intervention of a selective-attention process, in which subjects focused attention on those dimensions of the stimuli that were relevant to solving a given categorization problem. Such a selective-attention process should be particularly efficient for separable-dimension stimuli.

Nosofsky (1984, 1986) formalized these early ideas involving exemplar-based generalization and selective attention within an integrated model. This

model, which is a generalization of the *context model* of categorization
proposed by Medin & Schaffer (1978), builds directly on the multi-
dimensional scaling–SCM framework discussed in the previous section.

According to the generalized context model (GCM), the evidence favoring
Category J given presentation of stimulus i is found by summing the
(weighted) similarity of stimulus i to all exemplars of Category J, and then
multiplying by the response bias for Category J. This evidence is then divided
by the sum of evidences for all categories to predict the conditional probabil-
ity with which stimulus i is classified in Category J:

$$P(R_J \mid S_i) = (b_J \sum_{j \in C_J} M_j \, \eta_{ij}) / (\sum_K b_K \sum_{k \in C_K} M_k \, \eta_{ik}). \qquad\qquad 2.$$

where η_{ij} denotes the similarity between exemplars i and j; b_J denotes the
Category J response bias; and M_j denotes the strength with which exemplar j
is stored in memory.

The relation between the decision rules in the GCM (Equation 2) and the
SCM (Equation 1) is readily apparent. However, because of the selective-
attention processes discussed by Shepard et al (1961), Shepard (1964), Tver-
sky (1977), Garner (1974), Medin & Schaffer (1978), and others, the η_{ij}
similarity parameters in Equations 1 and 2 may be not invariant across the
identification and categorization paradigms.

Nosofsky (1984, 1986) adopted the Individual Differences Scaling
approach to multidimensional scaling (Carroll & Wish 1974) as a theory for
explaining attention-based changes in similarities. The distance between ex-
emplars i and j (d_{ij}) in a multidimensional psychological space is given by

$$d_{ij} = [\sum_m w_m \mid x_{im} - x_{jm} \mid^r]^{1/r}, \qquad\qquad 3.$$

where x_{im} is the psychological value of exemplar i on dimension m; the value
of r defines the distance metric (e.g. r = 1, city-block; r = 2, Euclidean); and
w_m ($0 \leq w_m$, $\Sigma w_m = 1$) is the "attention-weight" given to dimension m.
Large values of w_m serve to "stretch" the psychological space along dimen-
sion m, and small values of w_m serve to "shrink" the space along that
dimension. The distance d_{ij} is converted to a similarity measure by using the
function

$$\eta_{ij} = \exp(-c \cdot d_{ij}{}^p), \qquad\qquad 4.$$

where c is a general sensitivity parameter; and the value of p defines the
similarity gradient (e.g. p = 1, exponential; p = 2, Gaussian).

The general approach to predicting and relating identification and

categorization in terms of this multidimensional scaling framework is as follows. First, by fitting the MDS-choice model (Equations 1, 3, and 4) to a set of identification-confusion data, a maximum-likelihood MDS solution is derived for a set of stimuli. This MDS solution can then be used in conjunction with the GCM (Equations 2, 3, and 4) to predict performance in any given categorization paradigm involving the same set of stimuli. Because the MDS solution will have been derived from the identification-confusion data, a minimum of parameters remain to be estimated for predicting categorization. The critical parameters tend to be the weights (w_m) in the distance function (Equation 3), which describe the role of the selective attention process in modifying similarities across identification and categorization.

Nosofsky (e.g. 1984, 1986, 1987, 1989, 1991c) has demonstrated numerous successful quantitative applications of the GCM, in situations involving both integral and separable-dimension stimuli. These demonstrations are important because they illustrate that the fundamental processes of identification and categorization can be understood within a unified theoretical framework, and that precise quantitative predictions of performance in each paradigm can be achieved within this framework. Furthermore, the predictions of categorization are achieved with a minimum of parameter estimation. Finally, the estimated attention-weight parameters vary in psychologically meaningful ways. In particular, Nosofsky (e.g. 1984, 1986, 1991c) has provided evidence that subjects often distribute attention over psychological dimensions so as to nearly optimize their categorization performance—i.e. maximize their average percentage of correct categorization choices. A variety of mechanistic models have recently been proposed for how the attention weights in the GCM may be learned trial by trial (e.g. Hurwitz 1990; Kruschke 1990).

The role of MDS in developing these theoretical relations is critical. Note that it is not "similarity" that is invariant across identification and categorization; rather, the MDS solution for the stimuli is invariant. Because of the selective-attention processes assumed to operate on the scaling representation, similarities among exemplars are systematically modified.

The MDS-based exemplar model accounts successfully for the effects of a number of fundamental variables on categorization performance. As one example, Nosofsky (1988c, 1991c) established experimental learning conditions in which the frequency of individual exemplars was manipulated. In the GCM, increasing the frequency of an exemplar is assumed to increase its "strength" in memory. Exemplar memory-strength is modeled by the M_j parameters in Equation 2. Because memory strength combines multiplicatively with interexemplar similarity, the GCM predicts an interactive effect of frequency and similarity on categorization performance. The interactive effect is observed: Classification accuracy and confidence increase

for exemplars that are presented with high frequency, and for items that are similar to the high-frequency exemplars. Little effect of frequency occurs for items that are dissimilar to the high-frequency exemplars. It is as if the high-frequency exemplar acts as a "magnet" in the psychological space, drawing nearby objects toward it.

Alternative models of categorization can also be formulated within an MDS framework. According to prototype models, classification is determined by the similarity of an item to the central tendency of the distributions of category exemplars in the multidimensionally scaled psychological space (e.g. Nosofsky 1987, 1991c; Reed 1972; Shin 1990). Prototype models tend not to fare as well as exemplar models, however, in their quantitative predictions of classification performance (see Nosofsky 1992, for a review).

The fuzzy logical model of perception (FLMP) of Massaro, Oden, and their colleagues (e.g. Massaro 1987; Massaro & Friedman 1990; Oden & Massaro 1978) can also be construed as an MDS-based prototype model, although here the prototype of a category is defined more generally as an "ideal point" in the psychological space rather than as the central tendency. In a typical experimental paradigm for testing the FLMP, the stimuli vary along M orthogonal continuous dimensions, and the subject is required to classify each object into one of K categories. According to the model, the probability that stimulus i is classified in Category J is given by

$$P(R_J \mid S_i) = \eta_{iPJ} / \Sigma \; \eta_{iPK}, \qquad\qquad\qquad 5.$$

where η_{iPJ} denotes the similarity (or "fuzzy logical degree of match") of stimulus i to the prototype of Category J. This similarity is given by the multiplicative rule

$$\eta_{iPJ} = \Pi \; s(i_m, J_m), \qquad\qquad\qquad 6.$$

where $s(i_m, J_m)$ denotes the similarity of stimulus i to Prototype J on dimension m. This interdimensional multiplicative rule for computing similarities between exemplars was also proposed by Medin & Schaffer (1978) in their original formulation of the context model, although they restricted attention to binary-valued dimensions.

The FLMP has accounted impressively for numerous phenomena involving forms of information integration in diverse domains (Massaro 1987). In most previous applications of the model, however, all the individual $s(i_m, J_M)$ values were treated essentially as free parameters. As noted by Nosofsky (1984, 1986), the multiplicative similarity rule (Equation 6) has a natural MDS interpretation that would allow for a much more parsimonious application of the FLMP. In particular, an interdimensional multiplicative rule arises

whenever $p = r$ in Equations 3 and 4. For example, when distance in psychological space is described by a city-block metric ($r = 1$), and similarity is an exponential decay function of psychological distance ($p = 1$), then we would have

$$\eta_{iPJ} = \exp\left[-c \left(\Sigma \ w_m \ | \ x_{im}\text{-}x_{Jm} \ |\right)\right]$$
$$= \Pi \exp\left(-c \cdot w_m \ | \ x_{im}\text{-}x_{Jm} \ |\right), \qquad\qquad 7.$$

which is Equation 6 with $s(i_m, J_m) = \exp(- c \cdot w_m \ | \ x_{im}\text{-}x_{Jm} \ |)$. Thus, by deriving in independent tasks similarity-scaling solutions for the objects under study, the FLMP could be applied to predict categorization with a minimum of parameter estimation.

Recently, Anderson (1990) proposed a rational model of categorization. According to the rational model, exemplars are grouped into clusters during the category learning process. The probability that an exemplar joins a cluster is determined jointly by the current size of each cluster, the similarity of the exemplar to the cluster's central tendency, and the value of a "coupling" parameter, which is a free parameter in the model. There are also mechanisms in the model for determining the probability that membership in each cluster signals a given category label. Roughly, the probability that stimulus i is classified in Category J is found by summing the similarity of i to each cluster's central tendency, weighted by the category-label J probability associated with the cluster. Similarity to the central tendency of each cluster is computed by using a multiplicative-similarity rule that is isomorphic to the one assumed in the context model and the FLMP. With the addition of some technical assumptions, Nosofsky (1991a) proved that in domains involving binary-valued dimensions, the rational model generalizes both the context model and the FLMP. Intuitively, when the value of the coupling parameter is zero, each exemplar forms its own cluster, and the rational model becomes the context model. By contrast, when the value of the coupling parameter is unity, the clusters that are formed correspond to prototypes for each of the experimentally defined categories, and the rational model is essentially the FLMP. For intermediate values of the coupling parameter, the rational model functions as a multiple-prototype model. A natural direction of future research will involve the use of MDS techniques in conjunction with the rational model, as I have described previously for the context model and the FLMP.

RECOGNITION   The MDS-based exemplar model (the GCM) has also been used to model old-new recognition memory performance (Nosofsky 1988a, 1991c; Nosofsky et al 1989). Following previous investigators (e.g. Gillund & Shiffrin 1984; Hintzman 1986), Nosofsky made the central assumption that recognition judgments are based on the overall summed similarity of an item

to all exemplars stored in memory. This summed similarity gives a measure of overall "familiarity," with higher familiarity values leading to higher recognition probabilities. Specifically, the familiarity for item i (F_i) is given by

$$F_i = \Sigma \Sigma M_k \, \eta_{ik}, \qquad\qquad 8.$$

where the M_k and η_{ik} parameters are defined as before (see Equation 2), and the sum is over all exemplars stored in memory. Nosofsky (1991c) demonstrated that by deriving MDS solutions for sets of objects, and using these MDS solutions in conjunction with the model (Equations 3, 4, and 8), fine-grained differences in old-new recognition judgments could be predicted on the basis of fine-grained differences in similarities among items.

Note that categorization and recognition are presumed to involve different decision rules. According to the exemplar model, categorization decisions involve a relative-similarity rule (Equation 2), whereas recognition decisions involve an absolute-similarity rule (Equation 8). Thus, the exemplar model can predict markedly different patterns of performance across the two tasks, as are often observed. However, a unified account of categorization and recognition is provided by the model in the sense that both judgments are assumed to be based on the similarity of an item to the exemplars in a multidimensionally scaled psychological space.

Same-Different Judgments and Reaction Time

Takane & Sergent (1983) and Sergent & Takane (1987) proposed and tested a scaling-based process model for jointly characterizing accuracy and reaction time data in "same-different" judgment tasks. The model has three main components. The representation component specifies the function used to compute distances among objects in a psychological space. Takane & Sergent assume that error is introduced into these distance judgments. The error component of their model specifies the nature of the error perturbations operating on the distances. The distribution of error perturbations is assumed to be log-normal in form, with variance that increases as the true distance increases. [This assumption is the same as the one used by Ramsay (1977) in his maximum-likelihood method for scaling similarity judgments.] Finally, the response component of their model relates observed reaction times and same-different judgments to the error-perturbed distances. If the judged distance exceeds a threshold then a "different" response is made, else a "same" response is made. Based on the log-normal assumption for the distribution of errors, the log of the distribution of "different" reaction times for each stimulus pair is assumed to be normal in form, with a mean that decreases linearly with the difference between the (log) distance and (log) threshold. Thus, "different" RTs get faster as stimuli become more dissimilar. By

contrast, the distribution of "same" RTs is assumed to have a mean that increases with the difference between (log) distance and (log) threshold, reflecting Podgorny & Garner's (1979) finding that "same" RT increases as stimuli become more dissimilar.

Using maximum-likelihood methods, Sergent & Takane (1987) fitted the model to same-different data obtained for a variety of stimulus sets. One of the central purposes of their study was to gain "information about similarity structure of stimulus sets as they actually emerge under conditions of speeded judgment process" (Sergent & Takane 1987:312). The argument is that similarity structure and the nature of dimensional interactions may be functions not only of stimulus characteristics, but also of perceptual processes. Similarity relations among objects may differ depending on whether the objects are processed under speeded or unspeeded conditions. Indeed, Sergent & Takane (1987) found that under their process-limited conditions, the best-fitting distance metric for a set of separable-dimension stimuli (circles varying in size, and in the orientation of a radial line) was Euclidean rather than city-block, in contrast to the usual finding obtained under process-unlimited conditions (for similar evidence, see Nosofsky 1985b).

Cognitive Processes and the Metric Axioms

In their well-known and elegant work, Tversky and his colleagues have called into question the psychological validity of the fundamental metric axioms underlying traditional MDS approaches (e.g. Gati & Tversky 1982; Tversky 1977; Tversky & Gati 1982). Using an extensive array of similarity data, including direct judgments and recognition confusions, Tversky's demonstrations suggest, for example, that similarities can be asymmetric, that stimuli can have differing degrees of self-similarity, and that similarity data often entail violations of the triangle inequality. As an alternative to spatial MDS models, Tversky (1977) proposed a set-theoretic model of similarity based on feature matching, which has been extremely influential and widely used, and which can account for the patterns of similarity data noted above.

I believe that some of the force of Tversky's demonstrations is diminished, however, when MDS representations are viewed as components of cognitive process models. As I have argued previously, observed behavior reflects only indirectly the underlying similarity representation. Process models that incorporate symmetric-similarity representations can predict asymmetric patterns of proximity data. A straightforward example involves identification-confusion data, which are often highly asymmetric (e.g. the probability of identifying object i as object j may be far greater than the probability of identifying object j as object i). Despite these asymmetries, the symmetric-similarity SCM usually accounts accurately for the structure of identification-confusion matrixes. It accounts for the asymmetries by virtue of the bias

parameters in the model (Luce 1963; Shepard 1957), as well as the nature of the decision rule itself (e.g. see Getty et al 1979).

A very general reason why similarity data are often asymmetric may be that in addition to the role of pairwise similarities, properties of individual objects play a fundamental role in cognitive processes. For example, suppose that in a categorization experiment a particular exemplar is presented with high frequency. According to the GCM, the exemplar receives a strong memory representation, and the strength with which that individual item is stored in memory plays a fundamental role in subsequent classification. According to the model, a strong item is activated by a weak item far more than the weak item is activated by the strong, leading to asymmetries in classification behavior.

Holman (1979) presented a series of hierarchically organized models for describing asymmetric proximity data. These models incorporate a symmetric-similarity function together with individual item bias functions. "Bias" is defined very generally as a property associated with an individual object. According to one of the stronger models he presents, the proximity of i to j [p(i,j)] is given by

$$p(i,j) = F[s(i,j) + r(i) + c(j)], \qquad\qquad 9.$$

where $s(i,j)$ is the symmetric similarity between i and j, $r(i)$ is the "row" bias for item i (the "subject" of the object pair), $c(j)$ is the "column" bias for item j (the "referent" of the object pair), and F is an increasing function. In general, we have $p(i,j) > p(j,i)$ whenever $r(i) + c(j) > r(j) + c(i)$. Various models that have successfully accounted for asymmetric proximities are special cases of this "additive similarity and bias model," including the additive version of Tversky's (1977) feature-contrast model, Krumhansl's (1978) distance-density model, and the SCM for predicting identification confusions. Carroll's (1976) hybrid model, which combines spatial and hierarchical components, is a symmetric special case of Equation 9. Nosofsky (1991b) reviews a wide variety of phenomena involving asymmetric proximities that appear to be readily interpretable in terms of symmetric similarities together with individual item biases, as described by Holman's (1979) model. [It should be noted in this section, however, that in tests of Krumhansl's (1978) model, Corter (1987) conducted a series of experimental manipulations involving stimulus density, but failed to observe effects of this variable on similarity judgments.]

Self-proximities are also bound to be influenced by properties of the individual objects. For example, in a same-different judgment task, it should take more time to respond "same" for a complex object than for a simple one. In their modeling of same-different judgments, Takane & Sergent (1983)

discuss a representation based on Equation 9 in which the bias terms are assumed to reflect stimulus complexity.

Another diagnostic used to question the psychological validity of MDS models is nearest-neighbor analysis of proximity data (Tversky & Hutchinson 1986). Low-dimensional spatial solutions are unable to account for patterns of proximity data in which a single item is the nearest neighbor (most proximal) to many other items in the set. Such data arise frequently in semantic domains that include a single focal element such as the superordinate of a category. However, as noted by Tversky & Hutchinson (1986), by augmenting the spatial representation with individual item-bias components to model the hierarchical structure of the set, one can readily account for such patterns of proximity data. One interpretation is that above and beyond "similarity," properties of individual objects play a fundamental role in cognitive processes.

According to the triangle inequality, for any three points a, b, and c, the psychological distance from a to c must be less than or equal to the sum of the distances from a to b and b to c. Although the triangle inequality cannot be tested directly on the basis of ordinal data, in a clever experimental design Tversky & Gati (1982) were able to infer systematic violations of the triangle inequality. These violations occurred in situations involving highly separable--dimension stimuli, in which objects a and b coincided on one dimension, and b and c coincided on a second dimension. Tversky & Gati provided corroborating evidence of these qualitative violations in a series of MDS analyses that showed that a value of $r < 1$ in the Minkowski power model (Equation 3) yielded a best fit to the similarity data. A process-interpretation for $r < 1$ is that, in making their similarity judgments, subjects systematically give greater attention weight to those dimensions along which stimuli are more similar (Tversky & Gati 1982:150). This process interpretation is consistent with Sjoberg & Thorslund's (1979) suggestion that, in making similarity judgments, subjects carry out an active search for the ways stimuli are similar.

Summary

In the first part of this chapter I discussed MDS-based models for predicting a variety of performances, including generalization, identification, categorization, recognition, same-different accuracy and reaction time, and similarity judgment. The MDS-based similarity representation is a fundamental component of these models. In the case of categorization, for example, it is important to specify whether the representation consists of a prototype, multiple prototypes, individual exemplars, and so forth. Furthermore, to apply the models, the representational objects must be located as points in the psychological scaling solution. But a complete account of performance in

each task also requires specification of the cognitive processes that operate on the similarity representation. Some of the critical processes that were discussed were the nature of the decision rule, the role of selective attention in modifying the structure of the psychological space, and the influence of individual item properties such as memory strength. Testing the process models and deriving the scaling representations are mutually dependent, and one cannot proceed without the other.

Even after specifying the processing mechanisms, however, a potential shortcoming of all the models just reviewed is that they involve deterministic scaling representations, in which each object is represented as a single point in the psychological space. More general cognitive-process models make use of probabilistic scaling representations, which I review in the second part of this chapter.

PROBABILISTIC MULTIDIMENSIONAL SCALING APPROACHES

Probabilistic MDS models represent individual objects as probabilistic distributions of points in a multidimensional space, an extension of Thurstone's (1927a) classic framework for scaling unidimensional psychological magnitudes. As in deterministic models, presentation of an object is assumed to give rise to some internal representation. Because of noise in the system, however, the same internal representation is not yielded on every trial. Rather, across trials, presentation of an object gives rise to a probabilistic distribution of internal representations. Conceptually, such probabilistic representations are necessary in situations in which there is a good deal of noise in the perceptual processing system. Also, probabilistic models are needed for situations in which there is uncertainty in subjects' memory for the previously presented objects, as might occur because of diffusion of memory traces over time.

Each of the deterministic MDS models discussed in Part 1 of this chapter can be generalized by allowing the single-point representations of the objects to become probabilistic in nature. In addition, once one allows probabilistic representations, a variety of new process models suggest themselves. In the following, I focus primarily on these new models.

Probabilistic Scaling of Distance Judgments

Zinnes & MacKay (1983) developed maximum-likelihood procedures to obtain estimates of the parameters in the Hefner (1958) model. In this model, each object is represented as an n-dimensional random-vector, where the values on each dimension have been drawn at random from independent normal distributions of equal variance. Thus, each object is characterized on each dimension by a location parameter (the mean of the distribution) and a

variability parameter. Although a given stimulus is assumed to have the same variance on each dimension, the variance associated with different stimuli can be unequal. When the model is applied at the level of individual subjects, one interpretation of the variance parameter is that it represents the subject's level of unfamiliarity with, or uncertainty about, the nature of the stimulus.

Because each stimulus has the same variance on each of its dimensions, the Hefner (1958) model is isotropic, in the sense that there are no dominant directions in the space. MacKay (1989) generalized the model to allow each stimulus to have different variances on each of its dimensions, yielding an anisotropic model. In addition, he allowed the coordinates of each stimulus to be correlated. Techniques for obtaining maximum-likelihood estimates of the parameters were proposed and tested.

It is assumed in these models that in judging the distance between objects i and j, a point from each of the object distributions is randomly and independently sampled, and the Euclidean distance between the points is computed. This momentary distance, d_{ij}, is a random variable, and is conceptually distinct from the distance between the means of the object distributions (D_{ij}), which Zinnes & MacKay (1983) term the "true" distance. The expected value of d_{ij}, $E(d_{ij})$, also differs from the true distance D_{ij}. Indeed, even if D_{ij} is zero, $E(d_{ij})$ will become indefinitely large as the variance of the object distributions approaches infinity.

Thus, in the Hefner model, the expected distance between objects is not related to the "true" distance in a simple, monotonic way. This property of nonmonotonicity can lead to highly pathological solutions if a deterministic MDS algorithm is used to analyze data generated from a probabilistic MDS process. As one example, Zinnes & MacKay (1983) constructed a configuration in which the objects were positioned along an inner and an outer hexagon, the variances of the points forming the inner hexagon being larger than those forming the outer hexagon. Simulated distance judgments were then used as input to a nonmetric (deterministic) scaling program and to the maximum-likelihood (ML) procedure developed by Zinnes & MacKay. The ML procedure accurately recovered the true configuration, but the deterministic model actually interchanged the positions of the inner and outer hexagons. The reason is that the expected value of the interpoint distances strongly reflected the large variances of the inner hexagon stimuli, so the deterministic program incorrectly "perceived" the inner hexagon to be large.

In general, by fitting alternative restricted versions of the general anisotropic model to sets of distance judgments, and systematically comparing the fits, one can statistically test hypotheses concerning the dimensionality of the space, the values of the coordinates, whether the space is isotropic or anisotropic, and whether individual stimuli have common variance-covariance structures.

General Recognition Theory

The general recognition theory (GRT) of Ashby, Townsend, and their associates (e.g. Ashby & Perrin 1988; Ashby & Townsend 1986; Kadlec & Townsend 1992) is a multidimensional generalization of signal detection theory (Green & Swets 1966) and of Thurstone's (1927b) law of categorical judgment. Besides assuming probabilistic internal representations, the critical assumption in this theory is that the observer establishes *decision boundaries* to partition the psychological space into response regions. Any internal representation or "perceptual effect" falling in Region A would lead to an A response. Most applications of the GRT have assumed that the perceptual effects are distributed as multivariate normal random variables, an assumption that I make in the following discussion.

FUNDAMENTAL CONSTRUCTS   Ashby & Townsend (1986) discuss a variety of fundamental constructs and their interrelations within the framework of the GRT. For simplicity, imagine a complete identification experiment in which there are two physically manipulated dimensions, A and B, with r levels on dimension A and q levels on dimension B that are factorially combined. Assume further that the psychological dimensions along which the objects are represented correspond to the physically manipulated dimensions. Thus, over trials, each stimulus gives rise to a bivariate normal distribution. On each trial the subject is required to identify the level on each dimension of the presented stimulus (or provide an informationally equivalent response).

Perceptual independence for a pair of dimensions in a particular stimulus holds if the perceptual effects of the two dimensions are statistically independent, which, for the bivariate normal distribution, occurs if there is zero correlation between the perceptual effects on each dimension. Note that perceptual independence is a property of an individual stimulus.

Perceptual separability holds if, across stimuli, the perceptual effects of a given level of one dimension do not depend on the level of the other dimension. Consider, for example, the set of stimuli A_iB_j constructed from dimension A at level i and dimension B at level j. Dimension A would be perceptually separable from dimension B if, for each i, the perceptual effects of A_i do not depend on the level of B. In the case of the normal distribution, this property holds if, for each i, the stimuli A_iB_1, A_iB_2, ..., A_iB_q have the same mean and variance on dimension A. Note that dimension A can be perceptually separable from dimension B without the converse relation holding. Also, whereas perceptual independence is a property pertaining to an individual stimulus, perceptual separability is a property pertaining to a set of stimuli.

Decisional separability on dimension A holds if a subject's decision about the level of dimension A does not depend on the value of the perceptual effect

associated with dimension B. This property holds if the subject's decision boundaries are perpendicular to the Dimension A coordinate axis (or, equivalently, parallel to the Dimension B axis). As is the case for perceptual separability, note that decisional separability can hold on Dimension A without holding on Dimension B, and vice versa.

Perceptual independence, perceptual separability, and decisional separability are all logically independent from one another. However, Ashby & Townsend (1986) and Kadlec & Townsend (1992) prove a number of fundamental theorems that allow the constructs to be interrelated by means of observable response probabilities in an identification experiment. For example, in an identification experiment with two levels on each of the two dimensions, *sampling independence* in stimulus A_iB_j holds if the probability of A2 and B2 both being reported given presentation of stimulus A_iB_j is equal to the product of the individual probabilities of A_2 being reported and B_2 being reported (given stimulus A_iB_j). Ashby & Townsend (1986) prove that if decisional separability holds on both dimensions, then sampling independence is equivalent to perceptual independence. This simple example is intended to give only a flavor of the rich web of interrelated concepts that the GRT provides for investigating the structure of subjects' internal representations of multidimensional stimuli. Other methods for investigating the properties of perceptual independence, perceptual separability, and decisional separability are discussed and illustrated by Ashby (1988) and Wickens & Olzak (1989).

MODELS OF CLASSIFICATION   The GRT provides a powerful and flexible language for expressing numerous different models of stimulus classification. These models differ in terms of the types of decision boundaries that the subject uses for partitioning the multidimensional space into response regions.

Ashby & Gott (1988) distinguish between independent-decisions boundaries and several types of information-integration boundaries (cf Shaw 1982). Imagine, for example, that there are two categories, A and B, composed of objects varying on two dimensions. Both categories of objects are distributed as bivariate normal random variables, with members of Category A tending to have low values on both of dimensions 1 and 2, and members of Category B having high values on both dimensions. According to an independent-decisions model, the subject would establish a separate criterion on each dimension for partitioning low versus high values. Given presentation of a stimulus, separate decisions would be made about its value on each dimension, and these decisions would then be combined in making a response. "Low-low" decisions would result in a Category A response and high-high decisions would result in a Category B response. Low-high and high-low decisions provide ambiguous information, so the subject would be forced to

guess. In terms of the GRT, this decision strategy corresponds to establishing two orthogonal boundaries that are parallel to the coordinate axes (i.e. decisional separability holds on both dimensions). Percepts falling in the lower-left quadrant would be classified in Category A, whereas percepts falling in the upper-right quadrant would be classified in Category B. Percepts falling in the remaining two quadrants provide ambiguous information and the subject must guess.

By contrast, according to information-integration models, subjects are able to combine information from both dimensions into an integrated percept, and a single decision is then made with regard to that integrated information. Ashby & Gott (1988) discuss a variety of information-integration models in terms of the types of decision boundaries they entail. A *minimum distance boundary* is a linear boundary that bisects and is perpendicular to the segment that connects the central tendencies (prototypes) of Categories A and B. Minimum distance bounds arise when classification decisions are based on distance to the prototype: If the percept is closer to the prototype of Category A, then respond A; else respond B. *General linear boundaries* generalize minimum distance bounds by allowing the slope and y-intercept of the linear boundary to be free parameters. These boundaries can be interpreted in terms of a (biased) prototype model in which differential weight is given to each dimension in calculating distance. *Optimal boundaries* (that maximize probability of correct classification) are those in which the subject computes the overall likelihood of the percept coming from Category distribution A or B, and responds with the category with greater likelihood. There are close formal relations between these optimal likelihood–based boundaries and the decision boundaries that are predicted by certain types of exemplar storage models (Estes 1986; Nosofsky 1990).

Ashby & Gott (1988) and Ashby & Maddox (1990, 1992) have conducted a number of experimental studies to investigate the types of decision boundaries that subjects adopt. Using a procedure known as the general recognition randomization technique, which involves the systematic addition of multivariate external noise to the prototypes of each category, they have obtained convincing evidence that in their paradigm: 1. subjects adopt information-integration strategies rather than independent-decisions strategies (Ashby & Gott 1988), even if the underlying perceptual dimensions are highly separable in nature (Ashby & Maddox 1990); 2. if given sufficient motivation and training, subjects can adopt decision boundaries that are highly nonlinear, and sometimes close to optimal (Ashby & Maddox 1992); and 3. rather than using probabilistic decision rules, subjects use decision rules that are deterministic in nature (or close to it). The latter finding means that each percept in the psychological space has an associated category response probability that is essentially 0 or 1, in contrast to the predictions of models that postulate

competing response tendencies such as Nosofsky's (1986) GCM. Finally, in recent work, Ashby & Lee (1991) demonstrated successful applications in which versions of the GRT performed as well or better than the SCM and GCM at predicting identification and categorization data. These applications were in standard designs that did not involve the introduction of external noise.

SIMILARITY   Ashby & Perrin (1988) proposed to model similarity judgments in terms of the GRT by assuming that the judged similarity of A to B is related to the proportion of the A distribution that overlaps the B response region. A virtue of the model is that it contains the general Euclidean scaling model (Young 1984b) as a special case. For example, in the GRT, differential weighting of dimensions corresponds to differential variances of the distributions of perceptual effects, and oblique dimensions correspond to dependencies (correlations) in the distributions of perceptual effects. Unlike the general Euclidean scaling model, however, the GRT similarity model is not constrained by the metric axioms. Ashby & Perrin (1988) demonstrated support for the model by conducting an experiment in which distance between the prototypes of distributions A and B was held constant, but overlap between the A and B distributions was varied across conditions. Overall similarity judgments were observed to increase as the proportion of overlap increased.

In my view, this application of the GRT seems reasonable as a model of the similarity between categories, or as a model of similarity between objects with substantial variability. But in numerous experimental situations, one judges the similarity between pairs of individual objects with essentially no psychological variability. The applicability of the GRT similarity model in these situations seems more limited. Another interesting challenge for the model would be to explain the exponential gradient of similarity discussed by Shepard (1987), as well as why the metric of psychological space depends systematically on the types of dimensions that compose the stimuli.

COMPARING THE GCM AND THE GRT   Because the GCM and the GRT are two MDS-based models that have been applied rigorously in recent years to relate similarity, identification, and categorization data, it is of some interest to compare and contrast them. First, in the GCM, each object is represented as a single point in psychological space, whereas the GRT represents each object as a probabilistic distribution of points. In situations involving substantial perceptual or memorial variability, the single-point assumption of the GCM clearly needs to be modified. Second, the GCM assumes a probabilistic decision rule, whereas the GRT incorporates a deterministic decision rule. Ashby & Gott (1988) provided convincing evidence of the use of deterministic decision rules in experiments involving the recognition randomization

technique, but the generalizability of these results to more standard designs is open to question. It may well be that the use of probabilistic versus deterministic decision rules depends on the experimental situation.

The most fundamental difference between the GCM and the GRT concerns the presumed nature of the category representation. In the GCM it is assumed that people classify items on the basis of their summed similarity to the exemplars of alternative categories. By contrast, in the GRT, it is assumed that people form "decision boundaries" to partition the multidimensional space into response regions. The GRT should be viewed as providing a general and powerful language for expressing alternative models of classification. To use the GRT to predict classification probabilities, one needs to specify the types of decision boundaries that the subject uses to partition the multidimensional space. In recent work, Ashby & Maddox (1992) propose that subjects adopt quadratic bounds, which is the form that likelihood-ratio bounds take when the category distributions are normal in form. They have also discussed (mainly as foils) independent-decisions bounds, minimum distance bounds, general linear bounds, and bilinear bounds. My view is that each different type of boundary that is assumed constitutes an alternative model of classification. An infinite variety of such models is available within the general framework provided by the GRT. Indeed, one could formulate an exemplar-similarity model in its framework by assuming an exemplar-similarity boundary: The decision rule is to classify a percept into Category A if its summed similarity to the Category A exemplars exceeds its summed similarity to the Category B exemplars, else classify it in Category B. Thus, with modifications in some of the technical differences noted above, the exemplar-based GCM can be expressed within the language of the GRT.

Multivariate Discrimination Methods

Ennis, Mullen, and their colleagues (e.g. Ennis 1988, 1992; Ennis & Mullen 1986; Ennis et al 1988; Mullen & Ennis 1987 have developed a number of multivariate models for discrimination and grouping methods, such as the duo-trio method and the triangular method. They have also extended these models to account for same-different judgments and identification performance. Among other things, Ennis (1988; Ennis et al 1988) showed how these models could be used to reconcile Nosofsky's (1985a,b) observations of a Gaussian similarity gradient with Shepard's (1987) proposed universal law of generalization.

I illustrate the nature of the discrimination modelling by reviewing Ennis & Mullen's (1986) multivariate Euclidean model for the triangular method. In the triangular method, the subject is instructed to select out of three stimuli (two sampled from one stimulus distribution and one from another) the stimulus that is perceptually different from the other two. In the Ennis &

Mullen (1986) model, each stimulus distribution is assumed to be multivariate normal in form. The stimuli sampled from each distribution are assumed to be mutually independently distributed. The decision rule is to group together the two stimuli that are the shortest Euclidean distance apart. A correct response occurs if these shortest-distance stimuli were the ones that were sampled from the same distribution.

Ennis & Mullen (1986) developed a mathematical formulation of the triangular-method model for the bivariate case, and used Monte Carlo simulations to evaluate the more general multivariate model. Of most general conceptual importance regarding their findings was that discrimination performance is not a function solely of the distance between the means of the stimulus distributions, but depends critically on such characteristics of the distributions as their dimensionality, correlation structure, relative orientation, and variances.

In an extension of these methods to account for same-different judgments, similarity, and identification, Ennis (1988, 1992; Ennis et al 1988) combined assumptions about the stochastic, multivariate representation of the stimulus objects with the kinds of distance-based similarity judgments assumed in the models of Shepard (1957, 1987) and Nosofsky (1986). Assume that a pair of stimuli has been presented and the subject must judge whether they are the "same" or "different." As described previously, it is assumed that each stimulus gives rise to a momentary psychological representation (i.e. a point) in the perceptual space. The distance (d) between these points is computed by using the Minkowski power model, and the similarity between the objects is then given by $g(d) = \exp(-d^{\alpha})$, where $\alpha > 0$ (cf Nosofsky 1986; Shepard 1957, 1987). In one version of the same-different model, Ennis et al (1988) take $g(d)$ to be the (unbiased) probability that the subject judges the pair of stimuli to be the "same" on the given trial.

To predict the probability that a pair of stimuli is judged "same" during the course of the experiment, one would compute the expected value of $g(d)$, $E[g(d)]$. (Note that because the stimulus representations are stochastic, the distance between stimuli is a random variable in the model.) Ennis et al (1988) provide expressions for $E[g(d)]$ in the case in which the stimuli are distributed as multivariate normal random variables. They also illustrate that the parameters of the stimulus distributions can be accurately recovered by fitting the model to generated matrixes of same-different judgments. Thus, the model provides a viable approach to obtaining probabilistic MDS solutions for sets of multidimensional stimuli.

In further analyses, Ennis et al (1988) investigated the effect of the multivariate stochastic portion of the model on the presumed form of the similarity gradient. In particular, suppose that one modeled a set of similarity data by using a deterministic MDS model, but that the similarity data had actually

been generated by the probabilistic MDS process discussed above. Ennis et al (1988) provided evidence that the gradient relating similarity to distance between points in the space could look Gaussian in form, even if the true similarity judgment function was exponential. (Intuitively, the Gaussian-distributed dispersions associated with each stimulus can swamp the exponential similarity function that operates within trials.) Thus, Nosofsky's (1985b) observation of a Gaussian similarity gradient (which was obtained within a deterministic MDS framework) can be reconciled with Shepard's (1987) proposed exponential law. Conceptually, Shepard's (1987) law concerns a cognitive similarity-judgment process that operates at the level of individual trials, but when the stimuli are highly confusable, one needs to also model the variability that is associated with the stimulus representations across trials.

FUTURE DIRECTIONS

The recent influx of probabilistic scaling approaches to the study of similarity and classification is a welcome development. In addition to the increased power and generality that is afforded by probabilistic scaling models, the fundamental assumption that objects give rise to probabilistic representations in perception and memory seems conceptually well motivated. With this increased power and flexibility, however, it becomes even more important to search for invariances across tasks when fitting these models to similarity data. Thus, the probabilistic scaling representation that is derived by fitting a model to a matrix of same-different data should be useful for predicting how subjects will identify, classify, and recognize the same set of objects.

What lies in the near future regarding the intersection between similarity scaling and cognitive process models? One direction likely to be pursued will involve the use of similarity scaling to constrain connectionist/distributed models of perception and cognition. The recent explosion of studies that demonstrate the potential power of connectionist models is slowly giving way to efforts to test these models rigorously on their psychological validity and predictive, quantitative accuracy. An impediment to developing rigorous tests is that there is often no associated theory of stimulus representation in these models. A particular form of input representation might be assumed a priori, or the investigators might search for an input representation that "works" (in the sense that when used with the model, it delivers the desired behavior).

The process-model approach to scaling that I advocate in this chapter could easily be incorporated in the connectionist-modeling domain. For example, suppose that one wanted to test the quantitative predictions of a given connectionist model of category learning. As a first step, one could fit the model to a set of identification learning data. This step would involve

searching for the input representation of the stimuli that maximized the likelihood of the data with respect to the model—the portion of the modeling in which a scaling representation is derived. Then, using the same basic connectionist architecture and scaling representation, one could use the model to predict category learning in situations involving the same set of objects. With an invariant scaling representation, we gain greater confidence that a successful connectionist model captures psychological processes in a meaningful way.

Literature Cited

Anderson, J. R. 1976. *Language, Memory, and Thought.* Hillsdale, NJ: Erlbaum

Anderson, J. R. 1990. *The Adaptive Character of Thought.* Hillsdale, NJ: Erlbaum

Ashby, F. G. 1988. Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Percept. Psychophys.* 44:195-204

Ashby, F. G., ed. 1992. *Multidimensional Models of Perception and Cognition.* Hillsdale, NJ: Erlbaum

Ashby, F. G., Gott, R. E. 1988. Decision rules in the perception and categorization of multidimensional stimuli. *J. Exp. Psychol.: Learn. Mem. Cognit.* 14:33-53

Ashby, F. G., Lee, W. W. 1991. Predicting similarity and categorization from identification. *J. Exp. Psychol.: General.* 120:150-72

Ashby, F. G., Maddox, W. T. 1990. Integrating information from separable psychological dimensions. *J. Exp. Psychol.: Hum. Percept. Perform.* 16:598-612

Ashby, F. G., Maddox, W. T. 1992. Complex decision rules in categorization: contrasting novice and experienced performance. *J. Exp. Psychol.: Hum. Percept. Perform.* In press

Ashby, F. G., Perrin, N. A. 1988. Toward a unified theory of similarity and recognition. *Psychol. Rev.* 95:124-50

Ashby, F. G., Townsend, J. T. 1986. Varieties of perceptual independence. *Psychol. Rev.* 93:154-79

Blough, D. S. 1988. Quantitative relations between visual search speed and target-distractor similarity. *Percept. Psychophys.* 43:57-71

Carroll, J. D. 1976. Spatial, non-spatial, and hybrid models for scaling. *Psychometrika* 41:439-63

Carroll, J. D., Arabie, P. 1980. Multidimensional scaling. *Annu. Rev. Psychol.* 31:607-49

Carroll, J. D., Wish, M. 1974. Models and methods for three-way multidimensional scaling. In *Contemporary Developments in Mathematical Psychology,* ed. D. H. Krantz, R. C. Atkinson, R. D. Luce, P. Suppes, pp. 57-105. San Francisco: W. H. Freeman

Cliff, N. 1973. Scaling. *Annu. Rev. Psychol.* 21:473-506

Corter, J. E. 1987. Similarity, confusability, and the density hypothesis. *J. Exp. Psychol.: General* 116:238-49.

Ennis, D. M. 1988. Confusable and discriminable stimuli: comment on Nosofsky 1986 and Shepard 1986. *J. Exp. Psychol.: General* 117:408-11

Ennis, D. M. 1992. Modelling similarity and identification when there are momentary fluctuations in psychological magnitudes. See Ashby 1992. In press

Ennis, D. M., Mullen, K. 1986. A multivariate model for discrimination methods. *J. Math. Psychol.* 30:206-19

Ennis, D. M., Palen, J., Mullen, K. 1988. A multidimensional stochastic theory of similarity. *J. Math. Psychol.* 32:449-65

Estes, W. K. 1986. Array models for category learning. *Cognit. Psychol.* 18:500-49

Garner, W. R. 1974. *The Processing of Information and Structure.* New York: Wiley

Gati, I., Tversky, A. 1982. Representations of qualitative and quantitative dimensions. *J.*

Exp. Psychol.: Hum. Percept. Perform. 8:325–40

Gescheider, G. A. 1988. Psychophysical scaling. *Annu. Rev. Psychol.* 39:169–200

Getty, D. J., Swets, J. B., Swets, J. A., Green, D. M. 1979. On the prediction of confusion matrices from similarity judgments. *Percept. Psychophys.* 26:1–19

Gillund, G., Shiffrin, R. M. 1984. A retrieval model for both recognition and recall. *Psychol. Rev.* 91:1–67

Gluck, M. A. 1991. Stimulus generalization and representation in adaptive network models of category learning. *Psychol. Sci.* 2:50–55

Gluck, M. A., Bower, G. H. 1988. Evaluating an adaptive network model of human learning. *J. Mem. Lang.* 27:166–95

Green, D. M., Swets, J. A. 1966. *Signal Detection Theory and Psychophysics.* New York: Wiley

Hefner, R. A. 1958. *Extensions of the law of comparative judgment to discriminable and multidimensional stimuli.* PhD thesis. Univ. Michigan

Heiser, W. J. 1988. Selecting a stimulus set with prescribed structure from empirical confusion frequencies. *Brit. J. Math. Stat. Psychol.* 41:37–51

Henley, N. M. 1969. A psychological study of the semantics of animal terms. *J. Verb. Learn. Verb. Behav.* 8:176–84

Hintzman, D. L. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychol. Rev.* 93:411–28

Holman, E. W. 1979. Monotonic models for asymmetric proximities. *J. Math. Psychol.* 20:1–15

Hurwitz, J. B. 1990. *A hidden-pattern unit network model of category learning.* PhD thesis. Harvard Univ.

Hutchinson, J. W., Lockhead, G. R. 1977. Similarity as distance: a structural principle for semantic memory. *J. Exp. Psychol.: Hum. Learn. Mem.* 3:660–78

Kadlec, H., Townsend, J. T. 1992. Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *J. Math. Psychol.* In press

Krumhansl, C. L. 1978. Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. *Psychol. Rev.* 85:445–63

Kruschke, J. K. 1990. *A connectionist model of category learning.* PhD thesis. Univ. Calif. Berkeley

Luce, R. D. 1963. Detection and recognition. In *Handbook of Mathematical Psychology,* ed. R. D. Luce, R. R. Bush, E. Galanter, 1:103–90. New York: Wiley

MacKay, D. B. 1989. Probabilistic multi-dimensional scaling: an anisotropic model for distance judgments. *J. Math. Psychol.* 33:187–205

Marley, A. A. J. 1992. Developing and characterizing multidimensional Thurstone and Luce models for identification and preference. See Ashby 1992, In press

Massaro, D. W. 1987. *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry.* Hillsdale, NJ: Erlbaum

Massaro, D. W., Friedman, D. 1990. Models of integration given multiple sources of information. *Psychol. Rev.* 97:225–52

Medin, D. L., Schaffer, M. M. 1978. Context theory of classification learning. *Psychol. Rev.* 85:207–38

Monahan, J. S., Lockhead, G. R. 1977. Identification of integral stimuli. *J. Exp. Psychol.: General* 106:94–110

Mullen, K., Ennis, D. M. 1987. Mathematical formulation of multivariate Euclidean models for discrimination methods. *Psychometrika* 52:235–49

Nosofsky, R. M. 1984. Choice, similarity, and the context theory of classification. *J. Exp. Psychol.: Learn. Mem. Cognit.* 10:104–14

Nosofsky, R. M. 1985a. Luce's choice model and Thurstone's categorical judgment model compared: Kornbrot's data revisited. *Percept. Psychophys.* 37:89–91

Nosofsky, R. M. 1985b. Overall similarity and the identification of separable-dimension stimuli: a choice model analysis. *Percept. Psychophys.* 38:415–32

Nosofsky, R. M. 1986. Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol.: General* 115:39–57

Nosofsky, R. M. 1987. Attention and learning processes in the identification and categorization of integral stimuli. *J. Exp. Psychol.: Learn. Mem. Cognit.* 13:87–109

Nosofsky, R. M. 1988a. Exemplar-based accounts of relations between classification, recognition, and typicality. *J. Exp. Psychol.: Learn. Mem. Cognit.* 14:700–8

Nosofsky, R. M. 1988b. On exemplar-based exemplar representations: reply to Ennis (1988). *J. Exp. Psychol.: General* 117:412–14

Nosofsky, R. M. 1988c. Similarity, frequency, and category representations. *J. Exp. Psychol.: Learn. Mem. Cognit.* 14:54–65

Nosofsky, R. M. 1989. Further tests of an exemplar-similarity approach to relating identification and categorization. *Percept. Psychophys.* 45:279–90

Nosofsky, R. M. 1990. Relations between exemplar-similarity and likelihood models of classification. *J. Math. Psychol.* 34:393–418

Nosofsky, R. M. 1991a. Relation between the

rational model and the context model of classification. *Cognit. Sci. Rep.* 39, Indiana Univ.

Nosofsky, R. M. 1991b. Stimulus bias, asymmetric similarity, and classification. *Cognit. Psychol.* 23:94–140

Nosofsky, R. M. 1991c. Tests of an exemplar model for relating perceptual classification and recognition memory. *J. Exp. Psychol.: Hum. Percept. Perform.* 17:3–27

Nosofsky, R. M. 1992. Exemplars, prototypes, and similarity rules. In *Essays in Honor of William K. Estes,* Vol. 1, ed. A. Healy, S. Kosslyn, R. Shiffrin. Hillsdale, NJ: Erlbaum. In press

Nosofsky, R. M., Clark, S. E., Shin, H. J. 1989. Rules and exemplars in categorization, identification, and recognition. *J. Exp. Psychol.: Learn. Mem. Cognit.* 15:282–304

Oden, G. C., Massaro, D. W. 1978. Integration of featural information in speech perception. *Psychol. Rev.* 85:172–91

Podgorny, P., Garner, W. R. 1979. Reaction time as a measure of inter- and intraobject visual similarity: letters of the alphabet. *Percept. Psychophys.* 26:37–52

Ramsay, J. O. 1977. Maximum-likelihood estimation in multidimensional scaling. *Psychometrika* 42:241–66

Reed, S. K. 1972. Pattern recognition and categorization. *Cognit. Psychol.* 3:382–407

Sergent, J., Takane, Y. 1987. Structures in two-choice reaction- time data. *J. Exp. Psych.: Hum. Percept. Perform.* 13:300–15

Shaw, M. L. 1982. Attending to multiple sources of information: I. The integration of information in decision making. *Cognit. Psychol.* 14:353–409

Shepard, R. N. 1957. Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika* 22:325–45

Shepard, R. N. 1958a. Stimulus and response generalization: deduction of the generalization gradient from a trace model. *Psychol. Rev.* 65:242–56

Shepard, R. N. 1958b. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. *J. Exp. Psychol.* 55:509–23

Shepard, R. N. 1964. Attention and the metric structure of the stimulus space. *J. Math. Psychol.* 1:54–87

Shepard, R. N. 1986. Discrimination and generalization in identification and classification: comment on Nosofsky. *J. Exp. Psychol.: General* 115:58–61

Shepard, R. N. 1987. Toward a universal law of generalization for psychological science. *Science* 237:1317–23

Shepard, R. N. 1988. Time and distance in generalization and discrimination: reply to Ennis (1988). *J. Exp. Psychol.: General* 117:415–16

Shepard, R. N. 1990. Neural nets for generalization and classification: comment on Staddon and Reid (1990). *Psychol. Rev.* 97:579–80

Shepard, R. N. 1991. Integrality versus separability of stimulus dimensions: evolution of the distinction and a proposed theoretical basis. In *Perception of Structure,* ed. J. Pomerantz, G. Lockhead. Washington, DC: APA. In press

Shepard, R. N., Chang, J. J. 1963. Stimulus generalization in the learning of classifications. *J. Exp. Psychol.* 65:94–102

Shepard, R. N., Hovland, C. I., Jenkins, H. M. 1961. Learning and memorization of classifications. *Psychol. Monogr.* 75:1–41

Shepard, R. N., Kannappan, S. 1991. Connectionist implementation of a theory of generalization. In *Advances in Neural Information Processing Systems,* ed. R. Lippmann, J. Moody, D. Touretzky, Vol. 3. San Mateo, CA: Morgan Kaufman. In press

Shepard, R. N., Kilpatric, D. W., Cunningham, J. P. 1975. The internal representation of numbers. *Cognit. Psychol.* 7:82–138

Shin, H. J. 1990. *Similarity-scaling studies of "dot patterns" classification and recognition.* PhD thesis. Indiana Univ.

Sjoberg, L., Thorslund, C. 1979. A classificatory theory of similarity. *Psychol. Res.* 40:223–47

Smith, J. E. K. 1980. Models of identification. In *Attention and performance,* ed. R. Nickerson, 8:129–58. Hillsdale, NJ: Erlbaum

Smith, J. E. K. 1982. Recognition models evaluated: a commentary on Keren and Baggen. *Percept. Psychophys.* 31:183–89

Staddon, J. E. R., Reid, A. K. 1990. On the dynamics of generalization. *Psychol. Rev.* 97:576–78

Takane, Y., Sergent, J. 1983. Multidimensional models for reaction times and same-different judgments. *Psychometrika* 48:393–423

Takane, Y., Shibayama, T. 1985. Comparison of models for stimulus recognition data. In *Proceedings of the Multidimensional Data Analysis Workshop.* Leiden: DSWO Press

Thurstone, L. L. 1927a. A law of comparative judgment. *Psychol. Rev.* 34:273–86

Thurstone, L. L. 1927b. Psychophysical analysis. *Am. J. Psychol.* 38:368–89

Townsend, J. T. 1971. Theoretical analysis of an alphabetic confusion matrix. *Percept. Psychophys.* 9:40–50

Townsend, J. T., Landon, D. E. 1982. An experimental and theoretical investigation of the constant-ratio rule and other models

of visual letter confusion. *J. Math. Psychol.* 25:119–62

Townsend, J. T., Landon, D. E. 1983. Mathematical models of recognition and confusion in psychology. *Math. Soc. Sci.* 4:25–71

Tversky, A. 1977. Features of similarity. *Psychol. Rev.* 84:327–52

Tversky, A., Gati, I. 1982. Similarity, separability, and the triangle inequality. *Psychol. Rev.* 89:123–54

Tversky, A., Hutchinson, J. W. 1986. Nearest neighbor analysis of psychological spaces. *Psychol. Rev.* 93:3–22

Wickens, T. D., Olzak, L. A. 1989. The statistical analysis of concurrent detection ratings. *Percept. Psychophys.* 45:514–28

Young, F. W. 1984a. Scaling. *Annu. Rev. Psychol.* 35:55–81

Young, F. W. 1984b. The general Euclidean model. In *Three-Mode Models for Data Analysis*, ed. H. Law, C. Snyder, R. McDonald, J. Hattie. New York: Praeger

Zinnes, J. L., MacKay, D. B. 1983. Probabilistic multidimensional scaling: complete and incomplete data. *Psychometrika.* 48:27–48