THE TIME COURSE OF PERCEPTUAL CATEGORIZATION

Thomas J. Palmeri

Vanderbilt University

December 20, 1999

To appear in M. Ramscar & U. Hahn (Eds.), Similarity and Categorization,

Oxford University Press.

address correspondences to:

Thomas J. Palmeri

Department of Psychology

301 Wilson Hall

Vanderbilt University

Nashville, TN 37240

Tel : (615) 343-7900

Fax : (615) 343-8449

e-mail : Thomas.J.Palmeri@Vanderbilt.edu

WWW: http://www.vanderbilt.edu/AnS/psychology/cogsci/palmeri/home.html

THE TIME COURSE OF PERCEPTUAL CATEGORIZATION

Perceptual categorization is a fundamental aspect of human cognition. Any time we decide that some visually presented object is a dog rather than a cat, a bottle rather than a jar, or a tree rather than a shrub, we are making a categorization decision based on the perceptual attributes of that object (see Sloman & Malt, this volume). The goal of this chapter is to review a class of theories of perceptual categorization with particular emphasis on how these theories account for the time course of these judgements. A key assumption of these theories is that perceptual categorization depends on the underlying perceptual similarity relations among objects. As should be apparent from the other chapters in this volume, this general assumption is not without some controversy (see Ahn, this volume; Hampton, this volume; Markman, this volume; Sloman & Malt, this volume).

Considerable progress has been made over the past two decades developing and testing formal models of categorization. A formal model takes a significant step beyond mere verbally stated theories by well specifying, in either mathematical or computational detail, the hypothesized psychological processes thought to underlie a particular behavioral phenomenon. In the case of perceptual categorization, this requires specifying the details of a variety of hypothesized cognitive processes thought to be involved in making a categorization judgment. Some of these include specifying what perceptual information is provided by the sensory system and how that information is represented, how that perceptual information is compared with what has been previously learned, how this previously learned category information is represented in memory, and how a categorization decision is made based on the comparison of perceptual information with the stored category representations. For example, is perceptual information represented in terms of features or dimensions, is the comparison of perceptual information to what has been learned based on similarities to stored category representations or is some other matching process used, are categories represented in terms of rules, abstract prototypes, or exemplars, and so on.

A complete review of the various formal models of perceptual categorization which have been proposed is clearly beyond the scope of the present chapter (see Ashby, 1992; Cohen & Massaro, 1992; Estes, 1994; Komatsu, 1992; Lamberts, 1997; Nosofsky, 1992a, 1992b; E.E. Smith & Medin, 1981).

However, it is probably fair to say that, until quite recently, nearly every formal model of perceptual categorization has attempted to account for categorization choice probabilities (which category will a given stimulus be classified into), but has not addressed the time course of such judgements. There are a number of very good reasons for theories of perceptual categorization to attempt to account for the time to make a category choice.

REASONS TO INVESTIGATE THE TIME COURSE OF CATEGORIZATION

By far, the most common psychological measures of cognitive phenomenon involve accuracy and response time. Clearly, a complete theory of a complex cognitive phenomenon such as perceptual categorization must provide a specification of the time required for each sub-process which is engaged during a task (from which can be generated response time predictions) as well as provide an account of the particular responses that are produced (from which can be generated accuracy predictions).

Generally, testing models of perceptual categorization that only account for response probabilities requires the use of experimental paradigms in which subjects make errors. Subjects are only provided limited amounts of training − the amount of training is carefully calibrated to raise subjects to only an intermediate level of expertise (e.g., Medin & Schaffer, 1978) − or subjects learn category structures with high degrees of overlap which cannot be learned perfectly − errors are required by how the categories are defined (e.g., Ashby & Gott, 1988). In real world situations, errors may often be observed in novices (e.g., young children or newcomers to a specialized domain of expertise), when categorization decisions are difficult because of some degree of category overlap (e.g., telling a Cabernet from a Merlot), when subjects are placed under extreme speed stress to make quick categorization decisions (e.g., classifying objects when driving at high speeds), or when perceptual information is unavailable (e.g., classifying a partially occluded object). However, in many experimental paradigms and natural settings, categories are well learned and errors may be quite rare. Yet, relatively large differences in the time needed to make a categorization judgment can be observed. As a classic example, although subjects rarely make errors in judging whether a robin is a bird or whether a chicken is a bird, subjects are significantly faster to judge the former statement than the latter one (E.E. Smith, Rips, & Shoben, 1974).

Another reason to investigate the time course of categorization judgments is that, in many cases, it has proven quite difficult to distinguish between formal process models based on response probability data alone. In some situations, models which make quite different assumptions about the underlying processes involved in a categorization decision have proved to be nearly indistinguishable based solely on their ability to account for response probabilities (e.g., Maddox & Ashby, 1993; McKinley & Nosofsky, 1995). Yet, once these models have been provided with the processing dynamics required to account for response times as well, they can and do make different predictions which are testable (e.g., Nosofsky & Alfonso-Reese, in press; Nosofsky, Alfonso-Reese, & Palmeri, 1997; Nosofsky & Palmeri, 1997a, 1997b).

This chapter discusses a class of recent models of categorization that have aimed to account for both categorization accuracy and response times. Of the numerous theoretical approaches to categorization which have been offered, only decision boundary models (e.g., Ashby, Boynton, & Lee, 1994; Ashby & Maddox, 1994; Maddox & Ashby, 1996) and exemplar-based models (e.g., Lamberts, 1995, 1998; Nosofsky & Palmeri, 1997a; Palmeri, 1997a, 1999a) have been extended to account for response time data. Largely because the theme of this volume is on the role of similarity in categorization, this chapter will focus exclusively on the recent exemplar-based approaches (see also Heit, this volume).

A FORMAL MODEL OF CATEGORIZATION: THE GENERALIZED CONTEXT MODEL

Although other formal models have been proposed which are based on prototypes (e.g., Homa, 1984; Homa, Sterling, & Trepel, 1981), rules (e.g., Nosofsky, Palmeri, & McKinley, 1994; Nosofsky & Palmeri, 1998; Palmeri & Nosofsky, 1993, 1995; Erickson & Kruschke, 1998), decision boundaries (e.g., Ashby & Gott, 1998; Maddox & Ashby, 1993), and association weights in connectionist networks (e.g., Gluck & Bower, 1988), exemplar models have been some of the most successful and most rigorously tested theories (e.g., Estes, 1994; Nosofsky, 1992a, 1992b), have had some of the most applicability in other domains of cognition (e.g., Goldinger, 1998; E.R. Smith & Zarate, 1992; Valentine & Endo, 1992), and, most importantly for present purposes, have made the furthest inroads into developing a complete account of the time course of categorization (e.g., Lamberts, 1995, 1998; Nosofsky & Palmeri, 1997a; Palmeri, 1997a, 1997b, 1999a; but see Markman, this volume, for a critique of traditional laboratory methods used to test theories of categorization).

Exemplar models assume that categories are represented in terms of the individual instances of a category which have been experienced and stored in memory (see Heit, this volume, for applications of exemplar models to experimental paradigms encouraging the use of pre-experimental background knowledge). Classifying an object into some category is based on how similar that object is to the stored exemplars of that category relative to the object's similarity to exemplars in other categories. Exemplar models are unique from most other formal models of categorization in that little abstraction takes place in the creation of the category representation. They can be contrasted with prototype models, rule-based models, and decision boundary models which all assume the creation of some sort of summary representation of a category which is stripped of information about the particular instances which have been experienced. This is not to suggest that exemplar models cannot allow abstraction to occur. People can certainly form an "image" or a description of the “prototypical” cat, dog, or flower – according to exemplar models these abstractions do not occur at the time of storage, rather they occur on-line in the service of some particular task (Barsalou, 1990). Arguably, there may be an adaptive advantage to storing exemplars rather than creating abstractions. Creating abstractions requires that the organism be prescient to what information will be required at a later time for survival. By contrast, retaining detailed information about particular instances allows the organism to generate flexible abstractions online which may have been unanticipated when the categories are first acquired.

While a number of variants of exemplar models have been proposed (e.g., Estes, 1986; Hintzman, 1986; Myers, Lohmeier, & Well, 1994), I will begin by summarizing the well known generalized context model (GCM; Nosofsky, 1984, 1986), a generalization of the context model of Medin and Schaffer (1978) which builds on classic models relating stimulus identification and categorization (e.g., Shepard, 1957, 1958; Shepard & Chang, 1963; Shepard, Hovland, & Jenkins, 1961). This model serves as the basis for the models of the time course of perceptual categorization outlined later in this chapter.

Formal Specification of the GCM

The GCM assumes that objects are represented as points in a multidimensional psychological space. These dimensions may represent such characteristics of an object as its size, shape, or color, as illustrated in Figure 1. When applying the model, these multidimensional representations serve as the inputs to the GCM. In some circumstances, the psychological dimensions are rather obvious and can be specified a priori, such as those shown in Figure 1 (see also Nosofsky, Gluck, Palmeri, McKinley, & Gauthier, 1994). In other cases, the psychological dimensions of the stimuli are less obvious, but may be determined using techniques such as multidimensional scaling (MDS; Shepard, 1980).[1]

Presented items and the stored exemplars are represented as points in this multidimensional psychological space. The coordinates of that point specify the values that object has on each of the psychological dimensions. Similarity between an item i and stored exemplar j is a function of distance, d_{ij}, in the psychological space. Formally, this distance is given by

$$d_{ij} = \left(\sum_{m=1}^{M} w_m \left| x_{im} - x_{jm} \right|^r \right)^{1/r}, \tag{1}$$

where x_{im} is the coordinate value of stimulus i on dimension m, w_m is a selective attention weight on dimension m (discussed later), and r specifies the distance metric. The well-known Euclidean distance metric is specified when $r = 2$, which is typically the case for relatively integral dimensions such as the saturation and brightness of colors; a city-block metric is specified when $r = 1$, which is typically the case for relatively separable dimensions such as form and color (see Garner, 1974; Shepard, 1964, 1991).

Similarity is a decreasing function of distance. Items that are close together in psychological space are similar to one another, those far apart are dissimilar. Formally, the similarity between item i and exemplar j, s_{ij}, is an exponentially decreasing function of distance

$$s_{ij} = \exp(-c\, d_{ij}), \tag{2}$$

where c is a scaling parameter (Shepard, 1987; see Tversky, 1977; Markman, this volume, Keane & Smyth, this volume, and Pothos & Chater, this volume, for alternative conceptualizations of similarity).

A critically important assumption of the GCM is that dimensions can be differentially attended to depending on their diagnosticity in a particular categorization context. For example, if all members of category A are black and all members of category B are white, then an optimal strategy (Nosofsky, 1984, in press) would be to attend more to the color of the items than to other dimensions. This is instantiated by stretching dimensions that are attended to and shrinking dimensions that are not attended to. This increases distances along the attended dimension and decreases distances along the unattended dimension, causing commensurate changes in similarities. Figure 1 illustrates the situation where color is attended to more than shape or size, causing the space to stretch along the color dimension and shrink along the unattended dimensions. Selective attention is formalized as weights, w_m, in the distance metric given in Equation 1. A relatively larger weight on a dimension reflects relatively more attention to that dimension; a weight of zero would make differences along that dimension completely irrelevant for purposes of similarity computations.

The evidence that an item is a member of a particular category is found by summing the similarities to all exemplars of that category (and may include other exemplars from background knowledge of related categories as well, see Heit, this volume). The more an item is similar to exemplars of a category, the greater the evidence that it belongs to that category. Formally, the evidence, E_{Ai}, that an item i belongs to category A is given by

$$E_{Ai} = \sum_{k \in A} s_{ik}, \tag{3}$$

where k indices all members of category A and s_{ik} is the similarity between item i and exemplar k defined above.

The probability of judging item i as a member of category A depends on how much evidence there is for category A relative to the other category options. Formally, in the case of two categories, A and B, the probability of classifying item i as a member of category A, P(A|i), is given by

$$P(A \mid i) = \frac{b_A\left(E_{Ai} + \beta\right)^{\gamma}}{b_A\left(E_{Ai} + \beta\right)^{\gamma} + b_B\left(E_{Bi} + \beta\right)^{\gamma}}, \tag{4}$$

where b_A is the response bias for category A, β is a background noise term which acts like a guessing term when evidence for any category is low, and γ is a response scaling parameter. Increasing γ makes response more deterministic[2] (McKinley & Nosofsky, 1996, Maddox & Ashby, 1993); while this response scaling parameter was initially a post hoc addition to the GCM, it has recently been provided a specific process interpretation, as discussed later (Nosofsky & Palmeri, 1997a). Without bias, background noise, or response scaling, Equation 4 reduces to

$$P(A \mid i) = \frac{E_{Ai}}{E_{Ai} + E_{Bi}}; \tag{5}$$

the probability of classifying item i as a member if category A is simply given by the ratio of the evidence for category A to the total evidence for any category.

The GCM has successfully accounted for a wide variety of fundamental categorization phenomenon (e.g., Nosofsky, 1984, 1986, 1988a) and has successfully described relations between categorization and other processes, such as stimulus identification and recognition memory (e.g., Nosofsky, 1986, 1987, 1988b, 1991; Nosofsky & Zaki, 1998; but see Palmeri & Flanery, in press, Palmeri & Nosofsky, 1995). A connectionist instantiation of the GCM, ALCOVE (Kruschke, 1990), has supplied a mechanism which learns the selective attention weights and associations between exemplars and categories. ALCOVE has accounted for a category learning data in a number of paradigms (e.g., Kruschke, 1992; Nosofsky et al., 1994; Nosofsky, Kruschke, & McKinley, 1992; Nosofsky & Palmeri, 1996; Palmeri, 1999b).

Although the GCM has been so successful, it is limited in not providing an account of the time course of categorization decisions. Two recent extensions of the GCM have addressed this shortcoming by incorporating dynamic assumptions of the evolution of perceptual representations (Lamberts, 1995, 1998), making assumptions about the time needed to retrieve and compare perceptual information with exemplars stored in memory (Nosofsky & Palmeri, 1997a; Palmeri, 1997a), and by incorporating a process account of categorization decision making (Nosofsky & Palmeri, 1997a).

THE EXTENDED GENERALIZED CONTEXT MODEL

Categorization decisions can vary according to the amount of time pressure one is put under. For example, although I may know that a dolphin is a mammal, under conditions in which I must make an extremely rapid judgement, I might instead classify a dolphin as a fish. This section will review some recent empirical and theoretical work by Lamberts (1995, 1998) which has extensively investigated the effects of time pressure on perceptual categorization.

As a static model, the GCM does not specify the time course of how perceptual representations come to be generated over time. Sensory and perceptual processes are clearly time-dependent (e.g. P.L. Smith, 1995), and it seems reasonable to suppose that some perceptual information might become available sooner than other information. The Extended Generalized Context Model (EGCM; Lamberts, 1995, 1998) specifies how perceptual representations evolve over time. The model has successfully accounted for experimental paradigms in which subjects must classify objects under varying amounts of speed stress.

As with the GCM, EGCM assumes that similarity between item i and exemplar j is a decreasing function of distance in psychological space. Formally, combining Equations 1 and 2,

$$s_{ij}(t) = \exp\left[-c\left(\sum inc_m(t)\, u_m \left| x_{im} - x_{jm} \right|^r \right)^{1/r} \right], \tag{6}$$

where $s_{ij}(t)$ is now a time-varying similarity measure. Whereas the GCM assumed a single selective attention weight for dimension m, w_m, the EGCM breaks this parameter up into two independent components, a time varying parameter, $inc_m(t)$, which specifies whether dimension m has been included

(1) or not (0) at time t, and a constant parameter, u_m, which is the utility (diagnosticity) of dimension m (akin to the original attention weights in the GCM). The remainder of the EGCM is identical to the original GCM (i.e., Equations 3 and 4)

The main unique aspect of the EGCM is its assumption that perceptual information is provided in a time-dependent manner. The perceptual representation of an object is not available instantaneously, nor is all the perceptual information about an object provided at once. Rather, some dimensions may be processed more rapidly than others. For example, especially salient dimensions may be available earlier than less salient dimensions. The inclusion parameter, $inc_m(t)$, indicates whether or not a particular perceptual dimension has been made available to later processing systems.

The EGCM assumes that the probability of including dimension m at time t, given that it was not yet included before time t, is a constant inclusion rate, q. The probability that an event occurs at time t given that it has not yet occurred is termed the hazard function of the process (Luce, 1986). The only probability density function with a constant hazard function is the exponential distribution. Therefore, the cumulative inclusion probability that dimension m is included at or before time t is given by

$$i_m(t) = 1 - \exp(-q_m t), \tag{7}$$

where q_m is the inclusion rate for dimension m.

It must be emphasized that the inclusion probabilities and utility values can be entirely independent of one another. Although a highly salient dimension may be included much more rapidly than another dimension, that salient dimension may be completely nondiagnostic for determining category membership. This independence allows the EGCM to predict reversals in categorization choice probabilities as a function of the amount of time provided to the subject to perceptually process that object. Under extreme time pressure (via response deadlines or rapid stimulus presentations), subjects may classify an object on the basis of the most salient dimensions, albeit inappropriately. With more processing time, subjects will have all dimensions available to them and may classify an object on the basis of the most diagnostic dimensions (see also Busemeyer & Townsend, 1993; Goldstein & Busemeyer, 1992).

To illustrate predictions of the EGCM, I will describe a general pattern of results Lamberts has observed across a series of experiments using varying stimuli, category structures, and experimental designs. Consider the category structure shown in Table 1 (Lamberts & Freeman, 1999a, 1999b). Stimuli in these experiments were realistic computerized renderings of lamps that varied in the shape of the base, upright, shade, and top. Abstract value 0 along each dimension is associated with category A, while abstract value 1 is associated with category B. The exception to this is stimulus A5 (1111) which is actually the modal prototype of category B. Other category structures Lamberts has tested contain such exceptional stimuli as well (e.g., Lambert, 1995).

Lamberts and Freeman (1999a) had subjects learn to classify the stimuli in Table 1 into the correct category. Later, in a second stage, the same stimuli were presented again to be classified, but only for a short period of time (between 100 and 300 ms in Experiment 1, between 33 and 200ms in Experiment 2). One of the most theoretically challenging result from these studies was the finding that the exception stimulus (A5 in the present example) was classified into the wrong category under extreme speed stress but the correct category under little or no speed stress. Note that if subjects had simply not learned to correctly classify the exception, which is clearly the most difficult stimulus to learn, then subjects would classify it incorrectly regardless of speed stress. These results are theoretically important because if stimulus information from all perceptual dimensions were available at the same time, such as is implicitly assumed by the GCM and other models, then there is absolutely no way to predict a categorization crossover without assuming that other parameters change systematically as function of time (see Lamberts, 1995).

By contrast, if dimensions are gradually sampled over time, as is assumed by the EGCM, then a categorization reversal can be predicted. In the present example, contrast the effects of processing time on classifying A1 (0001) versus A5 (1111). For A1, so long as more than just the shade (dimension 3) and the top (dimension 4) have been included (perceptually sampled), subjects will tend to classify that stimulus as a member of category A. Stimulus A1 can be classified greater than chance quite rapidly, with minimal perceptual information, and so no category reversal would be expected as a function of presentation time or response deadline. Stimulus A5, on the other hand, must be sampled completely before it can be correctly classified. If three or fewer dimensions are sampled, then A5 would match exemplars in category B, and be classified incorrectly. The EGCM predicts just such patterns of categorization reversals, as were empirically observed.

In applications of the EGCM to experimental data, Lamberts and colleagues have not only accounted for qualitative patterns of results such as those described above, but have rigorously fitted the model to complete sets of classification probabilities for training and transfer stimuli at a variety of exposure durations and response deadlines. For example, using a variety of category structures, Lamberts (1995, 1998) trained subjects to classify a set of stimuli and later tested them on the training stimuli and on new transfer stimuli under a wide range of response deadlines (ranging from very quick deadlines of 100ms, to moderate deadlines of 600ms, to no deadline at all). Levels of response deadlines were either blocked separately (Lamberts, 1995) or mixed together within a block (Lamberts, 1998). By allowing separate inclusion rate parameters and utility parameters for each dimension, the EGCM was able to provide an excellent account of classification probabilities under the entire range of response deadlines, with all parameters remaining fixed across all manipulations of deadline in various experiments. The EGCM accounted for changes in response probability as a function of response deadline because of differential dimensional sampling as a function of time.

The original GCM confounds dimensional utility and dimensional salience in weighting psychological dimensions in computations of perceptual similarity. In fits of the GCM to experimental data, the best-fitting value of a selective attention weight could reflect the relative diagnosticity of that dimension, or the salience of that dimension, or both. As a further test of the independence of utility and salience assumed by the EGCM, Lamberts (1995, 1998) varied the assignment of physical dimensions to abstract dimensions in a category structure. Dimensional utility should be entirely a function of the abstract category structure – dimensions that are highly diagnostic for determining category membership should have high utility regardless of how they are physically instantiated. Dimensional salience should be entirely a function of the physical instantiation of the stimuli − dimensions that are easy to process should have high salience, hence faster inclusion rates, regardless of their diagnosticity for determining category membership. In experiments reported by Lamberts (1995, 1998), two different groups of subjects learned the same abstract category structure, with the same type of stimuli, but with different mappings of physical dimensions onto abstract dimensions. For example, while dimension 1 may be highly diagnostic for both groups, thereby having high dimensional utility, dimension 1 might be highly salient for one group (say the shape of the shade of a lamp), but might be less salient for the other group (say the shape of the base of a lamp). Lamberts (1995, 1998) successfully fitted the EGCM to data from both groups simultaneously, constraining all parameter values for the two groups of subjects to be the same (with inclusion rates specified for the particular mapping of physical to abstract dimensions for each group).

Other Evidence for the Time-Varying Nature of Perceptual Processing

The key assumption of the EGCM is that similarity changes systematically over time because of the time-dependent sampling of perceptual dimensions. Other evidence also suggests that perceptual similarity varies systematically as a function of time. For example, Sergent and Takane (1987) carried out multidimensional scaling analyses on reaction time data obtained in same-different judgements from a variety of stimulus set: colors varying in saturation and brightness, circles varying in size and orientation of a radial line, parallelograms of varying size and tilt, and rectangles varying in height and width. Under tachistoscopic presentation and instructions stressing both speed and accuracy, Sergent and Takane revealed that the psychological space of the various stimulus sets were quite different from those derived under conditions of unlimited viewing and unspeeded instruction. Divergence from unspeeded conditions occurred at the level of dimensionality of the space, distance metric, and dimensional orientation. Similarly, Klein (1982) found that stimulus variables which had no effect on perceived similarity under unlimited viewing conditions had significant effects when viewing time (exposure) was limited.

Goldstone and Medin (1994) investigated how similarities of scenes were influenced by the kind of correspondence between features of objects in the scenes and by the amount of time provided to make a similarity judgment. To borrow an example they provided, suppose in scene A there is a boy in a white shirt throwing a football, and in scene B there is a boy in a brown shirt throwing a baseball. The match between the feature brown of the football in scene A and the feature brown of the shirt of the boy in scene B constitutes a match out of place (MOP). The objects these features belong to do not correspond because they are dissimilar and play different thematic roles in the scene. Suppose that the boys in both scenes wore green pants. The match between the features green in both scenes constitutes a match in place (MIP). While MIPs should increase similarity between scenes more than MOPs, time is required to develop the appropriate correspondences between the various objects in a scene to determine whether matches between features constitute a MIP or a MOP. Hence, MOPs may have some influence on similarity when time is quite limited, but their influence should decrease as more time is allowed. Experimental work demonstrated that both MIPs and MOPs increase the similarity between scenes, but MIPs increased similarity more. With sufficient time, MOPs became much less influential, and MIPs became more influential. Under time pressure, however, MOPs could have more influence than MIPs under certain circumstances (see Markman, this volume, for further discussion of the role of relational properties in judgments of similarity).

Finally, research on decision making under risk and under uncertainty has shown that judgments and preferences change under conditions of time pressure as well (see Svenson & Maule, 1993). For example, Ben Zur and Breznitz (1981; see also Wright, 1974) had subjects choose gambles under three different levels of time pressure. When decision time was limited, subjects chose the less risky alternative (a high probability of a small gain) over a risky one (a low probability of a large loss). In concert with claims made by Lamberts (1995, 1998), one could argue that large losses, and other extremely negative consequences, are very salient properties of a gamble, even though they have a very low probability of occurrence (see also Svenson & Edland, 1987). Under time constraints, highly salient negative consequences are processed quite rapidly and loom large in subjects' decision making. Similar claims for the importance of dimensional salience in probabilistic inference were made by Wallsten and Barton (1982). In making inferences, subjects under time pressure used only the most salient subset of dimensions used by subjects not under pressure. They proposed a model of in which features were processed in order from most to least salient (see also Busemeyer & Townsend, 1982).

THE EXEMPLAR-BASED RANDOM WALK MODEL

While the EGCM focused on the temporal aspects of perceptual processing and its influence on similarity computations, the Exemplar-Based Random Walk Model (EBRW; Nosofsky & Palmeri, 1997a; Palmeri, 1997a) introduced dynamics into the memory retrieval and decision making components of the GCM. The basic architecture of the EBRW is quite similar to that of the GCM, so this section will highlight those elements of the EBRW that are unique. It should be noted that the EBRW assumes the same static perceptual process as the GCM; although perceptual processing is not assumed to occur instantaneously, the EBRW has assumed that perceptual processing takes the same amount of time for all stimulus attributes, and thereby assumes that all perceptual information is available at the same time. Given Lamberts' work, this is clearly a limitation of the EBRW.

The EBRW assumes that objects are represented as points in a multidimensional psychological space, that categories are represented in terms of stored exemplars, and that similarity between an item and an exemplar is a decreasing function of distance in the space. Within the EBRW, these assumptions, which are borrowed from the GCM, are merged with assumptions about the time course of memory retrieval as a function of practice, which are borrowed from the instance theory of automaticity (Logan, 1988, 1990, 1992; see Palmeri, 1997a). The EBRW goes beyond either of these models by incorporating a random walk process to generate categorization decisions (Link, 1975; Link & Heath, 1975; Luce, 1986).

I will provide a brief summary of how categorization decisions are made by the EBRW and then discuss each component in some detail. According to the EBRW, when an object is presented to be categorized, all stored exemplars in memory race to be retrieved with rates proportional to their similarity to the presented object. Although highly similar exemplars are most likely to be retrieved, because of the stochastic nature of the retrieval process, dissimilar exemplars may be retrieved with some nonzero probability as well. In a binary-choice context, the category label associated with the retrieved exemplar provides incremental evidence for either a category A or category B decision. In general, categorization is not based on a single retrieval, however. Rather, this incremental evidence drives a random walk decision process. As shown in Figure 2, a random walk accumulates evidence over time for one of the two possible category options. If a category A exemplar is retrieved, the random walk counter moves up one step toward the A barrier; otherwise, if a category B exemplar is retrieved, the counter moves down toward the B barrier. The time to take each step of the random walk, ΔT, is a function of the memory retrieval time. If a retrieval causes the counter to reach one of the barriers, then a response is made, otherwise another memory retrieval is made. As will be discussed in some detail, response times generated by the EBRW are a function of how quickly exemplar information can be retrieved from memory and how consistently the random walk counter moves toward only one of the threshold barriers.

I will begin by discussing the dynamics of the memory retrieval component of the EBRW, which necessitates briefly describing Logan's (1988, 1992) instance theory of automaticity, upon which many of these assumptions were based. I will then discuss the random walk component of the model, and how memory retrieval interfaces with this decision process. Finally, I will summarize a number of experimental results which support the EBRW.

Instance Theory of Automaticity

Instance theory (Logan, 1988, 1990, 1992) aims to explain the behavioral changes that are observed during the acquisition of a cognitive skill. Automatic processes are typically characterized as being unconscious, fast, require few processing resources, take place without intention, and are generally contrasted with strategic, or intentional processes (e.g., Kahneman & Treisman, 1984; Logan, 1985; Shiffrin, 1988). Traditional accounts of the development of automaticity posited a decrease in the demand for limited processing resources as the source for seemingly automatic behavior (e.g., LaBerge &

Samuels, 1974; Posner & Snyder, 1974; Shiffrin & Schneider, 1977). Instance theory, by contrast, aimed to understand the deployment of automaticity in processing terms.

Instance theory assumes that the development of automaticity is largely a memory phenomenon. Tasks that are not well learned require conscious application rules, strategies, or algorithms for their completion. For example, a novice searching the woods for prized Chanterelle mushrooms must make regular reference to a set of complex rules for telling them apart from the highly similar, yet quite poisonous, Jack O'Lantern mushroom. Although these rules may come to be internalized, rather than requiring reference to a field manual or some other external aid, categorization still involves a deliberate use of rule-based knowledge. An expert, by contrast, is generally able to recognize a Chanterelle and ignore the Jack O'Lantern with little or no effort, and seemingly without recourse to rules. According to instance theory, this automatic categorization is a result of remembering previous instances that have been stored in memory.

Although this example makes obvious the usefulness of the instance theory for understanding categorization, historically, the areas of categorization and automaticity shared fairly little communication. One reason for this is possibly methodological. Whereas most categorization studies have focused on accuracy, most automaticity studies have focused on response time; whereas most categorization studies have manipulated similarities between stimuli, most automaticity studies have not; whereas most automaticity studies have manipulated the amount of processing resources available for accomplishing some test, most categories studies have not. One of the unique aspects of the EBRW is that it provides a unified account of both perceptual categorization and automaticity (Palmeri, 1997a, 1997b, 1999a).

Instance theory assumes that strategic processes and memory retrieval occur simultaneously. While a person has a choice over what particular strategies they may employ to solve a particular task, memory retrieval occurs without intention. In processing terms, strategic processing and memory retrieval race to completion, with the winner determining the response. The development of automaticity is seen as a shift from the need to use slow, effortful, strategic processes to the ability to rely on direct retrieval of instances from memory.

Instance theory assumes that practice at a task does not causes the algorithm to speed up. Every time a task, such as perceptual categorization, is performed on some object, an instance of the object and its solution are stored in memory. When an object is again presented, stored instances of that object race to be retrieved from memory. As more instances of a particular object are stored, retrieval gets faster and faster (see Logan, 1988, 1992). This is because memory retrieval is also conceptualized as a race; probabilistically, for every additional runner that enters a race, the winning time gets marginally faster. Eventually, memory retrieval becomes fast enough to complete before the strategic processes can be executed. Therefore, according to instance theory, the development of automaticity is due to the storage of additional instance information in memory, which causes retrieval to eventually dominate performance. As acknowledged by Logan (1988), one shortcoming of instance theory was that it did not incorporate similarity-based retrieval. Rather, only identical instances could be retrieved from memory.

Retrieval Assumption in the EBRW

Following the GCM, categories are represented in terms of stored exemplars. Following instance theory, practice causes additional exemplars to be stored in memory, and the addition of these exemplars causes memory retrieval times to speed up.[3] Every time an instance is categorized, a trace of that instance with its category label is stored in memory. When item i is presented, all exemplars race to be retrieved with rates proportional to their similarity to that item, s_{ij}. Largely for reasons of analytic tractability, the retrieval times are assumed to be exponentially distributed.[4] Therefore, the probability density that exemplar j is retrieved at time t given probe item i is given by

$$f(t) = s_{ij} \exp(-s_{ij}\, t)\,, \tag{8}$$

where s_{ij} is the similarity computed using Equation 2.

The assumption of exponentially distributed retrieval times leads to some interesting mathematical results (see Bundesen, 1993; Townsend & Ashby, 1983). Assume there are n independent exponentially distributed memory retrieval processes racing in parallel with rates s_{i1}, s_{i2}, …, s_{in}. We are interested in determining the probability that a particular exemplar wins this race (i.e., is retrieved from memory) and the probability density that this exemplar wins this race at time t. First, the probability that exemplar j wins the race given probe item i is simply given by

$$P(j \mid i) = \frac{s_{ij}}{\sum_k s_{ik}}, \tag{9}$$

where k indices all exemplars taking part in the race. This formula may be recognizable as a form of Luce's (1963) choice rule. As Bundesen (1993) noted, Luce's choice rule emerges as a natural consequence of assuming a race among exponentially distributed processes.

The minimum of a race among n exponentially distributed retrieval times (i.e., the retrieval time density function of the winner of the race) is exponentially distributed with a rate equal to the sum of the individual retrieval rates. The expected completion time for an exponentially distributed random variable with rate s_{ij} is given by

$$E[T \mid i] = 1/s_{ij}. \tag{10}$$

Therefore, the expected completion time for the winner of a race among n exponentially distributed random variables with rates s_{ij} is given by

$$E[T \mid i] = \frac{1}{\sum_{k=1}^{n} s_{ik}}. \tag{11}$$

Consider the case where some stored exemplars belong to category A and some belong to category B. Then, we can rewrite Equation 11 as

$$E[T \mid i] = \frac{1}{\sum_{k \in A} s_{ik} + \sum_{k \in B} s_{ik}}, \tag{12}$$

which from Equation 3 can be rewritten as

$$t_x = E[T \mid i] = \frac{1}{E_{Ai} + E_{Bi}}. \tag{13}$$

The retrieval time is inversely proportional to the total summed similarity of an item to exemplars as memory. As an object becomes more familiar, information is retrieved more rapidly.

Equation 9 gave the probability of retrieving a particular exemplar when presented with object i. We are interested in the probability of retrieving an exemplar from category A or category B when presented with object i, which will be used to determine the direction the counter in the random walk will move. Because retrievals are independent, the probability that an exemplar from category A wins the race is simply given by

$$P(A \mid i) = \sum_{j \in A} P(j \mid i) = \frac{\sum_{j \in A} s_{ij}}{\sum_{k} s_{ik}} = \frac{\sum_{j \in A} s_{ij}}{\sum_{j \in A} s_{ij} + \sum_{j \in B} s_{ij}}, \tag{14}$$

which, using Equation 3, can be rewritten as

$$P(A \mid i) = \frac{E_{Ai}}{E_{Ai} + E_{Bi}}, \tag{15}$$

which is Equation 5, the simplified response rule from the GCM. Hence, the probability that the random walk counter moves in a particular direction is a function of the relative similarity of the probe item i to the stored exemplars of category A.

Random Walks Component of the EBRW

According to the EBRW, the results of memory retrieval provides evidence for a random walk decision process (Link, 1975; Luce, 1986). Random walks, and their continuous-time relatives, diffusion processes, have been very successful as models of choice reaction time in a variety of cognitive phenomenon (e.g., Busemeyer, 1985; Diederich, 1997; Link, 1975; Link & Heath, 1975; Ratcliff, 1978; Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1996; P.L. Smith, 1995; Strayer & Kramer, 1994a, 1994b). In a random walk, evidence gradually accumulates over time for one of the category choices. As shown in Figure 2, accumulated evidence is given by the position of a counter along the vertical axis and time is given by the position along the horizontal axis. A response is made when the random walk counter reaches one of the threshold barriers (A or -B). Although the random walk discussed here is formulated for the two-category case, there are ways of extending it to multiple-category cases as well (e.g., Palmeri, 1997a; Ratcliff & McKoon, 1997).

In an unbiased random walk, the probability of moving up toward A or down toward B would be equal. However, in most categorization contexts, there will be some tendency to move toward one category barrier rather than the other. According to the EBRW, movement of the random walk counter is driven by information retrieved from memory. If a category A exemplar is retrieved, the counter moves up, if a B exemplar is retrieved, the counter moves down. The probability of moving up or down is given by Equation 15. The time to make each step of the random walk is a function of the retrieval time. However, it should be apparent that as more and more exemplars are added to memory, the mean retrieval time gets smaller and smaller; that is, in Equation 13, $E[T|i]$ goes to zero as the number of exemplars in memory approaches infinity. Therefore, in the EBRW, the time to make each step is given by

$$\Delta T = \alpha + E[T \mid i] = \alpha + \frac{1}{E_{Ai} + E_{Bi}}, \tag{16}$$

where α is a constant time associated with each step of the random walk.

Loosely borrowing the terminology used with continuous-time diffusion processes, I will refer to the speed with which the counter moves toward one particular category barrier as the drift rate of the random walk. Panel A of Figure 3 illustrates a case in which there is only a slight drift toward the A barrier, while Panel B illustrates a case in which there in a relatively large drift. As the certainty that an item belongs to category A increases, drift rate increases, which causes response times to increase.

Figure 3 also illustrates a very beneficial property of the random walk. In many cases, to insure reasonable levels of accuracy, the location of the barriers will be set relatively far from the starting point, as shown by the solid lines in the figure. If speed is stressed, however, then the barriers are moved in closer to the starting point. This is in keeping with the notion that under speed stress less information is required before a response is made. If the barriers are moved in, as shown by the dashed lines in Figure 3, the likelihood of errors increases. Especially when drift rate is low, as shown in Panel B, the probability of making an error increases (the counter hits the B barrier before it can return to the correct A barrier). Note, also, that if the barriers are places sufficiently far from the starting point, errors can be made to be exceedingly rare, yet significant differences in response times are still produced.

In addition, the random walk also correctly predicts the shapes of response time distributions which are typically observed. As shown in Figure 3, the predicted response time distributions start at zero, are skewed to the right, and have a long right hand tail which extends out to positive infinity. Other response time models, such as counter models (Pike, 1966, 1968), have often been criticized because they make erroneous predictions of the shape of the response time distributions (Ratcliff, 1978).

These positive points of the random walk aside, one might question whether this complexity is really necessary. Why not base a classification decision on the first exemplar retrieved from memory? Are multiple samples really necessary? Consider the stimulus configuration shown in Figure 4. Consider classification response times for stimuli A2, which is close to exemplars from category A, and for stimulus A4, which is far from the exemplars of category B. One would certainly expect that A4 would be classified more accurately than A2. According to the EBRW, when a stimulus is presented to be classified, all exemplars stored in memory race to be retrieved with rates proportional to their similarity to the presented item. As specified by Equation 15, given this probe stimulus, the probability of retrieving an A exemplar is given by the relative summed similarity to exemplars of category A. Compared with A2, A4 is far from the exemplars of category B. Therefore, the probability of retrieving an exemplar from category A is greater for A4 than for A2. Hence, just a single exemplar retrieval can predict superior classification of A4 than A2, without the need to posit a random walk. In fact, as discussed earlier, the original GCM (without a response scaling parameter) can be given this process interpretation in terms of a similarity-based race to retrieve of exemplars from memory.

But, what about response times? Recall from Equation 13, that the retrieval times are given by $1/(E_{Ai}+E_{Bi})$, the inverse of the absolute summed similarity of i to exemplars from both categories. A2 is very similar to B1, B2, B3, A1, A2, and A3, so its absolute summed similarity will be quite high. By contrast, A4 is really only similar to itself, so its absolute summed similarity will be quite low. Notice that based on a single exemplar retrieval, A4 should be classified more slowly than A2, which seems quite unlikely. So, how can the EBRW predict fast, accurate responses for A4 and slow, inaccurate response for A2? Errors are relatively straightforward. A2 has a smaller drift rate than A4, so the probability of hitting the correct barrier will be smaller for A2 than A4. Response times depend both on the time to retrieve each exemplar, which will be faster for A2 than A4, and on the number of steps required to reach a barrier, which will be greater for A4 than A2. In general, all else being equal, the latter component will be more important than the former, so A4 will be classified more quickly than A2.

As mentioned above, the original GCM can be provided a process interpretation in terms of a single exemplar retrieval. Yet, the EBRW assumes multiple retrievals. Although the EBRW provides a superior account of classification response times, might its account of classification accuracy be somehow compromised? As revealed by Nosofsky and Palmeri (1997a), when the category barriers are equidistant from the starting point, classification probabilities predicted by the EBRW are simple given by

$$P(A \mid i) = \frac{E_{Ai}^{\gamma}}{E_{Ai}^{\gamma} + E_{Bi}^{\gamma}}, \tag{17}$$

where γ is the distance of the barriers from the starting point. With the addition of background noise (see Nosofsky & Alfonso-Reese, in press), this becomes the general response rule used by the GCM (see Equation 4). Work subsequent to the original publication of the GCM revealed that the γ parameter was a necessary, although largely post hoc, addition to the model. This parameter now has a process interpretation in terms of the distance from the starting point in a random walk; Nosofsky & Palmeri (1997a; see also Nosofsky, in press) found that values of γ greater than one generally provided a superior fit to the experimental data reported in conjunction with the original development of the GCM (Nosofsky, 1986).

Tests of the EBRW

I will now briefly summarize a couple of experiments that have tested the ability of the EBRW to account for experimental data. In the first experiment of Nosofsky and Palmeri (1997a), detailed response time data was collected from three subjects in a speeded classification task over several sessions. Stimuli were colors of constant red hue varying in saturation and brightness; the approximate Munsell color configuration for the twelve colors is shown in Figure 5. Subjects learned to classify each of these colors into category A (circle symbols) or category B (square symbols). On each trial, one of the twelve colors was randomly displayed, the subject classified it as an A or a B, and then corrective feedback was supplied. Subjects completed thirty blocks of trials for each of five sessions. Subjects also completed three sessions of similarity ratings in which they made subjective similarity judgments of each pair of colors.

The multidimensional scaling configuration for one of the subjects (Participant 3) is shown in both panels of Figure 6. For this subject, as for the other subjects tested, the MDS configuration did not map perfectly onto the idealized representation shown in Figure 5 (see Sergent & Takane, 1987). Observed mean classification response times, averaged across the final four sessions, are proportional to the size of the circle surrounding the stimulus number in the left panel of Figure 6. Small circles indicate fast responses, large circles indicate slow responses. In general, response times were highly systematic. Stimuli near the boundary between category A and B were classified quite slowly (e.g., stimuli 3 and 12), whereas stimuli far from the boundary were classified quite rapidly (e.g., stimuli 7 and 10). Figure 7 displays mean classification response time as a function of training, averaged across all twelve stimuli. The speed-up in observed response time followed the ubiquitous power law of practice (Newell & Rosenbloom, 1981; see, however, Palmeri, 1999a).

The EBRW was fitted to the observed response time data by minimizing the sum of squared deviations between observations and predictions. The free parameters of the model were a dimensional attention weight, w, from Equation 1, the scaling parameter, c, from Equation 2, the constant time associated with each step of the random walk, α, from Equation 16, the distance of the barriers from the starting point in the walk, A, a base response time, T_R, which reflects the mean residual times for stimulus processing and response execution, and an intercept parameter, k, which scales the random walk times into units of milliseconds. These six parameters were used to fit the 42 observed data points in Figures 6 and 7.

The right panel of Figure 6 displays the EBRW predictions using the same format as was used to display the observed data. A simple comparison between the left (observed) and right (EBRW) panels of Figure 6 reveals the excellent fit of the EBRW. According to the EBRW, when classifying stimuli which are highly similar to both categories, such as stimuli 3 and 12, probes of memory tend to retrieve exemplars from both categories. This causes the random walk to wander between the barriers before threshold is reached, leading to relatively prolonged responses. By contrast, when classifying stimuli that are similar only to a single category, such as stimuli 7 and 10, probes of memory tend to retrieve exemplars from just one category. This causes the random walk to move in consistent fashion toward the correct barrier, leading to relatively rapid responses. Figure 7 displays the speed-ups in response time predicted by the EBRW, which were also highly consistent with the observed data. According to the EBRW, with practice, more instances of each stimulus are stored in memory. Retrieval times get faster as more repetitions of an instance are stored in memory (Logan, 1988, 1992), causing the random walk to accumulate evidence more quickly.

Although the EBRW predicts that stimuli far away from the contrast category will generally be classified more quickly than stimuli close to the contrast category, distance from a "category boundary" is not the only factor determining classification response times. According to the EBRW, if two stimuli are equally distant from the contrast category, yet one of them is more familiar than the other, the familiar stimulus should be classified more quickly than the unfamiliar stimulus. This prediction is in stark contrast to other recent models of perceptual categorization that assume that response times are solely a function of distance from the category boundary (e.g., Ashby et al., 1994; Ashby & Maddox, 1994; Maddox & Ashby, 1996); according to such models, similarity to particular category instances plays no direct role in categorization judgments. The second experiment of Nosofsky and Palmeri (1997a) tested the influence of stimulus familiarity on classification response times.

The stimuli used in this experiment were also colors varying in saturation and brightness. The approximate Munsell configuration in shown in Figure 8. Subjects learned to classify these stimuli, with feedback, during an initial set of training blocks. The critical stimuli were B7 and B8. For one group of subjects, B7 was not presented during training (Condition U7); for another group, B8 was not presented during training (Condition U8). Following training, subjects classified all eight stimuli as quickly as possible without making errors. Note that while B7 and B8 are approximately the same distance from the "category boundary", their level of familiarity is different in the two conditions. The EBRW predicts that, within a condition, for two stimuli which are approximately the same distance from the category boundary, the more familiar one will be classified more quickly. In addition, across conditions, a stimulus will be classified more quickly in the condition in which it is more familiar.

The MDS solution obtained from similarity ratings of the color stimuli is shown in both panels of Figure 9. The boundary of equal summed similarity to training exemplars in both categories is also shown. Although this boundary is not identical across the two conditions (because the set of training stimuli does differ between the two conditions), the differences in its location are extremely slight. Therefore, distance from the "category boundary" for all stimuli can be considered to be equated across the two conditions. Mean classification response time is proportional to the size of the circles in the figure. As expected, stimuli close to the boundary were classified more slowly than stimuli far from the boundary. However, in condition U7, in which stimulus B7 is unfamiliar, B8 was classified more quickly than B7, and in condition U8, in which stimulus B8 was unfamiliar, B7 was classified more quickly than B8. In addition, across conditions, B7 and B8 were classified more quickly in the condition in which they are the familiar stimulus. These results are consistent with the predictions of the EBRW, but inconsistent with the predictions of models that assume that distance from the boundary is the only important factor in determining categorization response times (Ashby et al., 1994).

SUMMARY

This chapter reviewed two recent models of the time course of perceptual categorization that extend the well-known exemplar-based generalized context model of categorization (Nosofsky, 1984, 1986). The EGCM (Lamberts, 1995, 1998) provides a dynamic account of how similarities between exemplars systematically change over time; the model successfully accounted for changes in categorization probabilities as a function of the time provided to make a categorization judgment. The EBRW (Nosofsky & Palmeri, 1997a; Palmeri, 1997a) provides a dynamic account of the memory retrieval and decision-making components of categorization; the model successfully accounted for categorization response times in a variety of tasks. Clearly, given the common genealogy of these two models, one important goal for future research will be to combine these two theoretical approaches in a way that results in a comprehensive, yet testable, theory of the time course of perceptual categorization.

REFERENCES

Ashby, F.G. (1992). Multidimensional models of categorization. In F.G. Ashby (Ed.), Multidimensional models of perception and cognition (pp. 449-483), Hillsdale: NJ: Erlbaum.

Ashby, F.G., Boynton, G., & Lee, W.W. (1994). Categorization response time with multidimensional stimuli. Perception & Psychophysics, 55, 11-27.

Ashby, F.G., & Gott, R.E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 33-53.

Ashby, F.G., & Maddox, W.T. (1994). A response time theory of separability and integrality in speeded classification. Journal of Mathematical Psychology, 38, 423-466.

Barsalou, L.W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T.K. Srull & R.S. Wyer (Eds.), Content and process specificity in the effects of prior experiences: Advances in social cognition, Vol. 3 (pp. 61-88). Hillsdale, NJ: Erlbaum.

Ben Zur, H., & Breznitz, S.J. (1981). The effects of time pressure on risky choice behavior. Acta Psychologica, 47, 89-104.

Bundesen, C. (1993). The relationship between independent race models and Luce's choice axiom. Journal of Mathematical Psychology, 37, 446-471.

Busemeyer, J.R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. Journal of Experimental Psychology, Learning, Memory, and Cognition, 11, 538-564.

Busemeyer, J.R., & Townsend, J.T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review, 100, 432-459.

Cohen, M.M., & Massaro, D.W. (1992). On the similarity of categorization models. In F.G. Ashby (Ed.), Multidimensional models of perception and cognition (pp. 449-483), Hillsdale: NJ: Erlbaum.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. Journal of Mathematical Psychology, 41, 260-274.

Erickson, M.A., Kruschke, J.K. (1998). Rules and exemplars in category learning. Journal of Experimental Psychology: General, 127, 107-140.

Estes, W.K. (1986). Memory storage and retrieval processes in category learning. Journal of Experimental Psychology: General, 115, 155-174.

Estes, W.K. (1994). Classification and cognition. Oxford University Press.

Garner, W.R. (1974). The processing of information and structure. New York: Wiley.

Gluck, M.A., & Bower, G.H. (1988a). From conditioning to category learning: An adaptive network model. Journal of Experimental Psychology: General, 117, 225-244.

Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. Psychological Review, 105, 251-279.

Goldstein, W.M., & Busemeyer, J.R. (1992). The effect of “irrelevant” variables on decision making: Criterion shifts in preferential choice? Organizational Behavior and Human Decision Processes, 52, 425-454.

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

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

Homa, D. (1984). On the nature of categories. Psychology of Learning and Motivation, 18, 49-94.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. Journal of Experimental Psychology: Learning, Memory, and Cognition, 7, 418-439.

Kahneman, D., & Treisman, A.M. (1984). Changing views of attention and automaticity. In R. Parasuraman & R. Davies (Eds.), Varieties of attention (pp. 29-61). New York: Academic Press.

Klein, R. (1982). Patterns of perceived similarity cannot be generalized from long to short exposure durations and vice versa. Perception & Psychophysics, 32, 15-18.

Komatsu, L.K. (1992). Recent views of conceptual structure. Psychological Bulletin, 112, 500-526.

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

LaBerge, D., & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. Cognitive Psychology, 7, 495-531.

Lamberts, K. (1995). Categorization under time pressure. Journal of Experimental Psychology: General, 124, 161-180.

Lamberts, K. (1997). Process models of categorization. In K. Lamberts & D.R. Shanks (Eds.), Knowledge, concepts and categories: Studies in cognition. Cambridge, MA MIT Press.

Lamberts, K. (1998). The time course of categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition, 24, 695-711.

Lamberts, K., & Freeman, R.P.J. (1999a). Categorization of briefly presented objects. Psychological Research, 62, 107-117.

Lamberts, K., & Freeman, R.P.J. (1999b). Building object representations from parts: Tests of a stochastic sampling model. Journal of Experimental Psychology: Human Perception and Performance, 25, 904-926.

Link, S.W. (1975). The relative judgment theory of two choice response time. Journal of Mathematical Psychology, 12, 114-135.

Link, S.W., & Heath, R.A. (1975). A sequential theory of psychological discriminations. Psychometrika, 40, 77-105.

Logan, G.D. (1985). Skill and automaticity: Relations, implications, and future directions. Canadian Journal of Psychology, 39, 367-386.

Logan, G.D. (1988). Toward an instance theory of automatization. Psychological Review, 95, 492-527.

Logan, G.D. (1990). Repetition priming and automaticity: Common underlying mechanisms. Cognitive Psychology, 22, 1-35.

Logan, G.D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 883-914.

Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), Handbook of mathematical psychology (pp. 103-189). New York: Wiley.

Luce, R.D. (1986). Response times: Their role in inferring elementary mental organization. New York: Oxford University Press.

Maddox, W.T., & Ashby, F.G. (1993). Comparing decision bound and exemplar models of categorization. Perception & Psychophysics, 53, 49-70.

Maddox, W.T., & Ashby, F.G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. Journal of Experimental Psychology: Human Perception and Performance, 22, 795-817.

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

McKinley, S.C., & Nosofsky, R.M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. Journal of Experimental Psychology: Human Perception and Performance, 21, 128-148.

McKinley, S.C., & Nosofsky, R.M. (1996). Selective attention and the formation of linear decision boundaries. Journal of Experimental Psychology: Human Perception and Performance, 22, 294-317.

Myers, J.L., Lohmeier, J.H., Well, A.D. (1994). Modeling probabilistic categorization data: Exemplar memory and connectionist nets. Psychological Science, 5, 83-89.

Newell, A., & Rosenbloom, P.S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), Cognitive skills and their acquisition (pp. 1-55). Hillsdale, NJ: Erlbaum.

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

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

Nosofsky, R.M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 87-109.

Nosofsky, R.M. (1988a). Similarity, frequency, and category representations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 54-65.

Nosofsky, R.M. (1988b). Exemplar-based accounts of relations between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 700-708.

Nosofsky, R.M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. Journal of Experimental Psychology: Human Perception and Performance, 17, 3-27.

Nosofsky, R.M. (1992a). Exemplar-based approach to relating categorization, identification, and recognition. In F.G. Ashby (Ed.), Multidimensional models of perception and cognition (pp. 449-483), Hillsdale: NJ: Erlbaum.

Nosofsky, R.M. (1992b). Similarity scaling and cognitive process models. Annual Review of Psychology, 43, 25-53.

Nosofsky, R.M. (in press). Optimal performance and exemplar models of classification. In M. Oaksford & N. Chater (Eds.), Rational models of cognition. Oxford University Press.

Nosofsky, R.M. & Alfonso-Reese, L. (in press). Effects of similarity, practice, and familiarity on speeded classification response times and accuracies: Further tests of an exemplar-retrieval model. Memory & Cognition.

Nosofsky, R.M., Alfonso-Reese, L., & Palmeri, T.J. (1997). Effects of similarity, practice, and familiarity on speeded classification response times and accuracies: Further tests of an exemplar-retrieval model. Indiana University Cognitive Science Technical Report Number 197, Bloomington: Indiana Univeristy.

Nosofsky, R.M., Gluck, M., Palmeri, T.J., McKinley, S.C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). Memory & Cognition, 22, 352-369.

Nosofsky, R.M., Kruschke, J.K., McKinley, S.C. (1992). Combining exemplar-based category representations and connectionist learning rules. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 211-233.

Nosofsky, R.M., Palmeri, T.J., & McKinley, S.C. (1994). Rule-plus-exception model of classification learning. Psychological Review, 101, 53-79.

Nosofsky, R.M., & Palmeri, T.J. (1996). Learning to classify integral-dimension stimuli. Psychonomic Bulletin & Review, 3, 222-226.

Nosofsky, R.M., & Palmeri, T.J. (1997a). An exemplar-based random walk model of speeded classification. Psychological Review, 104, 266-300.

Nosofsky, R.M., & Palmeri, T.J. (1997b). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. Perception & Psychophysics 59, 1027-104.

Nosofsky, R.M., & Palmeri, T.J. (1998). A rule-plus-exception model for classify objects in continuous-dimension spaces. Psychonomic Bulletin & Review, 5, 345-369.

Palmeri, T.J. (1997a). Exemplar similarity and the development of automaticity. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 324-354.

Palmeri, T.J. (1997b). An exemplar-based random walk model of perceptual categorization. In M. Ramscar, U. Hahn, E. Cambouropolos, & H. Pain (Eds.), Proceedings of the Interdisciplinary Workshop On Similarity And Categorisation (pp. 181-187), Edinburgh, Scotland: University of Edinburgh.

Palmeri, T.J. (1999a). Theories of automaticity and the power law of practice. Journal of Experiment Psychology: Learning, Memory, and Cognition, 25, 543–551.

Palmeri, T.J. (1999b). Learning hierarchically structured categories: A comparison of category learning models. Psychonomic Bulletin & Review, 6, 495-503.

Palmeri, T.J., & Flanery, M.A. (in press). Learning about categories in the absence of training: Profound amnesia and the relationship between perceptual categorization and recognition memory. Psychological Science.

Palmeri, T.J., & Nosofsky, R.M. (1993). Generalizations by rule models and exemplar models of category learning. In W. Kintsch (Ed.), Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society, pp. 794-799. Hillsdale, NJ: Erlbaum.

Palmeri, T.J., & Nosofsky, R.M. (1995). Recognition memory for exceptions to the category rule. Journal of Experiment Psychology: Learning, Memory, and Cognition, 21, 548-568.

Pike, A.R. (1966). Stochastic models of choice behaviour: Response probabilities and latencies of finite Markov chain systems. British Journal of Mathematical and Statistical Psychology, 19, 15-32.

Pike, A.R. (1968). Latency and relative frequency of response in psychophysical discrimination. British Journal of Mathematical and Statistical Psychology, 21, 161-182.

Posner, M.I., & Snyder, C.R.R. (1975). Attention and cognitive control. In R.L. Solso (Ed.), Information processing and cognition: The Loyola symposium (pp. 55-85). Hillsdale, NJ: Erlbaum.

Ratcliff, R. (1978). A theory of memory retrieval. Psychological Review, 85, 59-108.

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. Psychological Review, 104, 319-343.

Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. Psychological Science, 9, 347-356.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. Psychological Review, 106, 261-300.

Sergent, J., & Takane, Y. (1987). Structures in two-choice reaction-time data. Journal of Experimental Psychology: Human Perception and Performance, 13, 300-315.

Shepard, R.N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. Psychometrika, 22, 325-345.

Shepard, R.N. (1958). Stimulus and response generalizations: Deductions of the generalization gradient from the trace model. Psychological Review, 65, 242-256.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. Journal of Mathematical Psychology, 1, 54-87.

Shepard, R.N. (1980). Multidimensional scaling, tree-fitting, and clustering. Science, 210, 390-398.

Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. Science, 237, 1317-1323.

Shepard, R.N. (1991). Integrality versus separability of stimulus dimensions: Evolution of the distinction and a proposed theoretical basis. In J. Pomerantz & G. Lockhead (Eds.), Perception of Structure, Washington, D.C.: APA.

Shepard, R.N., & Chang, J.J. (1963). Stimulus generalization in learning of classifications. Journal of Experimental Psychology, 65, 94-102.

Shepard, R.N., Hovland, C.I., & Jenkins, H.M. (1961). Learning and memorization of classifications. Psychological Monographs, 75 (13, Whole No. 517).

Shiffrin, R.M. (1988). Attention. In R.A. Atkinson, R.J. Hernstein, G. Lindzey, & R.D. Luce (Eds.), Stevens' handbook of experimental psychology (pp. 739-811). New York: Wiley.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Smith, E.E., & Medin, D.L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.

Smith, E.E., Rips, L.J., & Shoben, E.J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 81, 214-241.

Smith, E.R., Zarate, M.A. (1992). Exemplar-based model of social judgment. Psychological Review, 99, 3-21.

Smith, P.L. (1995). Psychophysically principled models of visual simple reaction time. Psychological Review, 102, 567-593.

Strayer, D.L., & Kramer, A.F. (1994a). Strategies and automaticity: I. Basic findings and conceptual framework. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 318-341.

Strayer, D.L., & Kramer, A.F. (1994b). Strategies and automaticity: II. Dynamic aspects of strategy adjustment. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 318-341.

Svenson, O., & Edland, A. (1987). Change of preferences under time pressure: Choices and judgements. Scandinavian Journal of Psychology, 28, 322-330.

Svenson, O., & Maule, A.J. (1993). Time pressure and stress in human judgment and decision making. Plenum: New York.

Townsend, J.T., & Ashby, F.G. (1983). Stochastic modeling of elementary psychological processes. Cambridge: Cambridge University Press.

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

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 44A, 671-703.

Wallsten, T.S., & Barton, C. (1982). Processing probabilistic multidimensional information for decisions. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 361-384.

Wright, P. (1974). The harassed decision maker: Time pressure, distractions, and the use of evidence. Journal of Applied Psychology, 59, 555-561.

ACKNOWLEDGEMENTS

This work was supported by Vanderbilt University Research Council Direct Research Support Grants. Correspondence concerning this article may be addressed to Thomas J. Palmeri, Department of Psychology, 301 Wilson Hall, Vanderbilt University, Nashville, TN 37240, email: Thomas.J.Palmeri@Vanderbilt.edu, World Wide Web: http://www.vanderbilt.edu/AnS/psychology/cogsci/palmeri/home.html.

FOOTNOTES

1. MDS techniques elicit similarity judgements from subjects for each pair of stimuli used in an experiment. Given the pairwise similarity matrix, MDS generates a multidimensional configuration of the stimuli which can best account for the collected similarity ratings under the assumption that judged similarity is monotonically related to distance in the space. Statistical arguments can be used to determine the minimal number of psychological dimensions which are required.

2. As γ approaches infinity, $P(A|i)=1$ if $E_{Ai}>E_{Bi}$, otherwise $P(A|i)=0$.

3. Also borrowed from the instance theory is the notion that in applications in which performance is initially governed by a set of rules, memory-based processes (as instantiated by the random walk component) races against the completion of the rule-based process (Palmeri, 1997a, 1999a).

4. Instance theory assumed retrieval times to be distributed as Weibulls (Logan, 1988, 1992).

Table 1.

Category structure used by Lamberts and Freeman (1999a, 1999b).

Category	Stimulus	Base	Upright	Shade	Top
A	A1	0	0	0	1
	A2	0	0	1	0
	A3	0	1	0	0
	A4	1	0	0	0
	A5	1	1	1	1
B	B1	1	1	1	0
	B2	1	1	0	1
	B3	1	0	1	1
	B4	0	1	1	1

FIGURE CAPTIONS

Figure 1. Schematic illustration of stimuli varying along three psychological dimensions (shape, size, and color). The change from the configuration in Panel A to that in Panel B illustrates the role of selective attention in modifying similarities among exemplars.

Figure 2. Schematic illustration of the Exemplar-Based Random Walk Model.

Figure 3. Example traces of the accumulation of evidence by a random walk. Panel A illustrates a relatively low drift rate (probability of moving toward the A barrier equals 0.55), Panel B illustrates a high drift rate (probability of moving toward the A barrier equals 0.75). Above each A barrier, in both panels, is a histogram of the times to hit that barrier. In both panels, the dashed line illustrates barrier positions under conditions of speed stress.

Figure 4. A multidimensional representation of an example category structure discussed in the text.

Figure 5. Illustration of the approximate Munsell color configuration used in Experiment 1 of Nosofsky and Palmeri (1997a)

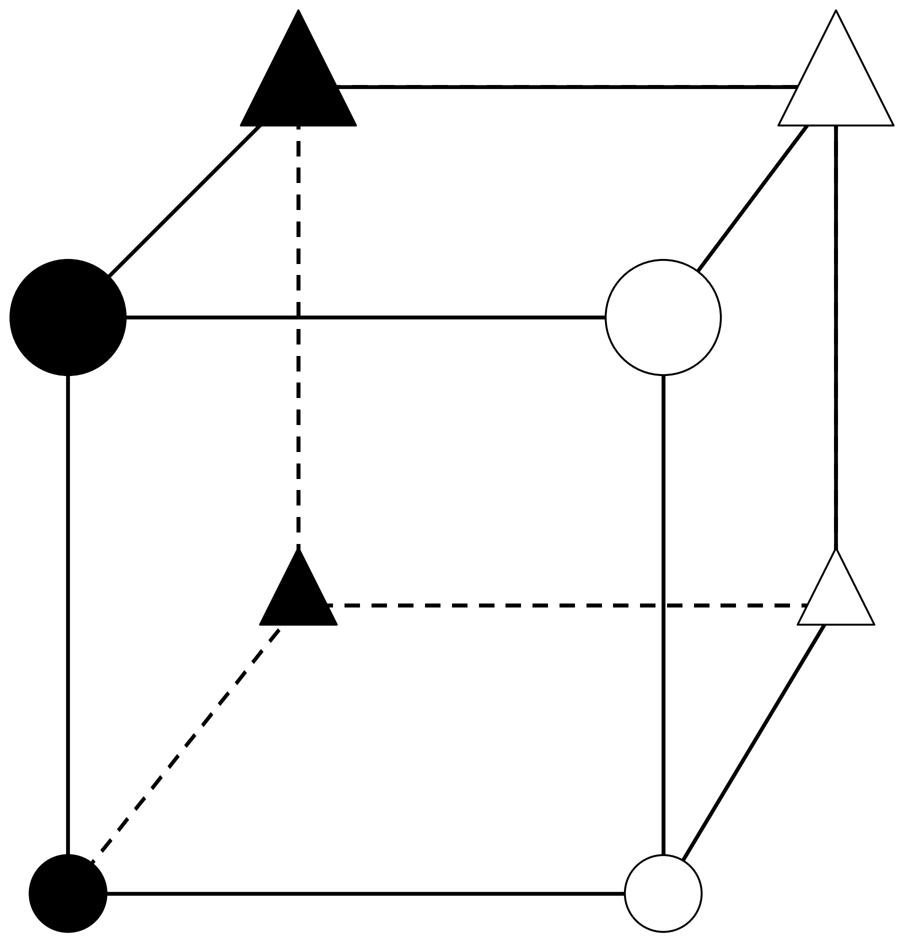
Figure 6. Multidimensional scaling solution of the color stimuli for Participant 3 from Experiment 1 of Nosofsky and Palmeri (1997a). The center of each circle represents the MDS coordinate of the color. The size of the circle is proportional to the mean classification response time for the color. The dashed line is a boundary of equal summed similarity to the exemplars in each category.

Figure 7. Mean response time to classify the colors averaged across blocks of trials for Participant 3 in Experiment 1 of Nosofsky and Palmeri (1997a). The circle symbols indicate the observed data. The solid line indicates the EBRW predictions.

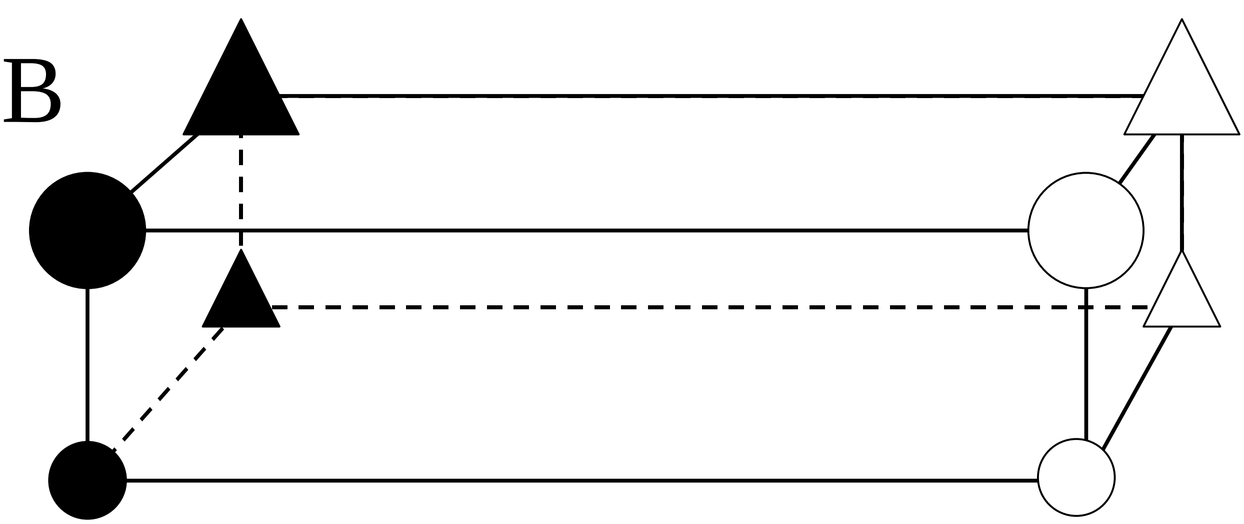
Figure 8. Illustration of the approximate Munsell color configuration used in Experiment 2 of Nosofsky and Palmeri (1997a)

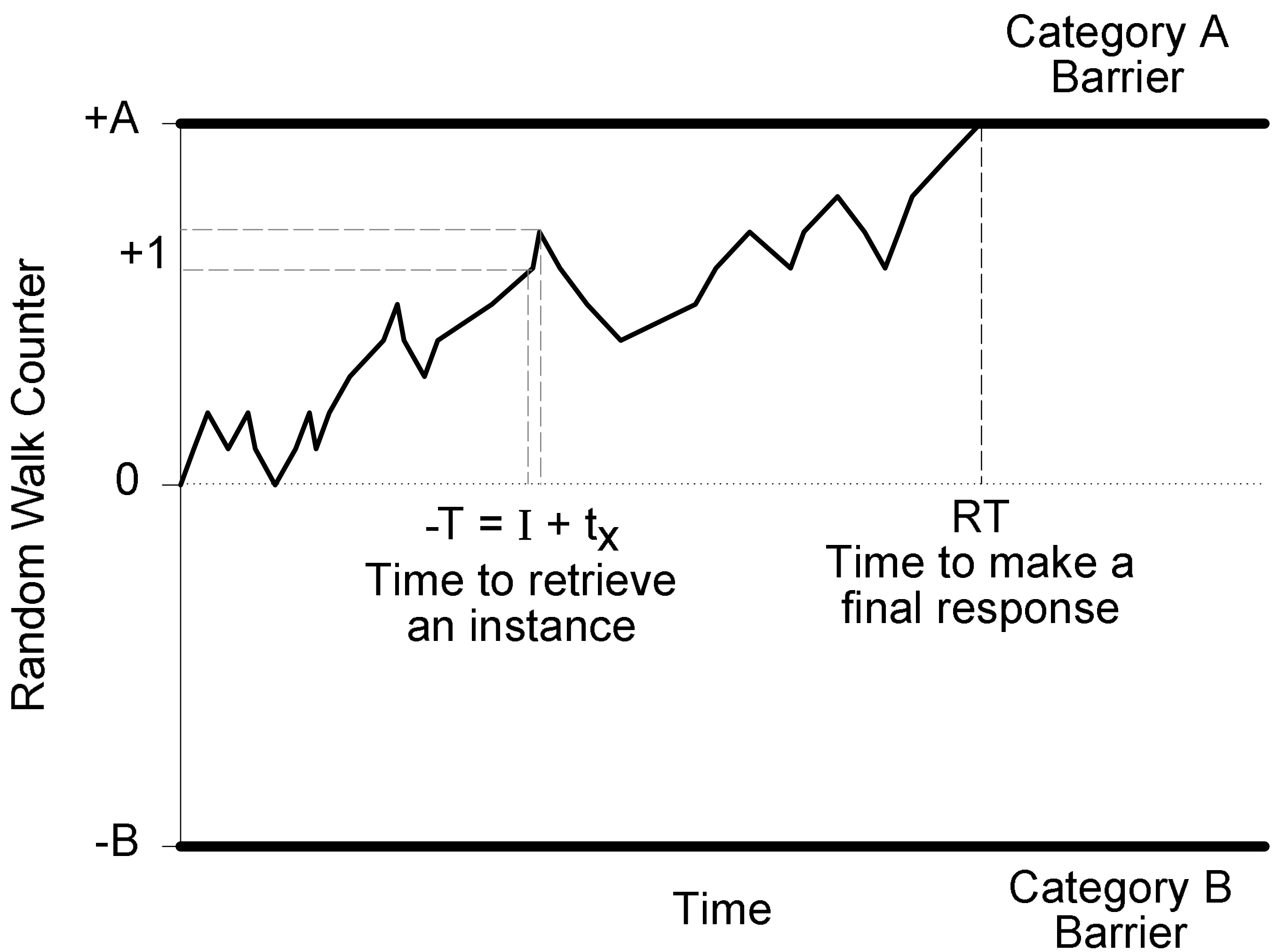
Figure 9. Multidimensional scaling solution of the color stimuli from Experiment 2 of Nosofsky and Palmeri (1997a). The center of each circle represents the MDS coordinate of the color. The size of the circle is proportional to the mean classification response time for the color. The solid line in each panel is a boundary of equal summed similarity to the exemplars in each category. The left panel is data from condition U7, in which stimulus B7 was unfamiliar. The right panel is data from condition U8, in which stimulus B8 was unfamiliar.

A

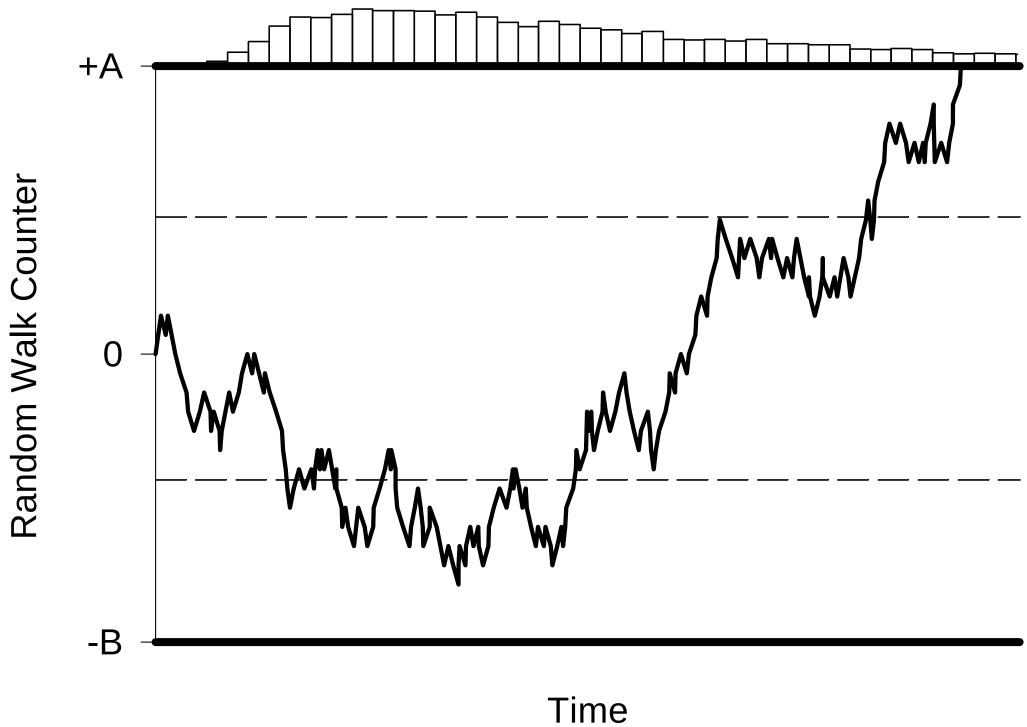


B

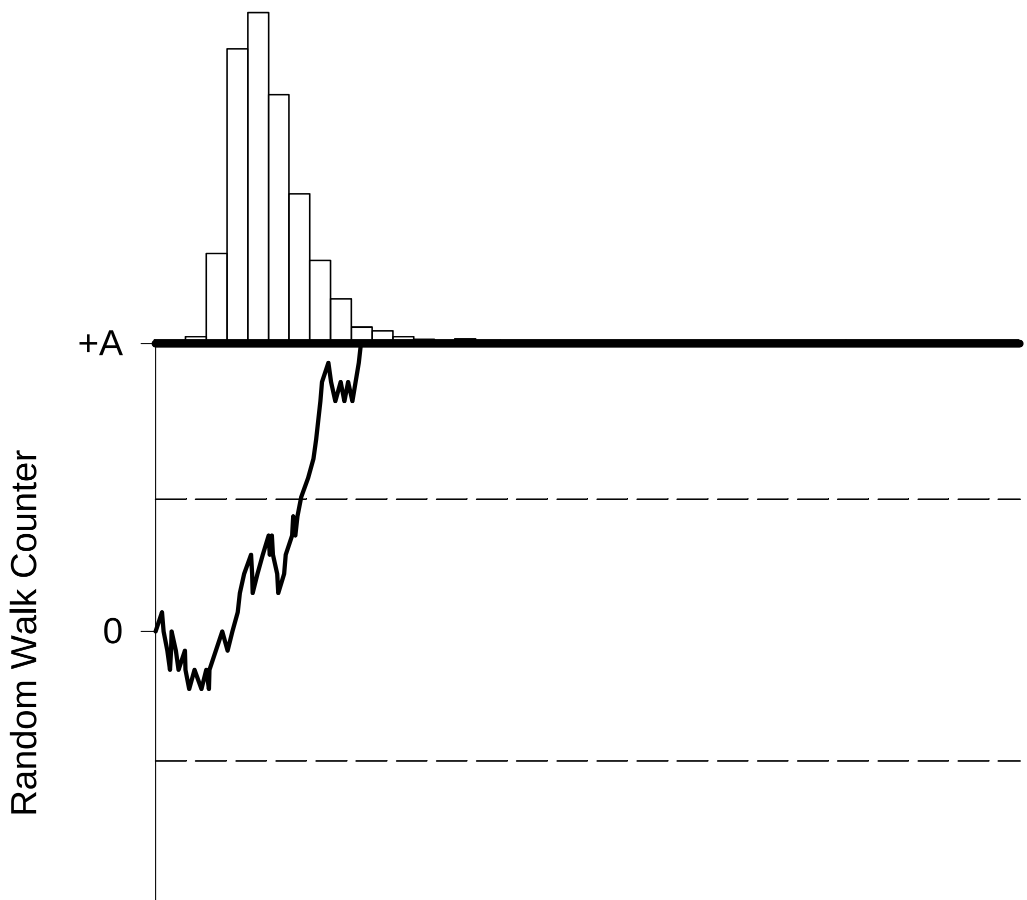


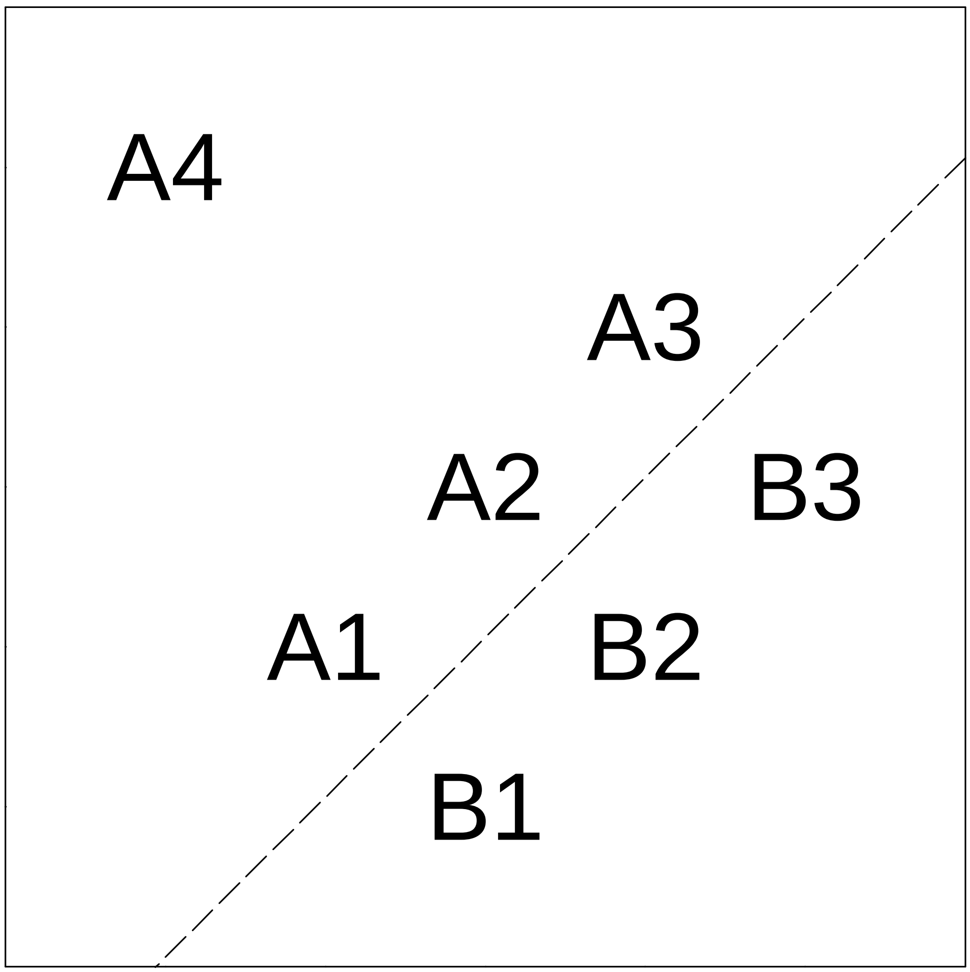

Category A
Barrier
+A
+1
0
-B
Random Walk Counter
-T = I + t_x
Time to retrieve
an instance
RT
Time to make a
final response
Time
Category B
Barrier

A

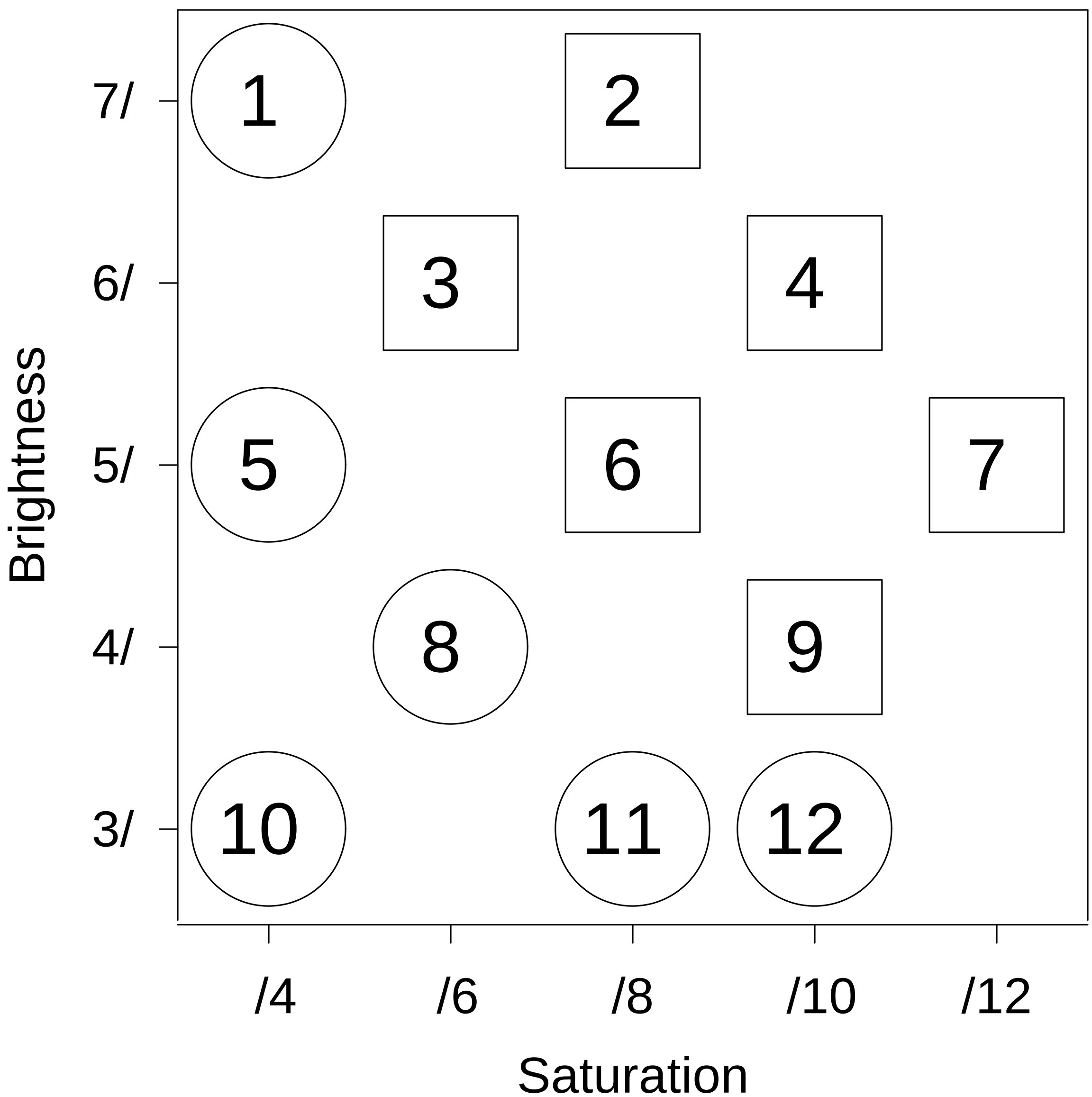


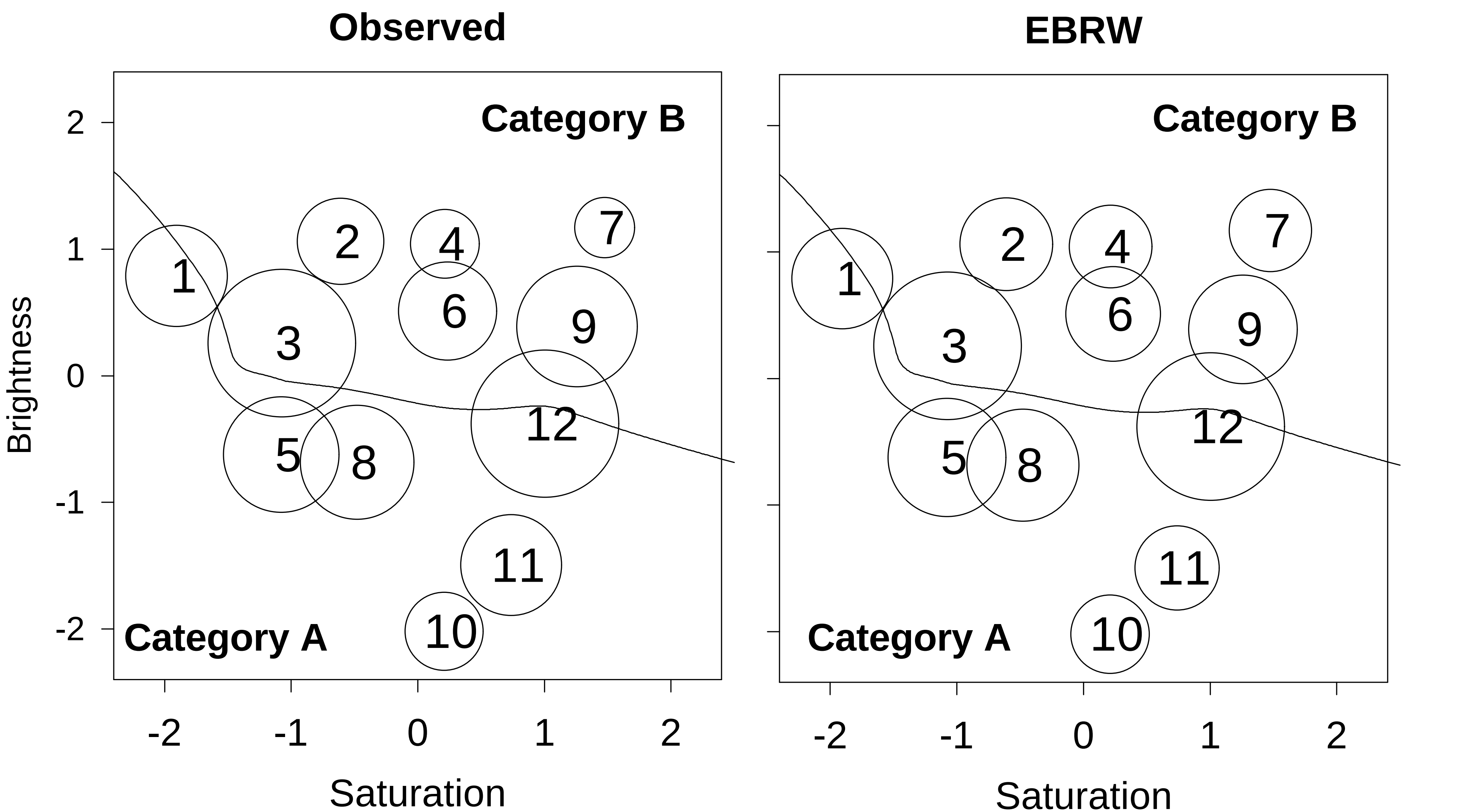
B

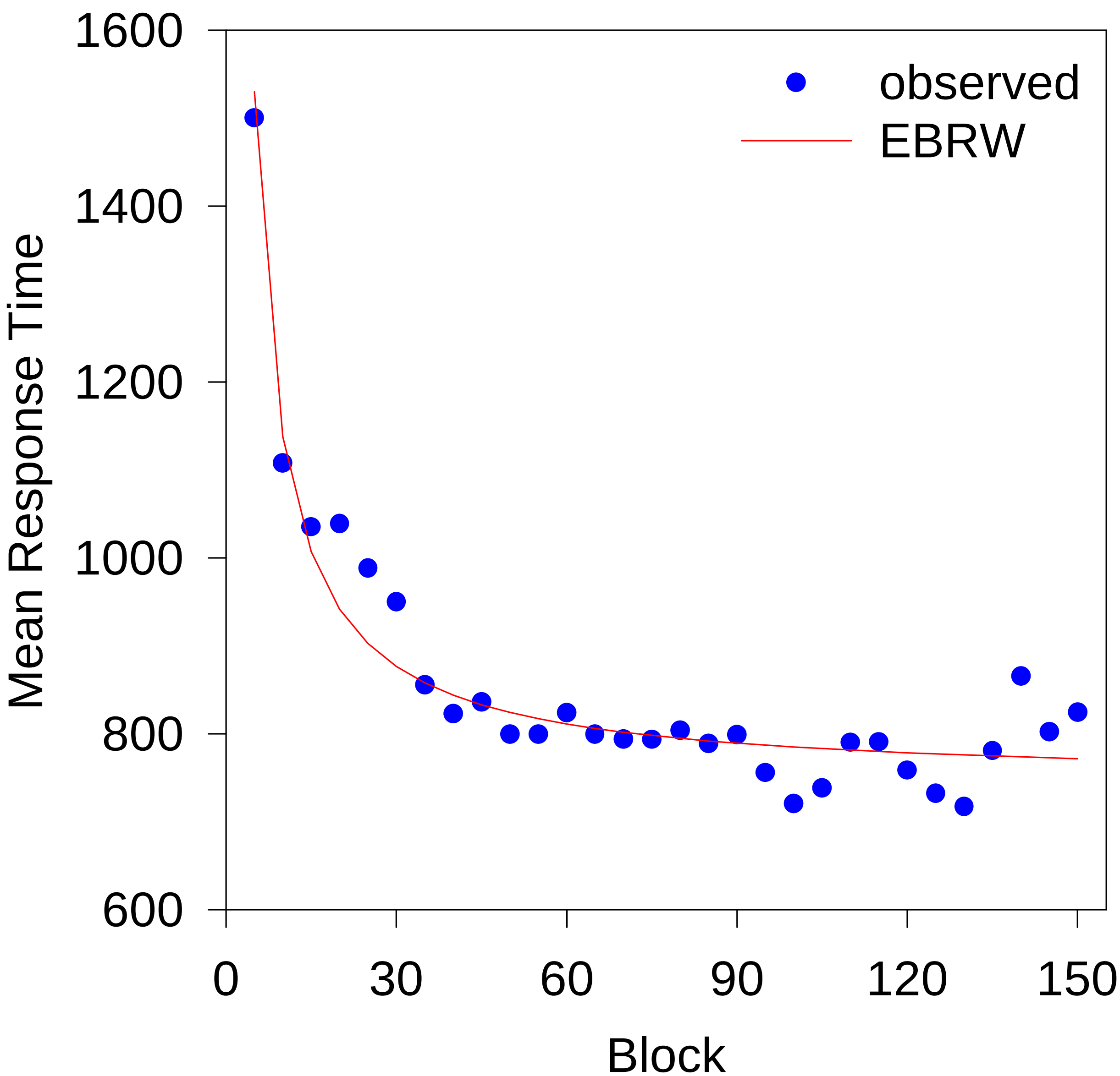




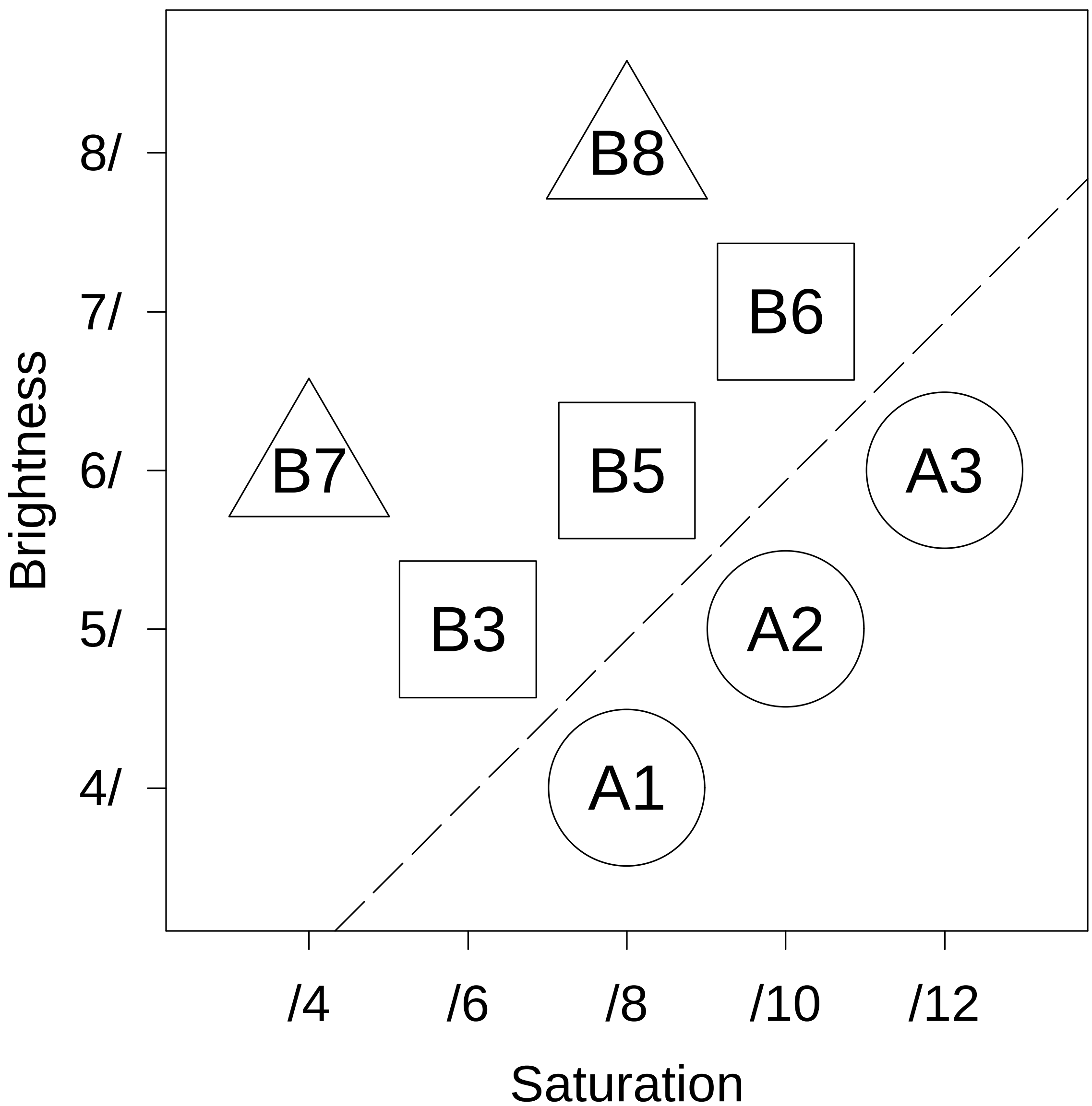
A4
A3
A2
B3
A1
B2
B1


1
2
3
4
5
6
7
8
9
10
11
12
7/
6/
5/
4/
3/
Brightness
/4
/6
/8
/10
/12
Saturation

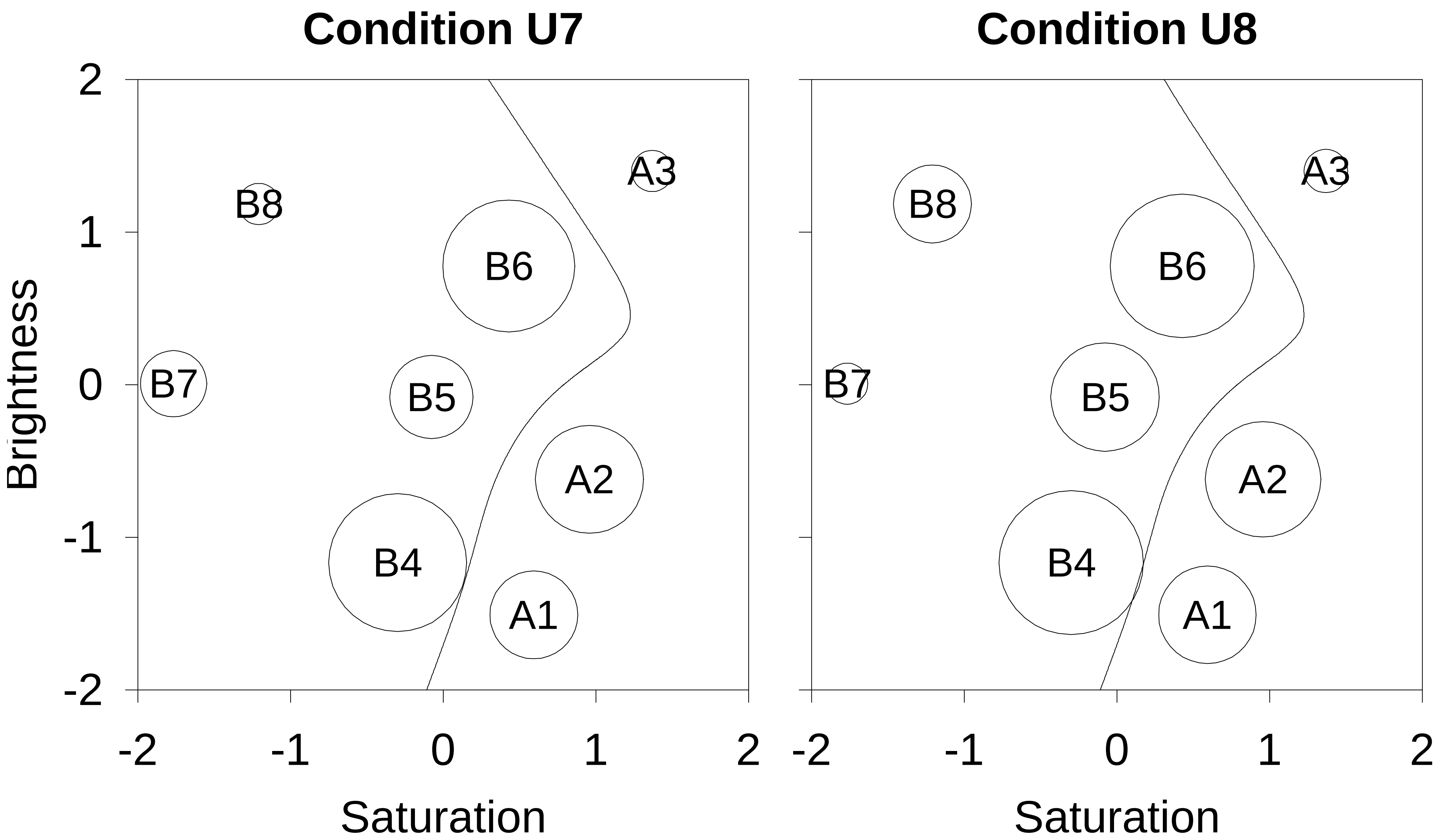

Observed
EBRW
Category B
Category A
1
2
3
4
5
6
7
8
9
10
11
12
Brightness
Saturation
2
1
0
-1
-2



1600
1400
1200
1000
800
600
Mean Response Time
0
30
60
90
120
150
Block
observed
EBRW



B8
B6
B7
B5
A3
B3
A2
A1
8/
7/
6/
5/
4/
Brightness
/4
/6
/8
/10
/12
Saturation


Condition U7
Condition U8
Brightness
Saturation
Saturation
2
1
0
-1
-2
-2
-1
0
1
2
-2
-1
0
1
2
B8
A3
B6
B7
B5
A2
B4
A1
B8
A3
B6
B7
B5
A2
B4
A1